CHAPTER SEVEN

Computational methods for the analysis of solution small-angle X-ray scattering of biomolecules: ATSAS

Haydyn D.T. Mertens*
European Molecular Biology Laboratory, Hamburg Unit, Hamburg, Germany
*Corresponding author: e-mail address: hmertens@embl-hamburg.de

Contents

Abstract

The ATSAS software suite provides a comprehensive set of programs for the processing, analysis and modeling of small-angle scattering data, tailored for but not limited to data acquired on biological macromolecules. In this review the major components and developments in the ATSAS package are described, with a focus on user driven application. Data reduction, analysis and modeling approaches and strategies will be introduced and discussed. At the time of writing the latest package, ATSAS 3.1, is freely available for academic users at: https://www.embl-hamburg.de/biosaxs/software.html.

Methods in Enzymology, Volume 678
ISSN 0076-6879
https://doi.org/10.1016/bs.mie.2022.09.019

1. Introduction

Small-angle X-ray scattering (SAXS) is a widely used technique in structural biology, with a multitude of specific applications developed for liquid state samples under near-native conditions (Brosey & Tainer, 2019; Da Vela & Svergun, 2020; Svergun, Koch, Timmins, & May, 2013; Trewhella, 2022). Although anisotropic methods are in use, primarily in the soft-matter community, the majority of biological solution SAXS experiments are conducted on isotropic solutions of biomacromolecules. As the requirements for robust sample preparation are arguably less demanding than for higher resolution methods, SAXS data can be readily acquired for many systems and provide a stand-alone approach to low-resolution characterization and an unmatched level of method complementarity. Indeed, a main strength of SAXS is in the validation and refinement of higher-resolution structural data, e.g., from Macromolecular X-ray crystallography (MX), Nuclear Magnetic Resonance (NMR) and cryogenic electron microscopy (cryoEM) (Kachala, Valentini, & Svergun, 2015; Kikhney & Svergun, 2015; Lipfert & Doniach, 2007; Mertens & Svergun, 2017). Additionally, SAXS is a powerful method for the characterization of conformationally flexible and disordered systems which are inaccessible to higher resolution structural techniques (Bernadó, Mylonas, Petoukhov, Blackledge, & Svergun, 2007; Hammel et al., 2021; Panjkovich & Svergun, 2016; Tria, Mertens, Kachala, & Svergun, 2015).

In a solution SAXS experiment a liquid sample typically under flow is exposed to a collimated X-ray beam. The scattered X-rays are normally isotropic and registered on a photon sensitive detector producing a 2D image. For biological SAXS experiments it is standard to azimuthally average the detector image and generate a 1D scattering intensity curve I(s), ($s = 4\pi\sin\theta/\lambda$, where 2θ is the angle between the scattered and the incident radiation, and λ is the X-ray wavelength). To improve data quality, replicate exposures are collected where possible and averaged. The average scattering of the background is then subtracted from that of the sample using a separate measurement of the solvent/buffer collected under identical instrumental conditions. This yields a difference scattering curve that describes the scattering of the particles of interest, with the contribution of the solvent, sample holder and parasitic scattering effects removed (Svergun et al., 2013).

The background-subtracted 1D scattering curve can be used to derive important structural characteristics of the particles in solution, such as the

radius of gyration (Rg) (Guinier, 1939), maximum particle dimension (D_{max}), pair distance distribution function P(r) (Glatter, 1977; Hansen, 2012; Svergun, 1992), excluded particle/Porod volume (V_p) (Porod, 1951) and molecular weight (MW) (Hajizadeh, Franke, Jeffries, & Svergun, 2018; Mylonas & Svergun, 2007; Orthaber, Bergmann, & Glatter, 2000; Rambo & Tainer, 2013). The average electronic structure of the particles or form-factor can also be recovered from a high-quality curve. In addition, interparticle interactions described by the so-called Structure-factor can be characterized. Advanced modeling procedures have been developed to extend beyond the simple parametric characterization of SAXS data including low-resolution ab initio methods for three-dimensional shape reconstruction (Franke & Svergun, 2009; Grant, 2018; Svergun, 1999; Svergun, Petoukhov, & Koch, 2001). Hybrid modeling approaches incorporating high-resolution models from other techniques such as X-ray crystallography are also available and are routinely applied to obtain atomistic representations (Panjkovich & Svergun, 2016; Petoukhov & Svergun, 2005). Polydisperse samples exhibiting either particle conformational sampling/disorder (e.g., protein flexibility) or distributions in particle size (e.g., assembly/oligomerization processes) can also be characterized through advanced modeling and fitting procedures that treat the sample as a mixture of defined components with specific volume fractions (Konarev & Svergun, 2018; Tria et al., 2015).

A number of dedicated software tools and packages have been developed for the processing, analysis and modeling of SAXS data, both in the soft matter and biological fields (Alves, Pedersen, & Oliveira, 2017; Breßler, Kohlbrecher, & Thünemann, 2015; Bressler, Pauw, & Thünemann, 2015; Brookes, Vachette, Rocco, & Pérez, 2016; Hopkins, Gillilan, & Skou, 2017; Liu, Morris, Hexemer, Grandison, & Zwart, 2012; Perkins et al., 2016; Schneidman-Duhovny, Hammel, & Sali, 2010). The most comprehensive and popular software suite used for biological SAXS is the cross-platform ATSAS package (Franke et al., 2017; Konarev, Petoukhov, Volkov, & Svergun, 2006; Manalastas-Cantos et al., 2021; Petoukhov et al., 2012; Petoukhov, Konarev, Kikhney, & Svergun, 2007), developed with a focus on ease of use and application. ATSAS is under continuous development by the team at EMBL (Hamburg), benefiting from the developers also supporting experiments at the P12 synchrotron SAXS beamline (PETRA III, DESY, Hamburg). The individual ATSAS programs provide an extensive set of utilities that can be applied to monodisperse and also polydisperse biological systems. The suite has been freely available for academic use since 2003 and has been downloaded more than 100,000 times by over

18,000 unique users. An online interface to the most popular programs is also available and submission to date is approaching 50,000 jobs annually by almost 1000 unique users. Below we will introduce the key underlying components of the engine that drives the modeling procedures of the mature ATSAS suite and also describe a number of the important data analysis tools required by the SAXS user community.

2. Calculation and simulation of scattering data

2.1 Computation of SAXS intensities and model fitting

A key feature of ATSAS is the use of spherical harmonics for the rapid computation of scattering amplitudes and isotropic SAXS intensities from high-resolution atomic structures, and in fitting predicted scattering curves to experim*ent*al SAXS data (Svergun, Barberato, & Koch, 1995). The workhorse program for these operations is CRYSOL. CRYSOL provides scattering intensities in several convenient forms, the first in units proportional to electrons-squared, and the second on an absolute scale per unit concentration [$I_{abs}(s)$ (cm^{-1})/c ($mg.mL^{-1}$)]. In addition, CRYSOL can incorporate wavelength dependent anomalous effects for the analysis of anomalous X-ray scattering experiments. The general execution of CRYSOL via the GUI or command line has been described in detail in our previous work and is also available via the online manual repository (https://www.embl-hamburg.de/biosaxs/manuals/), here we will outline how to execute CRYSOL for anomalous scattering calculations using the common command line interface.

The options for running an ATSAS tool such as CRYSOL can be obtained by adding the -h "help" flag to the executable:

```
$> crysol -h
Usage in the batch mode:
crysol [<Inp_File1>]...[<Inp_FileK>] [<Dat_File>]
[-param1 <param1>]...[-paramN <paramN>]
where the arguments Inp_File can be either the
names of a PDB format file (extension .pdb or .ent)
or CRYSOL save file (extension .sav). Wildcards
(e.g. m?.sav or 1a*.ent or *.pdb) are permitted, in
which case the program makes a loop over these files
Dat_File is experimental scattering data file name.
When no experimental file name is given the intensity
is predicted with default parameters (no fitting)
```

```
The following program options can be given as
keys with parameters
-lm Maximum order of harmonics (default 15)
-fb Order of Fibonacci grid (default 17)
-sm Maximum scattering vector (default 0.5)
-ns # points in computed curve (default 51)
-un Angular units
-dns Solvent density (default 0.334 e/A**3)
-dro Contrast of hydration shell (0.03 e/A**3)
-kp Whether to keep output files other than
  (i) in fitting mode: *.log and *.fit
  (ii) in prediction mode: *.log, *.alm and *.int
  (default No, i.e. other files will be deleted)
-cst Subtract constant (default no)
-eh Account for explicit hydrogens
-err=N Write old-style fit file without
  information on experimental errors
-old Read a PDB file with an old atom naming
  (pre 2008?)
-nmr # of NMR conformer
-cid Chain identifier(s) (0: all chains)
-el Absorbing element (for anomalous SAXS)
-en Energy, absorption edge (for ASAXS)
-ff File with anomalous form factors (ASAXS)
-p Prefix for the output file names
-v Print version information and exit
-h Print this help and exit
Examples:
crysol 6lyz.pdb -lm 20
Calculate scattering intensity from the PDB file
6lyz.pdb with Lmax = 20 and without fitting
crysol mod*.pdb exp.dat -un 2
Process PDB files with the names beginning with
"mod" and fit experimental data exp.dat with
scattering vector given in inverse nanometres.
crysol *.sav lyzexp.dat
Restore the scattering intensity from all
sav files in the current directory and fit
the experimental data in the file lyzexp.dat
```

From the detailed list of options one can see the relevant flags for anomalous scattering (i.e., -el, -en and -ff), alongside the standard options for model (*.pdb and *.cif format) and scattering data file (*.dat) input. Anomalous effects occur when the wavelength of the incident radiation causes electronic transitions of a particular element to occur, i.e., the radiation wavelength is close the absorption edge. When this condition is met the incident radiation is partially absorbed and electrons are excited to higher-energy states. This leads to a reduction in scattering intensity (James, 1950). From this anomalous effect distance information can be extracted and has been successfully used in both X-ray crystallography (Hendrickson, 2014) and SAXS (Miake-Lye, Doniach, & Hodgson, 1983; Stuhrmann & Notbohm, 1981), however, the change observed in the SAXS signal is typically very low and requires an optimized instrumental set-up. Below the command line arguments required for the prediction of anomalous scattering intensities from the metalloprotein parvalbumin (PDB: 4cpv), with calcium atoms substituted by terbium (absorption edge 7514 eV) is shown. The calculation is conducted at the absorption edge, and a two additional points far from it (3000 eV and 11,000 ev):

```
$> crysol -lm 50 -ns 1024 -el tb -en 7514 parvalbumin_Tb.pdb
$> crysol -lm 50 -ns 1024 -el tb -en 3000 parvalbumin_Tb.pdb
$> crysol -lm 50 -ns 1024 -el tb -en 11000 parvalbumin_Tb.pdb
```

The generation of a difference curve (Fig. 1), obtained by subtracting the high energy scattering profile (i.e., 11,000 eV) from that at the absorption edge (7514 eV), provides structural information on the distances between the two Tb metal centers in the protein.

A detailed description of using ASAXS for biomolecular studies is presented by Gruzinov et al. (2021), with a detailed discussion on the effect of anomalous scattering on SAXS parameters such the radius of gyration and the conversion via Fourier Transformation to real-space.

The previous example demonstrates the basic usage of CRYSOL for the computation of theoretical scattering curves, with the expansion of the program for ASAXS computation also shown. In the following example we demonstrate the fitting procedure of CRYSOL for the validation of a structural model of a protein from X-ray crystallography with the SAXS curve obtained in solution.

Here we compare the predicted scattering profile from the high-resolution crystallographic structure of Bovine Serum Albumin (BSA) 4F5S to the experimental data (bsa_sec.dat) collected at the P12 BIOSAXS beamline (PETRA III, Hamburg). The data for this sample was obtained

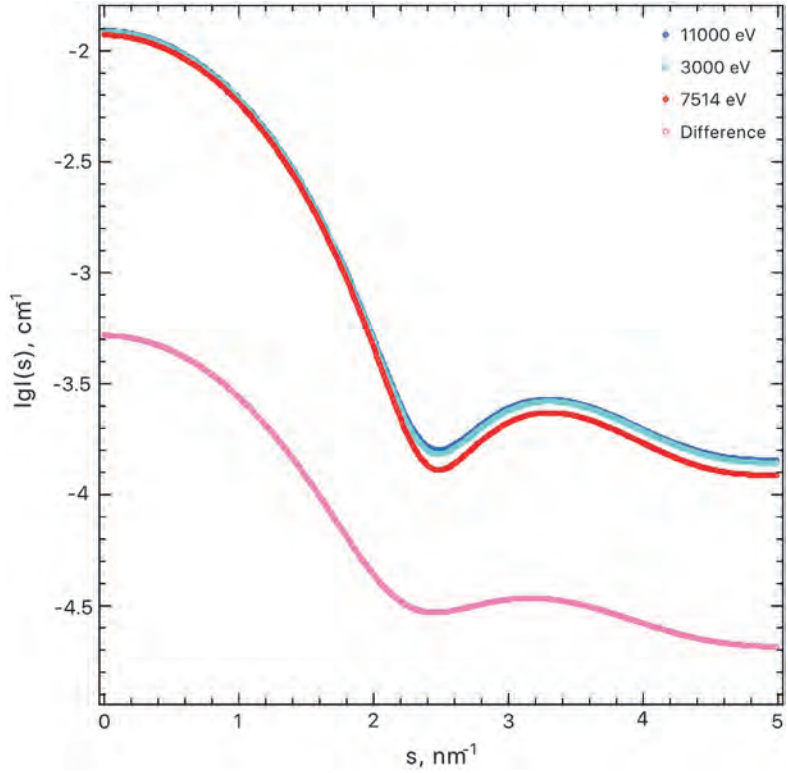

Fig. 1 ASAXS analysis of Parvalbumin. Predicted scattering curves from PDB ID. 4CPV at several energies, with both Calcium centers replaced by Terbium. The difference curve (pink circles) results from subtraction of the curve computed at the absorption edge of Terbium (7514 eV) from that computed far from the edge at 11000 eV.

by in-situ Size-exclusion chromatography (SEC) coupled SAXS, producing purified monomeric BSA during the measurement:

```
$> crysol -lm 50 -ns 1024 -cst 4F5S.pdb bsa_sec.dat
```

This command instructs CRYSOL to compute the predicted intensities from the 4F5S pdb structure file, using 50 spherical harmonics and 1024 points. The fit of the predicted profile to the experimental data is optimized through adjustment of the average atomic radius (adjusting the excluded volume used in the calculation, i.e., the volume occupied by the protein in a continuum of solvent) and the contrast of the thin water layer at the surface of the protein (the hydration layer of ordered water molecules, distance from the bulk solvent). An additional constant is applied to account for mismatch between the measured buffer and sample solutions (i.e., the buffer composition of both is not exactly the same) (Fig. 2).

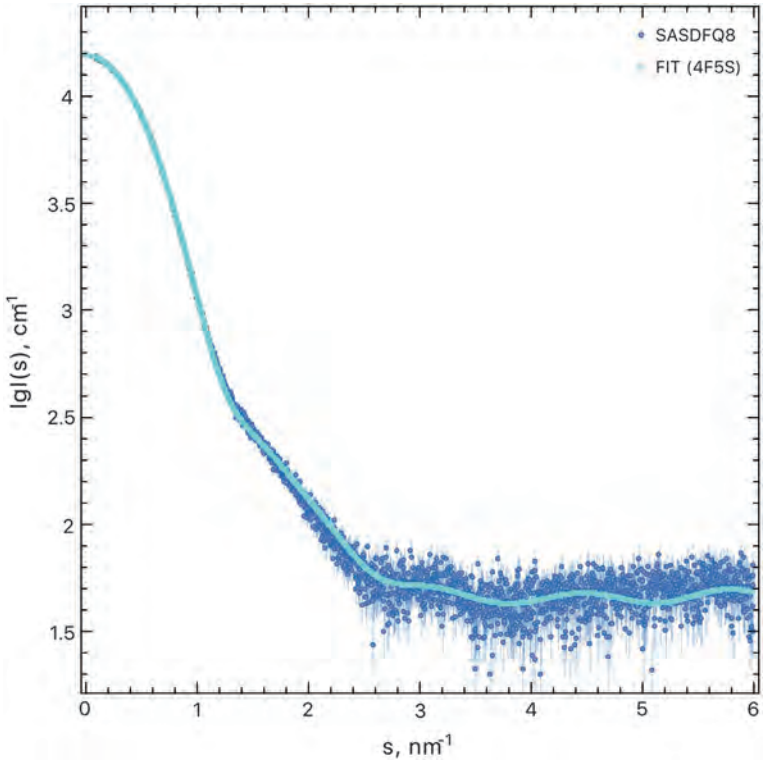

Fig. 2 CRYSOL fitting procedure for the comparison of the crystallographic structure of BSA with solution SAXS data. The experimental data was collected at the P12 SAXS beamline (EMBL, PETRA III, Hamburg) and is available at https://www.sasbdb.org/ (accession number: SASDFQ8).

The key component of the computation is the speed of the spherical harmonics approximation, operating in 3D polar (i.e., spherical) coordinates rather than the historical Debye point-to-point approach. The sum of these orthogonal angular functions $Y_{lm}(\Omega)$, the multipole expansion, provides an expression (Eq. 1) for the scattering density, $\rho(r)$ of a particle in solution that describes both gross structural features (lower order terms) and finer detailed structure (higher order terms) (Stuhrmann, 1970):

$$\rho(r) = \sum_{l=0}^{L} \sum_{m=-l}^{l} \rho_{lm}(r) Y_{lm}(\omega) \tag{1}$$

Where $(r,\omega) = (r, \theta, \varphi)$ are spherical coordinates and

$$\rho_{lm}(r) = \int_{\omega} \rho(r) Y^{*}_{lm}(\omega) d\omega \tag{2}$$

are the radial functions. The scattering amplitude is represented in reciprocal space as:

$$A(\mathbf{s}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} A_{lm}(s) Y_{lm}(\Omega) \tag{3}$$

Where the momentum transfer $s = 4\pi \sin(\theta/2)/\lambda$ with θ the scattering angle and λ the wavelength, Ω is the solid angle in reciprocal space, and the truncation value L defines the resolution. The partial amplitudes $A_{lm}(s)$ are calculated as:

$$A_{lm}(s) = 4\pi i^{l} \sum_{j=1}^{N} f_j(s) j_l(sr_j) Y^{*}_{lm}(\omega_j) \tag{4}$$

where the sum runs over all elements with coordinates $rj = (rj, \omega j) = (rj, \theta j, \varphi j)$ and $fj(s)$ are the corresponding form factors, and $j_l(sr)$ are spherical Bessel (Stuhrmann, 1970) functions.

The orthogonal properties of the spherical harmonics allows for the following expression of the spherically averaged scattering intensity $I(s)$:

$$I(s) = 2\pi^2 \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm}(s)|^2 \tag{5}$$

Eq. (5) provides a convenient mechanism for the rapid computation of particle scattering intensity as a sum of independent contributions (e.g., from radially separated scattering centers with known atomic form factors). In addition, the multipole expansion eliminates cross-terms from the calculation. Eqs. (1)–(5) also facilitate the reverse procedure, going from a scattering pattern to low resolution structure.

2.2 Simulation of 2D SAXS intensities

ATSAS also provides tools for the simulation of SAXS intensities as recorded on a 2D detector. The image simulation program, IMSIM, provides SAXS patterns in 2D (frames/images), as would be registered from the scattering of a sample on a modern detector (Franke, Hajizadeh, & Svergun, 2020). These images are invaluable for the testing of data reduction and analysis/modeling

approaches and also as an aid to experimental design and beamline optimization.

To generate a simulated 2D SAXS image of a protein molecule with IMSIM, the first step is to compute the 1D intensity profile on an absolute scale (cm^{-1}) using CRYSOL for example (see previous section). The simulation is based upon this curve and uses a statistical simulation approach such that the intensities and error estimates reflect that obtained from experimental data. A number of options are available as listed below:

```
$> imsim -h
Usage: imsim [OPTIONS] <ABSFILE>
Known Arguments:
  ABSFILE                  Calculated scattering intensity in I
[cm^-1] / conc[mg/ml]
Known Options:
      --seed=<INT>          Set the seed for the random number generator
      --background=<FILE>  Calculated background intensity in I
[cm^-1] / conc[mg/ml] (default: built-in flat background)
      --detector=<NAME>    Detector name, defines pixel size and
mask dimensions (default: Pilatus1M)
      --detector-distance=<m>Sample-detector-distance  in  meter
(default: 1.0)
      --flux=<N>           Number of incoming virtual photons per
second (default: 10^12)
  -p, --scattering-probabilitProbability that a virtual photons
scatters (default: 1e-6)
      --exptime=<T>        Exposure time in seconds (default: 1)
  -c, --concentration=<C>  Factor proportional to concentration
of sample (default: 0)
      --wavelength=<m>     X-ray wavelength in meter (default: 0.1e-9)
      --axis               Do  not  simulate  but  write  axis  for
radial averaging
      --mask               Do not simulate but write a centered
detector mask for radial averaging
      --mask-offset-x=<X>  Mask  offset  in  horizontal  direction
(default: 0)
      --mask-offset-y=<Y>  Mask  offset  in  vertical  direction
(default: 0)
      --header             Do not simulate but write header infor-
mation suitable for IM2DAT
```

```
  -o, --output=<FILE>       Output file name (default: basename of
the input file, extension depends on other options)
  -f, --format=<FORMAT>     Output format, one of: TIFF, EDF, CBF,
MSK, ...
      --atomic              Use atomic scattering without hydration
shell only
  -h, --help                Print usage information and exit
  -v, --version             Print version information and exit
```

Here we show an example for the simulation of the 2D SAXS image of BSA (PDB ID. 4F5S):

1. Computation of theoretical 1D SAXS intensities from PDB file, with a maximum s value of 1.0 $Å^{-1}$, 5001 points and 50 spherical harmonics.
   ```
   $> crysol -sm 1.0 -ns 5001 -lm 50 4F5S
   ```
2. Simulation of the 2D image from the 1D SAXS curve on absolute scale (4F5S00.abs), using the sim.sh helper script. The key options used such as X-ray wavelength, flux and exposure time, as well as desired output such as an axis calibration file and mask file are shown.
   ```
   $> bash sim.sh 4F5S00.abs
   --- * Helper script sim.sh * --
   #! /bin/bash
   # Shorthand for common options that are the same for all calls to
   IMSIM.
   PROFILE="--detector=pilatus6m      --detector-distance=3.1      --
   wavelength=0.124e-9 --flux=1e12 --scattering-probability=1e-3"
   SAMPLE_EXPTIME=01.00
   BUFFER_EXPTIME=10.00

   abs=${1}
   prefix=$(basename ${abs} .abs)

   # MASK and AXIS do not depend on random are the same everywhere.
   mask=${prefix}_mask.msk
   imsim ${PROFILE} ${abs} --mask --mask-offset-y=-1100 -o ${mask}
   axis=${prefix}_axis.dat
   imsim ${PROFILE} ${abs} --axis -o ${axis}

   for CONCENTRATION in 0.5 2.0 8.0 ; do
     # Prepare output file names.
     buffer=${prefix}_buffer_c${CONCENTRATION}_e${BUFFER_EXPTIME}
     sample=${prefix}_sample_c${CONCENTRATION}_e${SAMPLE_EXPTIME}
   ```

```
  sub=${prefix}_sub_c${CONCENTRATION}
  norm=${prefix}_norm_c${CONCENTRATION}

  # Do all simulations in parallel.
  imsim ${PROFILE} --exptime=${BUFFER_EXPTIME} -c 0.0 ${abs} -o
${buffer}.tiff &
  imsim ${PROFILE} --exptime=${BUFFER_EXPTIME} -c 0.0 ${abs} -o
${buffer}.txt --header &
  imsim ${PROFILE} --exptime=${SAMPLE_EXPTIME} -c ${CONCENTRATION}
${abs} -o ${sample}.tiff &
  imsim ${PROFILE} --exptime=${SAMPLE_EXPTIME} -c ${CONCENTRATION}
${abs} -o ${sample}.txt --header
  wait

  # Obtain the actual position of the incident beam.
  x=$(grep beam-center-x ${sample}.txt | cut -d: -f2 | tr -d [:
space:])
  y=$(grep beam-center-y ${sample}.txt | cut -d: -f2 | tr -d [:
space:])

  # Radially average two-dimensional images to one-dimensional
scattering data.
  im2dat -x ${x} -y ${y} -beamstop-mask ${mask} --axis-data ${axis}
${buffer}.tiff ${buffer}.txt -o ${buffer}.dat
  im2dat -x ${x} -y ${y} -beamstop-mask ${mask} --axis-data ${axis}
${sample}.tiff ${sample}.txt -o ${sample}.dat

  # Subtract background from sample and normalize by concentration.
  datop sub ${sample}.dat ${buffer}.dat -o ${sub}.dat
  datop div ${sub}.dat ${CONCENTRATION} -o ${norm}.dat

  # (optionally) clean up the resulting images to save some space.
  # rm -f *.tiff
done
```

Support for performing operations at the level of 2D images has been recently introduced into ATSAS, including the image operations program IMOP, which follows a similar strategy to that supplied by the data operations program DATOP for 1D scattering data (Franke et al., 2017). IMOP supports operations such as addition and subtraction of images, as well as AND, OR and XOR operations for binary masks. IMOP can be employed for cross-validation of data reduction operations.

Once a simulated 2D image has been generated, data reduction programs such as IM2DAT can be used to azimuthally average the image and provide 1D SAXS curves for standard analysis, with appropriate errors and statistics based on the desired input. The data reduction protocol is outlined in the following section.

3. Primary data processing

As mentioned in Section 1 it is typical in biological SAXS data analysis for the acquired intensities to be represented as 1D profiles/curves. A number of data reduction and processing steps are performed to obtain the desired SAXS profile, and the ATSAS package provides both an automated and semi-automated/manual workflow for these procedures. The standard data reduction steps are outlined in Fig. 3.

The initial step post 2D image acquisition is the azimuthal/radial integration of the scattering intensities. This is performed using the IM2DAT program (Konarev et al., 2006), where a calibration file based on an appropriate standard such as Silver Behenate ($AgC_{22}H_{43}O_2$, $d_{00} = 5.84$ nm) and a mask file (masking out of detector regions not contributing to the intensities, e.g., detector edges, beamstop shadow and space between modules) are used to reduce the dimensionality of the data. Error estimates are based on Poisson photon counting statistics, and outliers in each integration ring section detected through a transformation step approximating a normal distribution (Anscombe, 1948; Iglewicz & Hoaglin, 1993). Sub-pixel analysis (i.e., pixel-splitting) has not been implemented as we aim to avoid the introduction of correlations between neighboring intensity estimates. The resulting 1D SAXS profile/curve produced is then passed to programs specifically operating on 1D data sets with care taken to ensure reliable error propagation.

3.1 Primary data processing of 1D SAXS profiles/curves

Correct data handling is essential for the delivery of reliable 1D SAXS data sets (Kikhney, Borges, Molodenskiy, Jeffries, & Svergun, 2020). Cross-validation at various stages during primary processing helps to ensure that error propagation is correctly conducted.

3.1.1 Basic operations on 2D and 1D scattering data

Tools have been introduced to allow data reduction and processing to be conducted on both 2D and 1D data. For 2D images the program IMOP

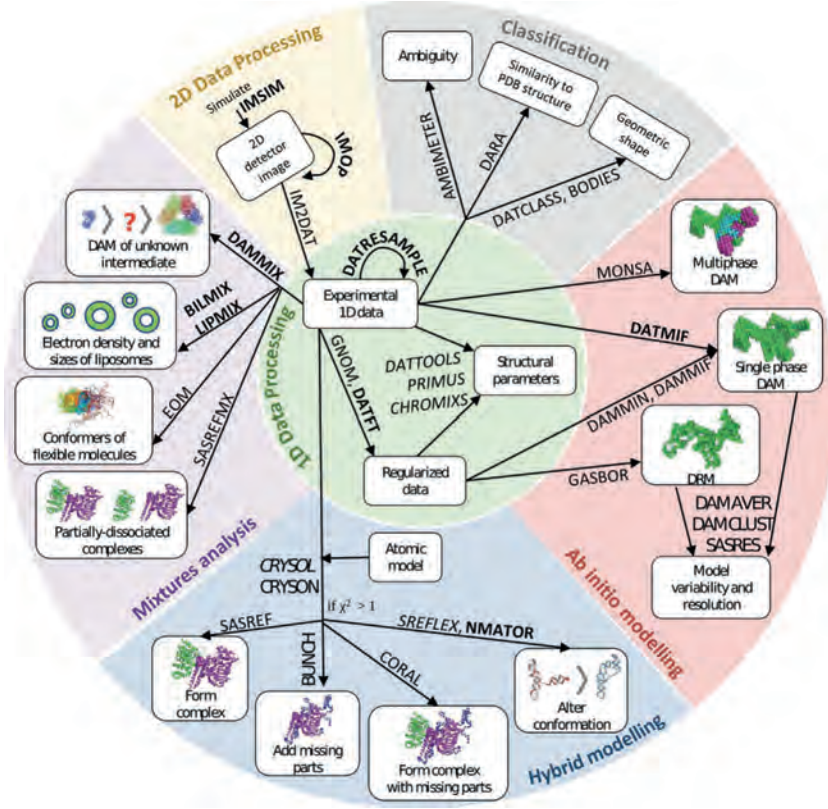

Fig. 3 The ATSAS suite of programs and procedures, showing both reduction/analysis/modeling procedures/steps and specific program modules. Also shown are the potential use cases involving monodisperse/polydisperse systems (DAM: dummy-atom model; DRM: dummy-residue model). *Reused with permission from IUCRJ Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C., Jeffries, C.M., Svergun, D.I., Franke, D. (2021). ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis.* Journal of Applied Crystallography *54:343–355. https://doi.org/10.1107/S1600576720013412.*

(image operations) can be used for basic mathematical operations such as image addition and subtraction, and also provides functionality for the masking of pixels. As described above, the reduction of 2D images to 1D intensity profiles is facilitated by the program IM2DAT. Processing of 1D scattering profiles is conducted using the command line based DATOOLS

suite, which has been designed for easy implementation into automated high-throughput small-angle scattering data processing procedures, such as the SASFLOW pipeline (Franke, Kikhney, & Svergun, 2012) employed at the P12 (EMBL, Hamburg) beamline. The DATOOLS (Franke et al., 2017) operations include functions for addition/subtraction/multiplication/division (DATOP, data operations) and for the extraction of SAXS invariants and other associated parameters as will be outlined in Sections 3.1.1–3.1.3. Additional helper tools are also available for the merging of data sets (DATMERGE, for data collected across different angular ranges, or the combining of low and high concentration data sets collected on the same samples), averaging (DATAVER, 1D data set averaging), and also for the statistical comparison of datasets (DATCMP, data comparison) to identify outliers. The components of the DATTOOLS suite have been integrated into the PRIMUS (Primary Data Analysis) GUI, such that those not familiar with command line usage can work visually with 1D SAXS data to perform all of the necessary data manipulations (Fig. 4).

The two primary SAXS invariant parameters extracted from SAXS data are the radius of gyration, R_g, and the forward scattering intensity (or intensity at zero angle), I_0. These parameters describe the overall average particle size and scattering mass, respectively. An example of using the command line DATRG protocol is shown below:

(i) Accessing the usage information/help:

```
$> datrg -h
Usage: datrg [OPTIONS] <DATAFILE(S)>
Radius of Gyration (Rg) from Guinier approximation
Known Arguments:
 DATAFILE           Data file
Known Options:
 -t, --type=<TYPE>   particle shape: GLOBULAR, ROD or FLAT (default
GLOBULAR)
  --first=<N>        index of the first point of the Guinier region
  --last=<N>         index of the last point of the Guinier region
 -h, --help          Print usage information and exit
 -v, --version       Print version information and exit
Mandatory arguments to long options are mandatory for short options
too.
Report bugs to <atsas@embl-hamburg.de>.
```

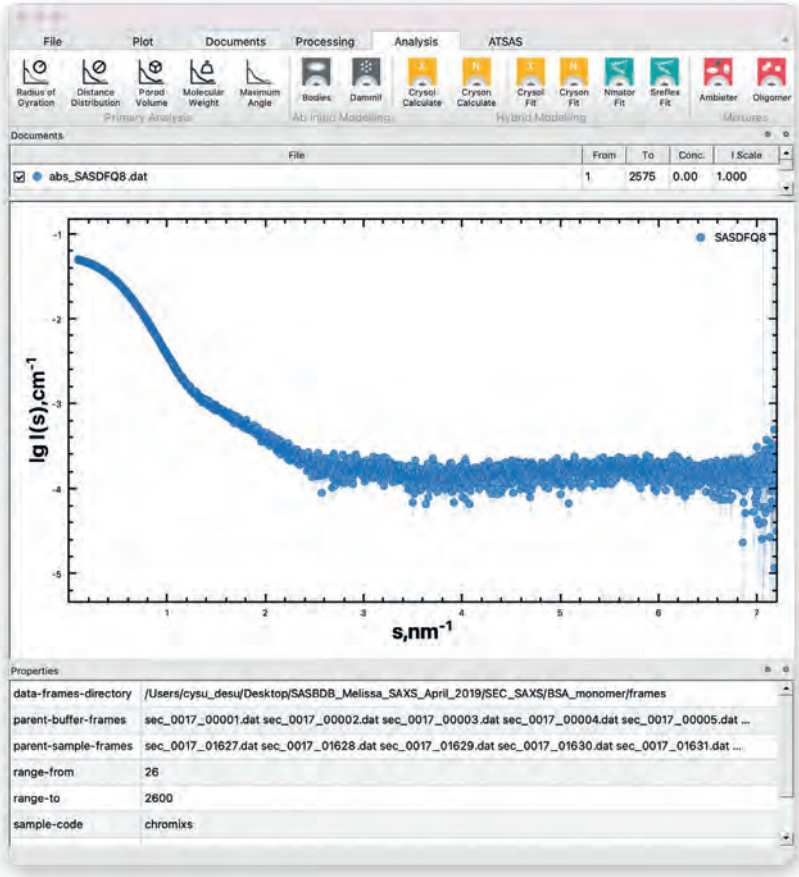

Fig. 4 PRIMUS GUI interface to ATSAS DATTOOLS suite of operations for data manipulation and analysis. Top panel shows the general GUI providing data manipulation and analysis tools alongside a data visualization canvas. Lower panels show the plot tools corresponding to the data visualization canvas (A), Guinier analysis/Radius of gyration analysis and corresponding residuals (B and E), real-space distance distribution (Indirect Fourier transformation) analysis and corresponding residuals (C and F), and the dimensionless Kratky plot representation (D).

(ii) Estimation of Rg and I0 parameters from an experimental data set (SASDFQ8.dat):

```
$> datrg --type globular --first 14 --last 135 abs_SASDFQ8.dat
 2.76215 0.188875E-02 0.499999E-01 0.202396E-04 0.123471E-01
```

The results are provided in columnar format as R_g, error R_g, I_0, error I_0 and fidelity.

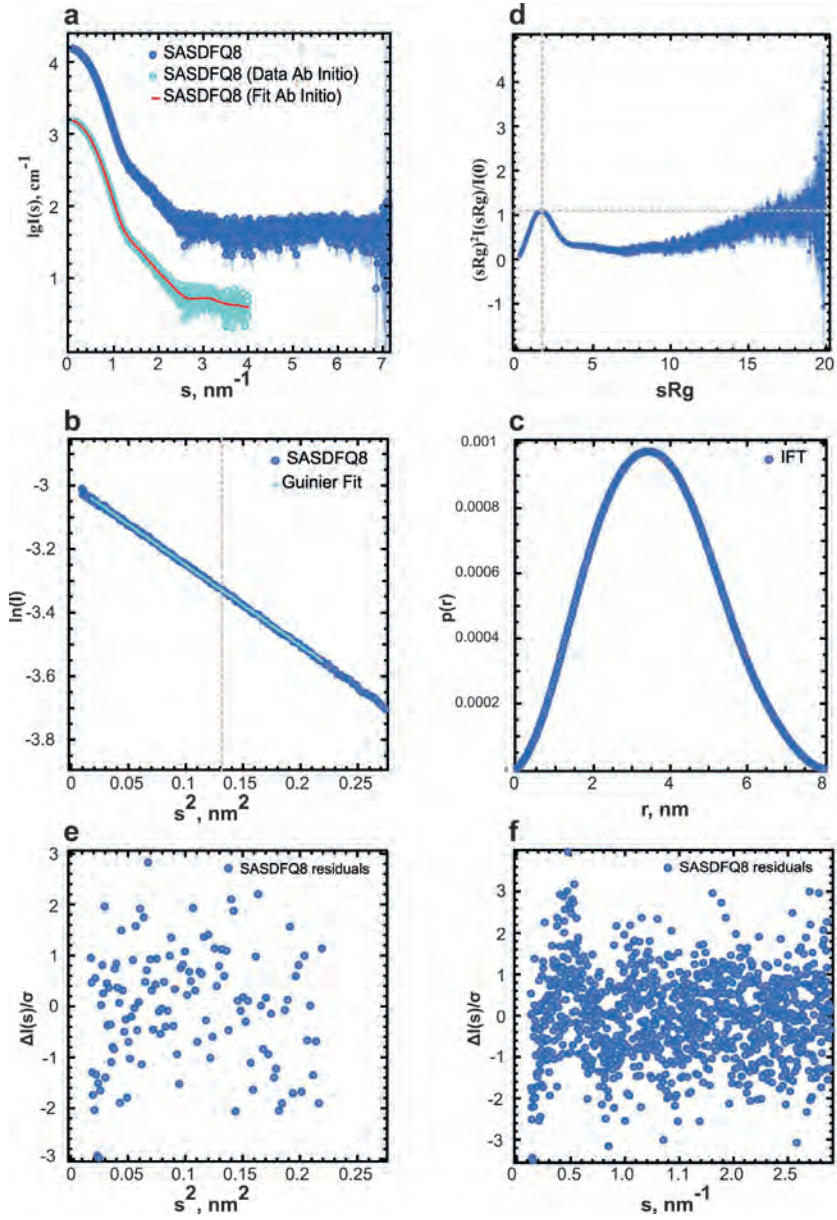
a
SASDFQ8
SASDFQ8 (Data Ab Initio)
SASDFQ8 (Fit Ab Initio)
lgI(s), cm-1
s, nm-1
d
(sRg)2I(sRg)/I(0)
sRg
b
SASDFQ8
Guinier Fit
ln(I)
s2, nm2
c
IFT
p(r)
r, nm
e
SASDFQ8 residuals
Δ(I(s)/σ
s2, nm2
f
SASDFQ8 residuals
Δ(I(s)/σ
s, nm-1

Fig. 4—cont'd

3.1.2 Indirect Fourier transformation and the real-space distance distribution

The SAXS intensities recorded in reciprocal space, I(s), are related to the scattering particles in real-space, *p(r)*, by Fourier transformation:

$$I(s) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin(sr)}{sr} \mathrm{d}r \tag{6}$$

The correlation of pointwise distances within the particles, weighted by the product of the excess scattering-length densities of each pair of points, can be described by the pair distance distribution function, *p(r)* (also referred to as the PDDF). The *p(r)* function is derived from *I(s)* by the inverse transform:

$$p(r) = \frac{r}{2\pi^2} \int_0^{\infty} sI(s) \sin(sr) \mathrm{d}s \tag{7}$$

As the measured intensities, *I(s)* are known only as a series of discrete points across a finite range ($0 < s < \infty$), *p(r)* is not typically determined directly from Eq. (2). The Indirect Fourier Transformation (IFT) approach, introduced by Glatter (1977), facilitates *p(r)* determination via the fitting of the experimental data with approximated functions prior to transformation. The ATSAS package expands upon this functionality in the program GNOM (Semenyuk & Svergun, 1991), where *p(r)* is parameterized by a set of analytical functions, and a regularization procedure is used to balance the fit to the experimental data and the smoothness of the real-space distribution. An example of the procedure applied to BSA is shown in Fig. 4. From the IFT approach, R_g and I_0 are also determined, providing an independent method to validate these parameters determined from Guinier analysis. In addition, IFT provides an estimate of the maximum particle dimension, D_{max}, and the hydrated particle volume (also referred to as the Porod volume, V_p).

Although, as previously stated, *p(r)* is not typically determined directly from Eq. (2), functionality in ATSAS is provided to perform a direct Fourier transform of I(s) with the help of additional constraints in the program DATFT. If I_0 and R_g can be reliably estimated via Guinier analysis, the truncated low angle data ($\lim \to 0$) can be extrapolated to 0. Extrapolation of the data at higher angles ($\lim \to \infty$) is approximated as $I(s) = s - n$ (where n is dependent on the particle form, with $n = 4$ for globular particles and $n = 2$ for fully flexible chain systems). These extrapolations are required to reduce

artificial oscillations in the *p(r)* function caused by the finite data range (Harris, 1978). DATFT can operate on scattering data for which no preprocessing has been conducted, however, optimal results are found for data having low-noise and uniformly distributed points.

Several helper tools have been introduced to validate experimentally determined p(r) functions (PDDFFIT) and to facilitate easy extraction of the distribution and the regularized/fit curve for visualization/plotting (OUT2POFR, OUT2FIT).

3.1.3 Estimation of molecular weight of proteins from SAXS data

I_0 is one of the most reliable invariant parameters extracted from experimental SAXS data. This parameter is directly related to the volume of scatterers (i.e., electrons for X-rays) and thus the molecular weight of the particles in solution. If the SAXS data has been placed on an absolute scale, for example relative to pure water or another reference sample (Mylonas & Svergun, 2007; Orthaber et al., 2000), and the concentration, c of the macromolecule of interest is accurately known (ideally within 1–2% of the true value), then the molecular weight of the macromolecule can in principle be determined from I_0/c:

$$MW = \left(N_{av}{}^{*} I_0/c\right)/\left(\mathrm{PSV}^{*}\Delta P\right)^2, \tag{8}$$

Where, $N_{av} = 6.022 \times 10^{23}$, PSV is the partial specific volume (cm^3/g) and ΔP is the scattering contrast (cm^{-2}) between the macromolecule and the solvent.

The determined molecular weight is a valuable parameter for sample characterization. Aggregation, oligomeric state, complex association and dissociation can all be explored/verified by the molecular weight values obtained from the experiment. Several ATSAS tools are available for absolute calibration (DATABSOLUTE) relative to pure water and molecular weight estimation (DATABSMW). The following shows an example of molecular weight estimation of BSA (MW = 66,000 Da) from DATABSMW:

```
$> databsmw -h
Usage: databsmw [OPTIONS] [FILE(S)]
Molecular weight from absolute scaling.
Known Arguments:
 FILE       Data file on absolute scale, divided by the concentra-
tion [mg/ml]
```

```
Known Options:
  --psv=<X>        Partial Specific Volume (default: 0.7425)
  --contrast=<X>   Contrast in units of 10^10 (default: 2.8086)
  --i0=<X>         I(0) of the input file
 -h, --help        Print usage information and exit
 -v, --version     Print version information and exit
Mandatory arguments to long options are mandatory for short options
too.
Report bugs to <atsas@embl-hamburg.de>.
$> databsmw --psv 0.74 --contrast 2.90 --i0 0.051 abs_SASDFQ8.dat
66690
```

An accurate concentration estimation is not always available for every protein sample, and in size-exclusion chromatography coupled (SEC-SAXS) data collected without UV/RI-detection (UV = ultraviolet wavelength radiation, RI = refractive index) concentration is unavailable. In such cases concentration independent approaches can be used to obtain molecular weight estimates. These approaches can additionally be used to check molecular weight estimates derived from I_0/c. ATSAS provides several tools for concentration independent MW estimation, using scattering invariants independent of data scale. DATPOROD and SAXSMoW calculate the Porod invariant (Q_p) to provide an estimate of the volume (V_p) of the sample, where MW is estimated as the quotient of V_p and the protein partial specific volume (Porod, 1951). The final MW estimate is obtained following several empirical corrections (Petoukhov et al., 2012; Piiadov, Ares de Araújo, Oliveira Neto, Craievich, & Polikarpov, 2019) implemented within these programs. Although technically not an invariant, the parameter V_c, volume of correlation, correlates with MW with both protein and RNA, and MW estimation based on V_c (Rambo & Tainer, 2013) has been implemented in the ATSAS tool DATVC. DATCLASS, a machine-learning method based on protein structures in the PDB, combines classification of shape with D_{max} and MW estimation from scattering data to provide another concentration independent approach to MW estimation (Franke, Jeffries, & Svergun, 2018). All of the concentration independent approaches to MW estimation outlined can be combined to as input for the DATMW tool (Hajizadeh et al., 2018), a Bayesian approach to generate a consensus MW value with a probability estimate, enabling users to examine the interval of possible values.

4. Structural modeling from SAXS data

Modeling procedures based on SAXS data have evolved tremendously in the last decade, moving beyond simple parameter extraction and analytical from-factor fitting to three-dimensional modeling and the deconvolution of complex mixtures. The first generally applicable shape modeling procedures developed were those based on ab initio (i.e., from first principles) modeling. In these approaches the average structure of particles in solution is determined directly from the measured SAXS intensities through comparison to a shape consisting of spheres or dummy atoms. Hybrid methods incorporate high-resolution structural data (e.g., models deposited at the PDB) and other biophysical and biochemical constraint information, using SAXS data to drive iterative refinement to optimize a fit between the computed scattering of the models and the experimental data. Approaches have also been developed for the analysis of flexible or unstructured systems (e.g., intrinsically disordered proteins/regions, IDPs/IDRs) and for the characterization and deconvolution of polydisperse systems including oligomeric mixtures and association/dissociation processes. Membrane proteins and lipidic/detergent systems are also becoming a significant target of solution scattering studies, and as such ATSAS provides several tools applicable to the modeling of these multi-contrast macromolecular complexes. Below a number of modeling procedures used for the characterization of biological macromolecules incorporating SAXS data are outlined.

4.1 Ab initio (shape) methods

In the absence of structural information, models of macromolecules can be reconstructed directly from SAXS intensities and constrained by penalties designed to limit the search space to biophysically plausible solutions. ATSAS provides a number of tools for the generation of fits to geometrical shapes (BODIES) (Konarev, Volkov, Sokolova, Koch, & Svergun, 2003), reconstruction of bead/dummy-atom shape models (DAMMIN (Svergun, 1999)/DAMMIF (Franke & Svergun, 2009)/MONSA (Svergun & Nierhaus, 2000)/DATMIF (Manalastas-Cantos et al., 2021)), and reconstruction of dummy-residue models of proteins by means of a globbic approximation of the average scattering of amino-acids (GASBOR) (Svergun et al., 2001). These methods have been in use for more than a decade and have proven invaluable to many SAXS based data analyses.

4.1.1 Direct modeling from experimental data

Several of the ATSAS ab initio bead modeling applications (DAMMIN, DAMMIF, GASBOR) require a regularized scattering profile as input, such as that produced during the fitting routine of the indirect Fourier transformation procedure (e.g., from GNOM). This indirect approach is historical and has been replaced by a more direct modeling procedure implemented in the program DATMIF (Dummy Atom Minimization Fast). DATMIF generates bead models through a direct fitting of the scattering data and thus incorporates the experimental error estimates. In the older ATSAS ab initio programs the penalties used to guide the solution toward "biological-like" models include those penalizing discontinuity of beads, compatibility with R_g, implicit looseness/compactness, and explicit anisometry (when symmetry has been requested). In DATMIF, the only modeling penalty applied is the Akaike information criterion (AIC) for minimizing the number of beads. This results in DATMIF minimizing the volume of the final model and results in compact protein-like model generation (Fig. 5). An example of the command line execution of DATMIF for the protein BSA is shown below:

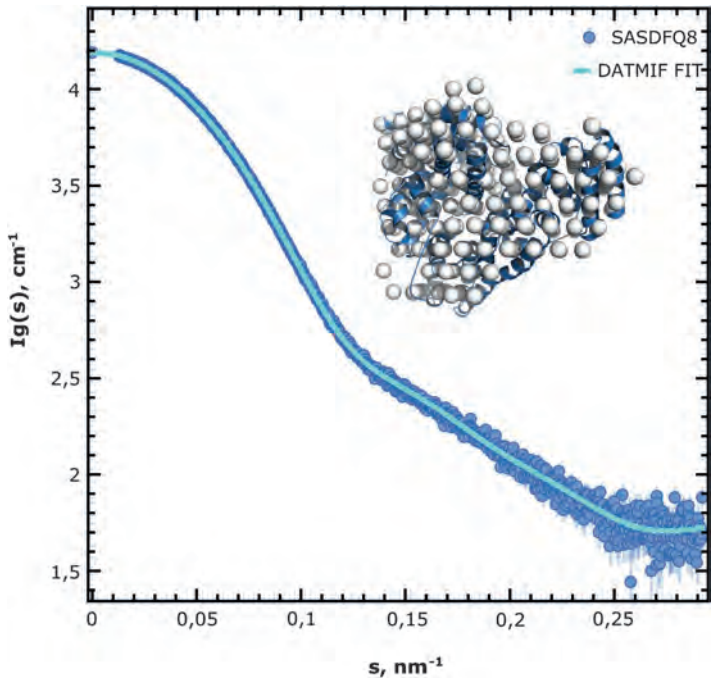

Fig. 5 DATMIF ab initio model of BSA. 1D scattering data (blue), overlaid with the corresponding DATMIF fit (cyan). Inset shows the DATMIF bead model superimposed on the high resolution crystallographic model of BSA (PDB ID 4F5S). Model generated from data deposited at www.sasbdb.org for entry SASDFQ8.

```
$> datmif -h
Usage: datmif [OPTIONS] <DATFILE>
Ab-initio shape determination of experimental small-angle scatter-
ing data
Known Arguments:
 DATFILE              Experimental data to model
Known Options:
 -h, --help            Print usage information and exit
 -v, --version         Print version information and exit
 -q, --quiet           Reduce verbosity level
  --seed=<INT>         Set the seed for the random number generator
 -u, --unit=<UNIT>     angular unit, inv. ANGSTROM, inv. NANOMETRE
or UNKNOWN (default: unknown)
  --constant=<VALUE>   constant to subtract (default: automatic)
  --first=<VALUE>      first data point (default: determined by
autorg)
  --last=<VALUE>       last data point (default: s * Rg <= 8)
 -p, --prefix=<PREFIX> output PREFIX prepended to any output
filename (default: datmif )
  --dmax=<D>           Estimate of the maximum overall distance? [A]
(required)
  --particle-bead-radius=dummy atom bead radius [A] (default: 4A)
 -s, --symmetry=<PXY>  particle symmetry
  --max-step=<ARG>     maximum number of temperature steps
(default: 500)
  --max-iter=<N>       maximum number of phase changes per temp.
step (default: #beads/10)
  --max-succ=<N>       maximum number of successful phase changes
before temp. is reduced (default: max-iter/10)
  --min-succ=<N>       minimum number of successful phase changes
before termination (default: max-iter/1000)
  --temp-schedule=<T> temperature schedule factor (0.0-1.0)
(default: 0.95)
Mandatory arguments to long options are mandatory for short options
too.
$>datmif --first 14 --last 1024 --unit nanometer --dmax 80 --prefix
datmif_bsa SASDFQ8.dat
```

4.1.2 Multiphase modeling of membrane proteins in solution

Membrane proteins are commonly solubilized in solution using lipids, lipidic scaffolds (e.g., nanodiscs, SMALp, saliproNPs) and detergents (Denisov, Grinkova, Lazarides, & Sligar, 2004; Glover et al., 2001; Seddon, Curnow, & Booth, 2004; Tanford & Reynolds, 1976). This is required for stability in aqueous solvents as a significant amount of surface area is hydrophobic and must be shielded from the polar environment to retain structure and function. Membrane mimics provide the mixed polar/apolar environment required to stabilize integral membrane proteins, however, differences in electron density of the polar and apolar regions of the mimics and also that of the protein yield a multiple scattering contrast system. This multi-contrast situation can be difficult to model unambiguously using the aforementioned ab initio methods, however, the program MONSA (Svergun, 1999; Svergun & Nierhaus, 2000), facilitates ab initio modeling of systems consisting of multiple phases with distinct contrasts, and can be used to model, e.g., detergent-solubilized transmembrane proteins.

MONSA can reconstruct models of membrane proteins solubilized in membrane mimics, and solves the ill-posed problem through the addition of information about the chemistry of the system. Knowledge based constraints on the search volume used by MONSA are provided through extensions made to the tool BODIES, with the three phase initial MONSA search volume (1. Protein, 2. Hydrophobic lipid/detergent tails, 3. Hydrophilic lipid/detergent heads) constrained by user specified hydrophobic and hydrophilic thickness of the non-protein phases on the basis of the chemistry of the lipid/detergent used (Fig. 6). The phase assignment at the boundaries of each pair of phases, including that between the protein core and the solvent, is variable, facilitating better optimization during modeling in MONSA.

An example showing how bodies can be used to set up the input search volume for MONSA is shown below (Protein in DDM detergent):

```
$> bodies -h
Usage: bodies [OPTIONS] <FILE>
Fit or predict the small angle scattering data or generate a Dummy
Atom Model (DAM) of various body types. Predict and DAM mode are
not available through command-line arguments but through dialog
mode only (start without arguments).
Known Arguments:
 FILE      Data file to fit
```

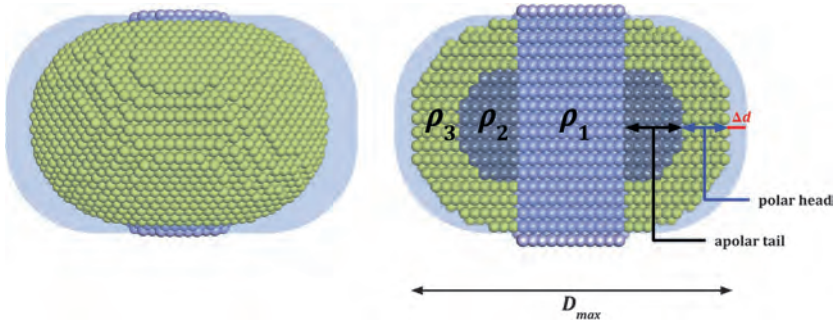

Fig. 6 Search volume created by BODIES for multiphase ab initio modeling with MONSA. The protein phase ρ_1 (blue), hydrophobic/apolar tail phase ρ_2 (gray) and hydrophilic/polar head phase ρ_3 (yellow) are shown within an ellipsoidal total search volume of maximum diameter D_{max}. Phase thicknesses and a variable boundary region $\triangle$d are specified by the user on the command line. A view of the surface and the internal structure of the phases are shown on the left and right, respectively.

```
Known Options:
 -p, --prefix=<PREFIX>  output prefix for .log and .fit
  --body=<BODY>   Name of the BODY type
  --rg=<X>        Expected Rg (default: estimated from data)
  --i0=<X>        Expected I0 (default: estimated from data)
  --volume=<X>    Expected volume (default: estimated from data)
  --first=<N>     First point of the Guinier range (default: 1)
  --last=<N>      Last point of the fit range (default: ~4 Shannon
channels)
  --seed=<INT>    Set the seed for the random number generator
 -q, --quiet      Reduce verbosity level
 -h, --help       Print usage information and exit
 -v, --version    Print version information and exit
$> bodies
Operation mode? Select one of: (f) fit, (p) predict, (d) dam, (v)
search volume (default: fit) ..................................... : d
Body type? Select one of: (1) ellipsoid, (2) rotation-ellipsoid, (3)
cylinder, (4) elliptic-cylinder, (5) hollow-cylinder, (6)
parallelepiped, (7) hollow-sphere, (8) dumbbell, (9) liposome, (10)
memebrane protein (default: ellipsoid) ........................... :
10
```

```
Value of parameter 'rmemb' [Angstrom]? ..... :
25
Value of parameter 'rtail' [Angstrom]? ..... :
26
Value of parameter 'rhead' [Angstrom]? ..... :
44
Value of parameter 'delta' [Angstrom]? ..... : 5
Value of parameter 'zcorona' [Angstrom]? ..... : 0
Dummy atom model symmetry? Select one of: (1) P1 (default: P1)
..... : p1
Number of dummy atoms? (default: 2000) ..... :
5000
DAM output file name? ..... :
vol.pdb
```

The membrane protein specific options are:

R_{memb}: approximate radius of the protein phase from the origin.

R_{tail}: radial distance of the hydrophobic tail phase (distance from the end of protein phase)

R_{head}: radial distance of the hydrophilic head phase (distance from the end of the tail phase)

Delta: distance of variable boundaries between phases

Z_{corona}: offset of membrane from center of protein phase/origin (i.e. height of membrane relative to membrane protein)

4.2 Hybrid modeling procedures

Ab initio modeling methods are useful in the absence of prior structural information, but if such information is available more advanced modeling procedures can be applied. Hybrid modeling methods are provided in ATSAS for the modeling of rigid-body and flexible systems, where partial models of structured domains and/or additional distance/orientation/biophysical information is included in the model refinement procedure.

4.2.1 Rigid-body modeling

Rigid-body methods incorporate structural information in the form of high-resolution structures from the PDB or high-scoring homology models (e.g., models obtained from sophisticated algorithms such as those used by AlphaFold) (Jumper et al., 2021) into the computational refinement

procedures. In the ATSAS programs BUNCH, SASREF and CORAL (Petoukhov et al., 2012; Petoukhov & Svergun, 2005), the rigid bodies introduced by the user are considered independent building blocks of known structure, and a solution is sought such that an optimal spatial arrangement of the building blocks is obtained that best fits the scattering data. The target function constructed for optimization additionally includes terms describing geometric criteria such as connectivity, symmetry and anisometry and penalties for steric clashes between backbone/main-chain atoms. SASREF provides functionality for simple rigid-body modeling, allowing refinement against many X-ray and neutron scattering data sets collected under a number of contrast conditions. BUNCH facilitates multidomain structure refinement with the addition of missing terminal and linker regions, modeled as dummy residues. CORAL (COmplexes with RAndom Loops) combines the methods of SASREF and BUNCH and facilitates a generally applicable hybrid-modeling approach, where rigid bodies for domains and missing region modeling can be combined with additional distance constraint information. CORAL allows users to specify fixed components and those free-to-move, and can impose symmetry as specified. An example of the input required for the program CORAL for the rigid body refinement of the hexameric transcriptional activator protein Rov C (Regulator Of Virulence interconnected with the Csr system, SASBDB ID SASDHP5) (Fig. 7.) is shown below.

(i) Configuration file (coral.con) listing the rigid bodies (PDB ID 6XZ5, split into four domains) and regions of missing sequence (NTER/LINK with number of residues) to be a modeled:

```
NTER 7
6XZ5_8to41.pdb
LINK 6
6XZ5_49to93.pdb
LINK 10
6XZ5_104to140.pdb
LINK 5
6XZ5_146to170.pdb
LINK 15
6XZ5_186to252.pdb
```

(ii) Options for CORAL (regular user mode), P32 symmetry and rigid body positions fixed:

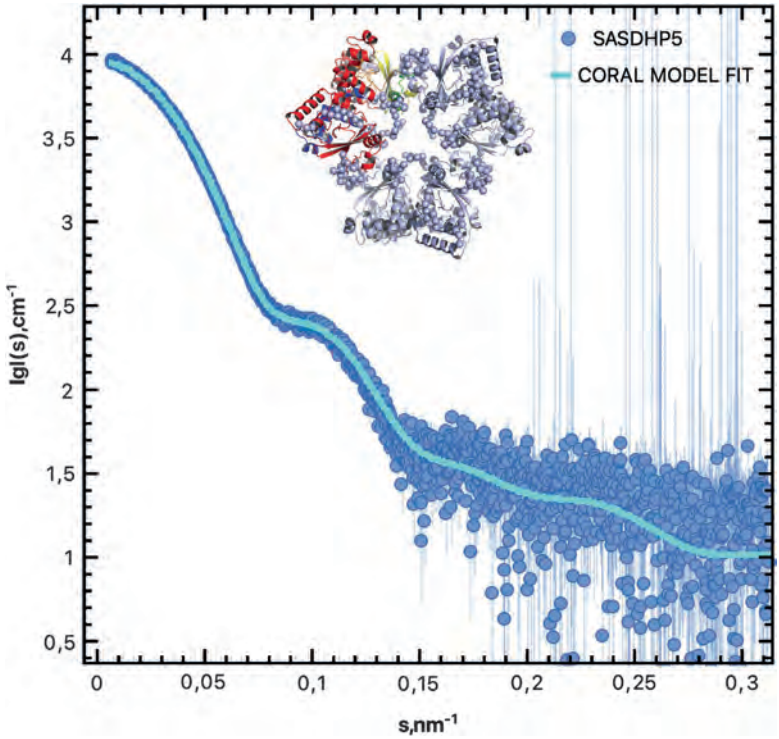

Fig. 7 CORAL rigid body refinement of RovC. 1D scattering data (blue), overlaid with the corresponding CORAL fit (cyan). Inset shows the hexameric model generated with P32 symmetry (Red, orange, yellow and blue cartoon shows the input rigid bodies for the refinement based on PDB ID 6XZ5 and modeled linker residues shown as beads). Model generated from data deposited at www.sasbdb.org for entry SASDHP5.

```
$> coral
*** Rigid body modelling of complexes with multidomain ***
*** proteins against multiple scattering curves.  ***
*** Symmetry/contacts can be taken into account.  ***
*** Linkers are represented as interconnected dummy ***
*** residues chains attached to the appropriate  ***
*** Ca atoms in rigid domains. Models with overlaps ***
*** and disconnected ones are penalized.    ***
*** Please reference: M.V. Petoukhov et al. (2012) ***
*** J. Appl. Cryst. 45, 342-350.      ***
*** Version 1.1 (r13837)        ***
*** Copyright (c) ATSAS Team       ***
```

```
*** EMBL, Hamburg Outstation, 2010-2015    ***
======= CORAL started on   29-Mar-2022 10:41:26
Computation mode (User or Expert) ..... <  User >: u
Log file name ......................... <  .log >: RovC
Project identifier ..................................... : RovC
Enter project description ............. : RovC hexamer refinement
P32 symmetry
File name with objects info ........... <  .con >: coral.con
Coordinates of the 1-st subunit evaluated from ......... :
6XZ5_8to41.pdb
 296 atoms read, center at -30.90 26.67 -19.06
Fix the subunit at original position? [
Y / N ] ............................... <  No >: y
Coordinates of the 2-nd subunit evaluated from ......... :
6XZ5_49to93.pdb
 373 atoms read, center at -22.26 30.15 -13.19
Fix the subunit at original position? [
Y / N ] ............................... <  No >: y
Coordinates of the 3-rd subunit evaluated from ......... :
6XZ5_104to140.pdb
 305 atoms read, center at -10.53 26.03 -15.61
Fix the subunit at original position? [
Y / N ] ............................... <  No >: y
Coordinates of the 4-th subunit evaluated from ......... :
6XZ5_146to170.pdb
 125 atoms read, center at -6.84 21.32 -1.02
Fix the subunit at original position? [
Y / N ] ............................... <  No >: y
Coordinates of the 5-th subunit evaluated from ......... :
6XZ5_186to252.pdb
 563 atoms read, center at -40.84 8.40 -8.72
Fix the subunit at original position? [
Y / N ] ............................... <  No >: y
Number of backbone atoms generated ..................... : 252
Symmetry: P1...19 or Pn2 (n=1,..,12) ... <  P1 >: P32
Input total number of scattering curves <  1 >: 1
Enter file name, 1-st experimental data <  .dat >: SASDHP5.dat
Fitting range in fractions of Smax ..... <  1.000 >: 1.0
...
```

4.2.2 Flexible body modeling and ensembles

In rigid-body modeling the high-resolution components are considered immutable, with no internal structural changes. However, it can be that significant structural plasticity is present for such bodies in solution, and so flexible modeling procedures can be applied. The ATSAS program SREFLEX (Panjkovich & Svergun, 2016) provides functionality for the conformational rearrangement of input high-resolution protein structures against experimental scattering data. During refinement the protein is morphed along the available Cartesian normal modes in an optimization procedure, with the goal of finding a set of conformations that best fits the data. Normal Modes Analysis (NMA) has been used to characterize interdomain motions in a number of protein systems (Alexandrov et al., 2005; Dobbins, Lesk, & Sternberg, 2008; Krebs et al., 2002; Tama & Sanejouand, 2001; Tobi & Bahar, 2005; Wako & Endo, 2017), using connected harmonic motions around a starting equilibrium to sample the available conformational space (Goldstein, 1980). An extended normal modes approach is that provided by NMATOR (Normal Modes Analysis in Torsion angle space) (Manalastas-Cantos et al., 2021), where the coordinated motions are described in torsion angle space. NMATOR facilitates NMA based refinement and conformational sampling of both protein and non-protein biomolecular systems, morphing structure through rotations of dihedral angles. NMATOR enables normal mode computation in torsion angle space, structure refinement along normal modes to optimize a fit to experimental SAXS data, and the generation of ensembles of conformations which can be used as input for the modeling and characterization of flexible systems. The latter two methods are also available in Cartesian space using SREFLEX.

Ensemble based approaches allow for the characterization of flexible and unstructured systems, where a single model usually provides an inadequate description of the solution behavior. ATSAS provides the Ensemble Optimization Method (EOM) (Bernadó et al., 2007; Tria et al., 2015) tool for the analysis of such systems, which is particularly suited to Intrinsically Disordered Proteins (IDPs) and macromolecules with Intrinsically Disordered Regions (IDRs).

The scattering intensity of an ensemble of conformationally polydisperse particles can be approximated as:

$$I(s) = \sum_k v_k I_k(s) \tag{9}$$

where $I_k(s)$ is the scattering intensity from the k^{th} component with volume fraction v_k (Konarev et al., 2006). For flexible protein systems an indirect

approach is used to approximate the contribution of the many components to the overall scattering. In EOM a large pool of possible conformations is generated (based on the sequence and any available structures corresponding to the sequence—PDB files) to approximate the large conformational space. The scattering of each member of the pool is computed and a genetic algorithm is then run to obtain the subset of conformations that best fit the experimental SAXS data, i.e., minimizing the discrepancy χ^2:

$$\chi^2 = \frac{1}{K-1}\sum_{j=1}^{K}\left[\frac{\mu I(s_j) - I\ \exp(s_j)}{\sigma(s_j)}\right]^2 \tag{10}$$

where $I_{exp}(s)$ is the experimental scattering, K is the number of experimental points, $\sigma_{(sj)}$ are standard deviations and μ is a scaling factor.

The conformers selected by the genetic algorithm provide the low-resolution sample space used for the generation of statistical distributions of parameters. These distributions (R_g, D_{max}, end-to-end distance and conformer volume) describe the degree of flexibility and compactness of the sample measured, relative to the corresponding distributions of unrestricted conformational freedom derived from the initial pool (Fig. 8). As the size-distributions are essentially probability density functions and describe the degree of entropy of the system, $H_b(S)$ (Shannon & Weaver, 1949), a measure of the states available (i.e., flexibility of the system) can be derived (see Tria et al., 2015 for details) (Tria et al., 2015) such that:

$$R_{\text{flex}} = -H_b(S)$$

Where a broad Gaussian-like distribution of parameters corresponds to high-uncertainty, i.e., $H_b(S)$ tends to -1 (similar to that expected for a pool of random conformers), thus the metric of flexibility R_{flex} approaches 1 (fully flexible). A narrow size distribution provides low uncertainty, with $H_b(S)$ tending to 0, and thus R_{flex} approaches 0 (low flexibility). The information content, or entropy, can therefore be used as a quantitative measure of flexibility with the metric R_{flex}. For convenience, R_{flex} can be reported as a percentage in the range 0 to 100%, with $R_{flex} = 100\%$ representing maximum flexibility. Using R_{flex}, the uncertainty of the selected ensemble distribution can be compared to that of the reference pool.

4.2.3 Quasi-atomistic modeling of membranes, detergents and membrane proteins

Membrane proteins solubilized in solution using lipids, lipidic scaffolds and detergents are readily measured in solution by SAXS. A significant issue to

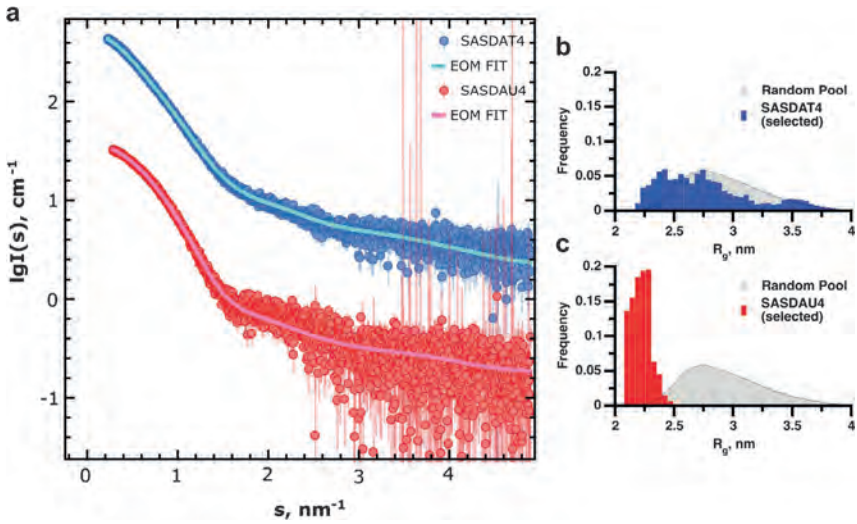

Fig. 8 EOM analysis of Urokinase Plasminogen Activator Receptor, uPAR wildtype (blue) and disulfide-bridged mutant (red). (A) Fits of the best ensembles selected by EOM are shown on the left for both constructs (data offset vertically for better visualization). R_g size-distributions for the selected ensembles and the initial random pool are shown on the right for wildtype (B) and mutant (C). The metric for flexibility is $R_{flex} = 81\%$ (wildtype) and 47% (mutant), relative to 85% for the pool. A broad distribution of compact and extended structures is selected for the wildtype case, whereas a narrow distribution of compact structures is selected for the mutant—indicating a stable compact structure in solution with little flexibility. Ensembles generated from data deposited at www.sasbdb.org for entries SASDAT4 (wildtype), and SASDAU4 (mutant).

consider for the measurement of membrane proteins in detergents is the separation of the scattered intensities of the protein-detergent-complex (PDC) from that of the free detergent in the sample. It is often not practical to match the detergent background of the buffer solution to that of the sample, as the equilibrium of free and bound detergent in the sample will lead to a different concentration of free detergent compared to the buffer containing the detergent (and no interacting PDC). Although extensive dialysis can be used to attempt to obtain a matched detergent concentration, the most robust approach is to measure the sample with in-line HPLC. Provided that the elution of the PDC and free detergent micelles is well separated, this facilitates extraction of an optimally buffer subtracted PDC SAXS curve.

The complex contrast situation mentioned in Section 4.1.2 for ab initio modeling also provides difficulties for hybrid modeling approaches, however, this can be overcome computationally again if appropriate chemical information is provided. In this case, structures of the detergent/lipid

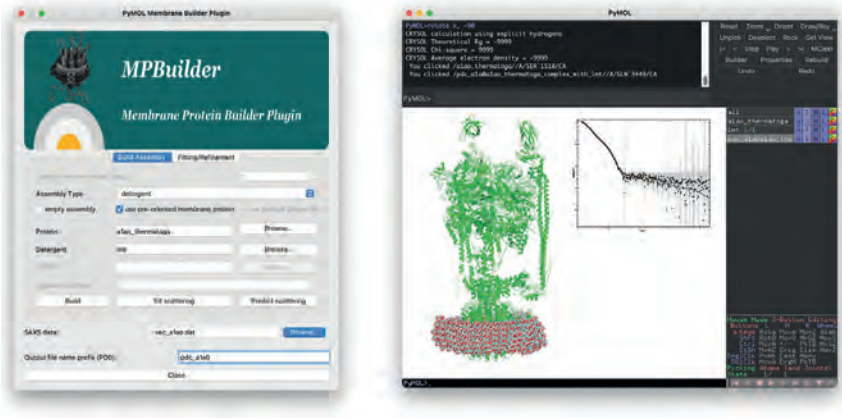

Fig. 9 MPBuilder interface to PYMOL. BUILD operation for the A1AO-ATP synthase from Therm. Thermophilus and a detergent corona of dodecylmaltoside (DDM) shown. Model generated from data deposited at www.sasbdb.org for entry SASDKK4.

components (PDB/CIF file formats) are used to accurately calculate the scattering intensities of the solubilizing components in solution. In the ATSAS package the PYMOL plugin MPBuilder (Molodenskiy, Svergun, & Mertens, 2021) provides a convenient interface for biological users to construct candidate membrane protein models from atomic structures in lipidic/detergent environments, and to calculate the fit of these models to experimental SAXS data (Fig. 9).

MPBuilder requires an input membrane protein structure file (PDB/CIF format), and for best results it is recommended that an alignment tool such as that provided by the PPM3.0 program at the OPM website (https://opm.phar.umich.edu/ppm_server) (Lomize, Pogozheva, Joo, Mosberg, & Lomize, 2012) be used to optimally position the transmembrane region with respect to the origin (center of the membrane or detergent corona to be modeled). A database of structures of commonly used detergents, lipids and protein scaffolds (nanodisc scaffold protein belts and saposin proteins) is provided in CIF format. The tool can be used in build or refinement mode, with the latter conducting a grid search of available parameters (e.g., number of detergent molecules—polar angles and density of detergent monomers, number of scaffold monomers, geometrical offset of scaffold monomers relative to the origin, and angle of scaffold monomers with respect to the membrane plane—e.g., XY).

In addition to MPBuilder, ATSAS provides a tool for the construction of liposomes and mixed micellar systems (e.g., bicelles). The program ELLLIP

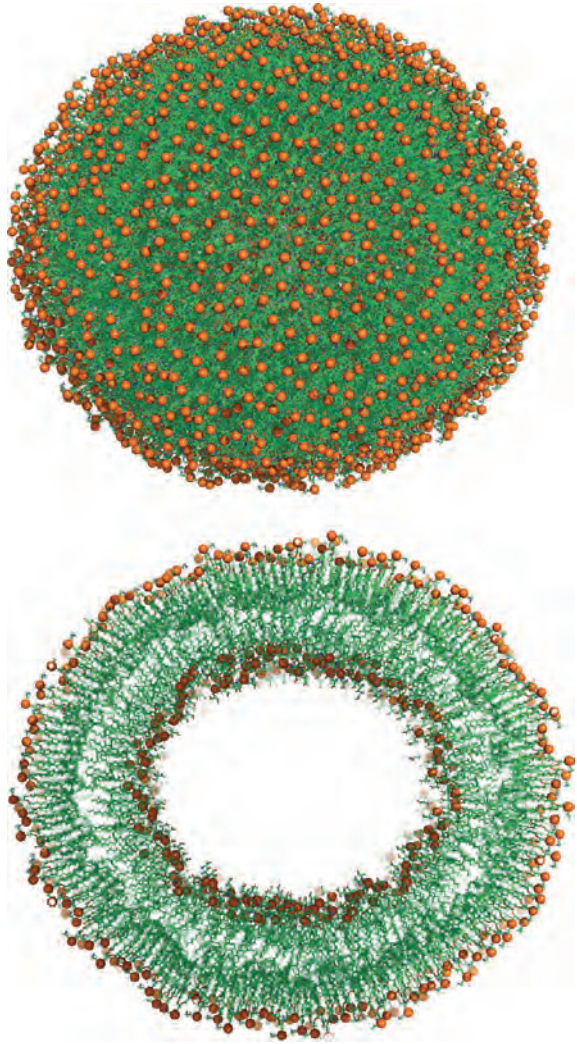

Fig. 10 Ellipsoidal POPE:POPS liposome generated with ELLLIP. Model constructed using the POPE (outer leaflet) and POPS (inner leaflet) PDB files extracted from the MPBuilder CIF database. Phosphorus atoms of the headgroups are shown as orange spheres.

(ELlipsoidal Lipid) (Petukhov et al., 2020) facilitates construction of ellipsoidal bilayered membrane particles (Fig. 10.), using an angular grid to optimize the positions of user defined input lipid molecules (PDB format). Further randomization can be conducted as required to reduce ideality and more closely resemble biophysically relevant lipid structures in solution. The

model membrane particles constructed can then be checked for agreement with scattering data with CRYSOL.

4.3 Polydisperse systems

In monodisperse ideal solutions, it is assumed that all particles present are identical. This facilitates application of the ab initio and hybrid modeling procedures outlined above, with an additional assumption that inter-particle interactions are minimal (i.e., sample is significantly dilute). These assumptions cannot be made when the sample measured is polydisperse. Analysis of such samples requires approaches that account for the different particle sizes/structures present and their corresponding volume fractions. For a mixture, similarly to that expressed in Eq. (9), the scattering profile can be represented as the volume-weighted sum of the scattering intensities of the individual components:

$$I(s) = \sum_{k=1}^{N} v_k I_k(s) \tag{11}$$

Where V_k and I_k correspond to the volume fraction and scattering profile (form-factor) of each distinct scattering component N. ATSAS provides the program OLIGOMER (Konarev et al., 2003) to analyze multiple sets of polydisperse scattering data and obtain an optimized solution that best describes the volume fractions of components present. Such an analysis can be conducted using pre-computed intensities (form-factors) from components with known three dimensional structure (i.e., PDB files), or individual scattering profiles previously measured, for each component. Association/dissociation processes and evolving systems are well-suited to this type of analysis.

4.3.1 Ab initio methods applied to polydisperse solutions

The ab initio methods described in Section 4.1 aim to reconstruct molecular shape information directly from scattering intensities. The success of this approach relies heavily on the assumption that samples are monodisperse and significantly dilute such that interparticle-interactions can be ignored. However, the ATSAS tool DAMMIX (Konarev & Svergun, 2018) has been developed specifically for the ab initio shape reconstruction of a transient component in an evolving polydisperse solution. In addition to the low-resolution structure the volume fraction of the intermediate is determined. From a set of k scattering curves representing time points, and

knowledge of the structure of the initial and final components, the structure/shape of the unknown intermediate component and its volume fraction are determined from Eq. (4) and the following relationship:

$$V_{start}(k) + V_{intermediate}(k) + V_{final}(k) = 1 \tag{12}$$

Where V_{start} $V_{intermediate}$ and V_{final} are the volume fractions of the starting structure, intermediate/unknown and final state, respectively.

DAMMIX can be applied to two-component evolving systems such as oligomerization from an initial monomeric state through to the formation of a large aggregate, and systems with multiple assembly states such as viruses and nanoparticles (Fig. 11).

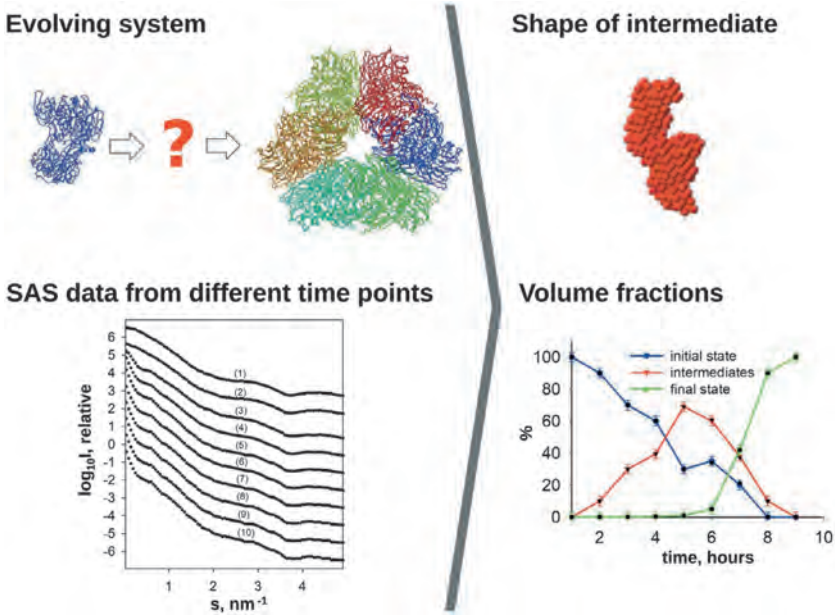

Fig. 11 DAMMIX facilitates low-resolution structure/shape reconstruction of an unknown intermediate in an evolving system. Knowledge of the initial and final state structure is required, and optimization is conducted using multiple SAXS data sets to determine the intermediate structure and the volume fractions of the initial, intermediate and final states at each time point. *Reused with permission from IUCRJ Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C., Jeffries, C.M., Svergun, D.I., Franke, D. (2021). ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis.* Journal of Applied Crystallography *54: 343–355. doi: 10.1107/S1600576720013412.*

4.3.2 Modeling polydisperse solutions of lipid vesicles

Lipid vesicles are often used as model biological membranes and scattering techniques are well suited to their analysis in solution (Heftberger et al., 2014; Konarev, Gruzinov, Mertens, & Svergun, 2021; Pabst, Rappolt, Amenitsch, & Laggner, 2000; Pencer & Hallett, 2000; Pérez & Koutsioubas, 2015). Several tools in ATSAS have been recently developed to facilitate the analysis of lipid vesicles, including methods for the reconstruction of electron density profiles across the lipid bilayer, and for the characterization of vesicle size distributions.

Scattering data from solutions of lipid vesicles can be used as input for the tools BILMIX (Konarev et al., 2020) and LIPMIX (Konarev et al., 2021) to generate profiles of the electron density, $\rho(z)$ across the membrane bilayer. An initial approximation of the SAXS intensities is made as the product of the form-factors of a thin spherical shell (FF_{TSS}) and a flat lipid bilayer (FF_{FLB}) (Kiselev, Lesieur, Kisselev, Lombardo, & Aksenov, 2002; Pencer, Krueger, Adams, & Katsaras, 2006). The former defines the size of the vesicles and the latter describes the electron density across the bilayer. The scattering profile of each vesicle k present in solution is then expressed as:

$$I_k(s) \cong \frac{1}{s^2}\left|\int FF_{\mathrm{TSS}}(s,\ r)_k D_{\mathrm{v}}(r)_k \mathrm{d}r\right|^2 \left|FF_{\mathrm{FLB}}(s)_k\right|^2 \sum_{i=1}^{M} w_i s_i^{\mathrm{FLB}}(s) \qquad (13)$$

Where $D_v(r)$ is the size distribution of the vesicles, and the last term (implemented only in LIPMIX) accounts for the presence of M distinct multilayers of occupancy w_i and inter-bilayer structure factor S_i^{FLB} (Zhang, Suter, & Nagle, 1994). The form factor of the flat lipid bilayer, $FF_{FLB}(s)$ is the Fourier transform of the electron-density profile $\rho(z)$:

$$\rho(z) = \sum_{i=1}^{2} A_i \left\{ \exp\left[\frac{-(z - z_{\mathrm{H}i})^2}{2\sigma_{\mathrm{H}i}^2}\right] + \exp\left[\frac{-(z + z_{\mathrm{H}i})^2}{2\sigma_{\mathrm{H}i}^2}\right]\right\} - \rho_r \exp\left(\frac{-z^2}{2\sigma_{\mathrm{c}}^2}\right) \qquad (14)$$

Where the Gaussian terms of width σ_{H1} and positioned at $\pm z_{H1}$ represent the polar lipid head groups, and the Gaussian term of width σ_{c} positioned at the center of the bilayer $z = 0$ represent the electron density of the apolar lipid core. Modeling of an asymmetric electron density profile is facilitated in BILMIX by the addition of two Gaussian terms of width σ_{H2} positioned at $\pm z_{H2}$.

LIPMIX and BILMIX can be used to model the size distribution of vesicles/liposomes and to determine bilayer electron-density profiles (Fig. 12).

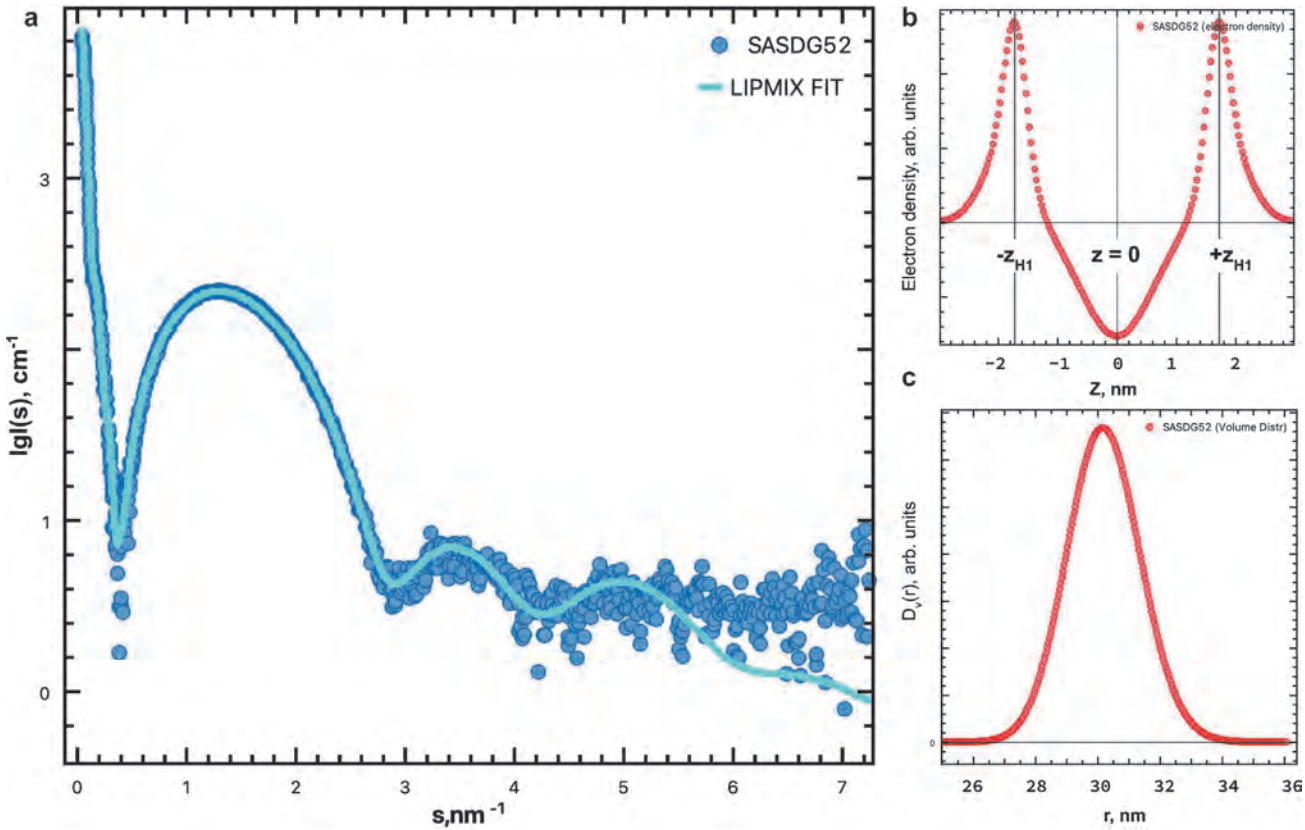

Fig. 12 LIPMIX modeling of the bilayer electron density profile and size-distribution of a 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) liposome solution, extruded through a 30 nm membrane. The model fit to the data (cyan line) is shown (A). The electron density profile (B) and volume distribution (C) are shown on the right. Data deposited at www.sasbdb.org for entry SASDG52.

This facilitates analysis of the behavior of model membranes in solution and can also be employed to examine potential interactions and disruption or morphological changes caused by additional species.

5. ATSAS summary

The ATSAS suite provides an extensive set of tools for reduction, analysis and structural modeling from experimental SAXS data. Not all available programs and tools are covered in this short review, however, potential users are encouraged to download the ATSAS suite and explore possible applications in their own SAXS analyses. ATSAS has been updated to include support of the mmCIF format, a new standard adopted by the PDB for structural information. A database initiated by the ATSAS team in collaboration with the PDB, the Small-Angle Scattering Biological DataBase (SASBDBD), has been successfully launched and adapted as a standard repository of biological scattering data (Kikhney et al., 2020). SASBDBD provides an excellent resource for new users to investigate the use of available data analysis and modeling procedures with curated experimental data. ATSAS is under constant development and will continue to be maintained at EMBL with the introduction of new tools and further improvements.

The latest ATSAS package can be installed locally (Windows/Mac OS/Linux) available at:

https://www.embl-hamburg.de/biosaxs/software.html.

Alternatively, many programs can be run on the EMBL Hamburg cluster via the ATSAS online interface:

https://www.embl-hamburg.de/biosaxs/atsas-online/

References

Alexandrov, V., Lehnert, U., Echols, N., Milburn, D., Engelman, D., & Gerstein, M. (2005). Normal modes for predicting protein motions: A comprehensive database assessment and associated web tool. *Protein Science*, *14*, 633–643. https://doi.org/10.1110/ps.04882105.

Alves, C., Pedersen, J. S., & Oliveira, C. L. P. (2017). Calculation of two-dimensional scattering patterns for oriented systems. *Journal of Applied Crystallography*, *50*, 840–850. https://doi.org/10.1107/S1600576717005179.

Anscombe, F. J. (1948). The transformation of poisson, binomial and negative-binomial data. *Biometrika*, *35*, 246–254. https://doi.org/10.2307/2332343.

Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*, *129*, 5656–5664.

Breßler, I., Kohlbrecher, J., & Thünemann, A. F. (2015). SASfit: A tool for small-angle scattering data analysis using a library of analytical expressions. *Journal of Applied Crystallography*, *48*, 1587–1598. https://doi.org/10.1107/S1600576715016544.

Bressler, I., Pauw, B. R., & Thünemann, A. F. (2015). McSAS: Software for the retrieval of model parameter distributions from scattering patterns. *Journal of Applied Crystallography*, *48*, 962–969. https://doi.org/10.1107/S1600576715007347.

Brookes, E., Vachette, P., Rocco, M., & Pérez, J. (2016). US-SOMO HPLC-SAXS module: Dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data. *Journal of Applied Crystallography*, *49*, 1827–1841. https://doi.org/10.1107/S1600576716011201.

Brosey, C. A., & Tainer, J. A. (2019). Evolving SAXS versatility: Solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Current Opinion in Structural Biology*, *58*, 197–213. https://doi.org/10.1016/j.sbi.2019.04.004.

Da Vela, S., & Svergun, D. I. (2020). Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution. *Current Research in Structural Biology*, *2*, 164–170. https://doi.org/10.1016/j.crstbi.2020.08.004.

Denisov, I. G., Grinkova, Y. V., Lazarides, A. A., & Sligar, S. G. (2004). Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. *Journal of the American Chemical Society*, *126*, 3477–3487. https://doi.org/10.1021/ja0393574.

Dobbins, S. E., Lesk, V. I., & Sternberg, M. J. E. (2008). Insights into protein flexibility: The relationship between normal modes and conformational change upon protein–protein docking. *Proceedings of the National Academy of Sciences*, *105*, 10390–10395. https://doi.org/10.1073/pnas.0802496105.

Franke, D., Hajizadeh, N. R., & Svergun, D. I. (2020). Simulation of small-angle X-ray scattering data of biological macromolecules in solution. *Journal of Applied Crystallography*, *53*, 536–539. https://doi.org/10.1107/S1600576720000527.

Franke, D., Jeffries, C. M., & Svergun, D. I. (2018). Machine learning methods for X-ray scattering data analysis from biomacromolecular solutions. *Biophysical Journal*, *114*, 2485–2492. https://doi.org/10.1016/j.bpj.2018.04.018.

Franke, D., Kikhney, A. G., & Svergun, D. I. (2012). Automated acquisition and analysis of small angle X-ray scattering data. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *689*, 52–59. https://doi.org/10.1016/j.nima.2012.06.008.

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., et al. (2017). ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*, *50*, 1212–1225. https://doi.org/10.1107/S1600576717007786.

Franke, D., & Svergun, D. I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, *42*, 342–346.

Glatter, O. (1977). A new method for the evaluation of small-angle scattering data. *Journal of Applied Crystallography*, *10*, 415–421. https://doi.org/10.1107/S0021889877013879.

Glover, K. J., Whiles, J. A., Wu, G., Yu, N., Deems, R., Struppe, J. O., et al. (2001). Structural evaluation of phospholipid Bicelles for solution-state studies of membrane-associated biomolecules. *Biophysical Journal*, *81*, 2163–2171. https://doi.org/10.1016/S0006-3495(01)75864-X.

Goldstein, H. (1980). Classical mechanics. *Addison-Wesley series in physics* (2nd ed.). Reading, Mass: Addison-Wesley Pub. Co.

Grant, T. D. (2018). Ab initio electron density determination directly from solution scattering data. *Nature Methods*, *15*, 191–193. https://doi.org/10.1038/nmeth.4581.

Gruzinov, A. Y., Schroer, M. A., Manalastas-Cantos, K., Kikhney, A. G., Hajizadeh, N. R., Schulz, F., et al. (2021). Anomalous SAXS at P12 beamline EMBL Hamburg: Instrumentation and applications. *Journal of Synchrotron Radiation*, *28*.

Guinier, A. (1939). La diffraction des rayons X aux tres petits angles; application a l'etude de phenomenes ultramicroscopiques. *Annales de Physique*, *12*, 161–237.

Hajizadeh, N. R., Franke, D., Jeffries, C. M., & Svergun, D. I. (2018). Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Scientific Reports*, *8*, 7204. https://doi.org/10.1038/s41598-018-25355-2.

Hammel, M., Rosenberg, D. J., Bierma, J., Hura, G. L., Thapar, R., Lees-Miller, S. P., et al. (2021). Visualizing functional dynamicity in the DNA-dependent protein kinase holoenzyme DNA-PK complex by integrating SAXS with cryo-EM. *Progress in Biophysics and Molecular Biology*, *163*, 74–86. https://doi.org/10.1016/j.pbiomolbio.2020.09.003.

Hansen, S. (2012). Bayesian methods in SAXS and SANS structure determination. In T. Hamelryck, K. Mardia, & J. Ferkinghoff-Borg (Eds.), *Bayesian methods in structural bioinformatics* (pp. 313–342). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-27225-7_13.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, *66*, 51–83. https://doi.org/10.1109/PROC.1978.10837.

Heftberger, P., Kollmitzer, B., Heberle, F. A., Pan, J., Rappolt, M., Amenitsch, H., et al. (2014). Global small-angle X-ray scattering data analysis for multilamellar vesicles: The evolution of the scattering density profile model. *Journal of Applied Crystallography*, *47*, 173–180. https://doi.org/10.1107/S1600576713029798.

Hendrickson, W. A. (2014). Anomalous diffraction in crystallographic phase evaluation. *Quarterly Reviews of Biophysics*, *47*, 49–93. https://doi.org/10.1017/S0033583514000018.

Hopkins, J. B., Gillilan, R. E., & Skou, S. (2017). BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*, *50*, 1545–1553. https://doi.org/10.1107/S1600576717011438.

Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers*. Milwaukee, Wis: ASQC Quality Press.

James, R. W. (1950). *The optical principles of the diffraction of X-rays*. Repr. Bell.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kachala, M., Valentini, E., & Svergun, D. I. (2015). Application of SAXS for the structural characterization of IDPs. In I. C. Felli, & R. Pierattelli (Eds.), *Intrinsically disordered proteins studied by NMR spectroscopy* (pp. 261–289). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-20164-1_8.

Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M., & Svergun, D. I. (2020). SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein Science*, *29*, 66–75. https://doi.org/10.1002/pro.3731.

Kikhney, A. G., & Svergun, D. I. (2015). A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters*, *589*, 2570–2577.

Kiselev, M. A., Lesieur, P., Kisselev, A. M., Lombardo, D., & Aksenov, V. L. (2002). Model of separated form factors for unilamellar vesicles. *Applied Physics A: Materials Science & Processing*, *74*, s1654–s1656. https://doi.org/10.1007/s003390201837.

Konarev, P. V., Gruzinov, A. Y., Mertens, H. D., & Svergun, D. I. (2021). Restoring structural parameters of lipid mixtures from small-angle X-ray scattering data. *Journal of Applied Crystallography*, *54*(Pt 1), 169–179. https://doi.org/10.1107/S1600576720015368.

Konarev, P. V., Petoukhov, M. V., Dadinova, L. A., Fedorova, N. V., Volynsky, P. E., Svergun, D. I., et al. (2020). BILMIX: A new approach to restore the size polydispersity and electron density profiles of lipid bilayers from liposomes using small-angle X-ray scattering data. *Journal of Applied Crystallography*, *53*, 236–243. https://doi.org/10.1107/S1600576719015656.

Konarev, P. V., Petoukhov, M. V., Volkov, V. V., & Svergun, D. I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, *39*, 277–286. https://doi.org/10.1107/S0021889806004699.

Konarev, P. V., & Svergun, D. I. (2018). Direct shape determination of intermediates in evolving macromolecular solutions from small-angle scattering data. *IUCrJ*, *5*, 402–409. https://doi.org/10.1107/S2052252518005900.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., & Svergun, D. I. (2003). PRIMUS: A windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, *36*, 1277–1282.

Krebs, W. G., Alexandrov, V., Wilson, C. A., Echols, N., Yu, H., & Gerstein, M. (2002). Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic. *Proteins: Structure, Function, and Bioinformatics*, *48*, 682–695. https://doi.org/10.1002/prot.10168.

Lipfert, J., & Doniach, S. (2007). Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annual Review of Biophysics and Biomolecular Structure*, *36*, 307–327. https://doi.org/10.1146/annurev.biophys.36.040306.132655.

Liu, H., Morris, R. J., Hexemer, A., Grandison, S., & Zwart, P. H. (2012). Computation of small-angle scattering profiles with three-dimensional Zernike polynomials. *Acta Crystallographica. Section A*, *68*, 278–285. https://doi.org/10.1107/S010876731104788X.

Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I., & Lomize, A. L. (2012). OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Research*, *40*, D370–D376. https://doi.org/10.1093/nar/gkr703.

Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., et al. (2021). ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography*, *54*, 343–355. https://doi.org/10.1107/S1600576720013412.

Mertens, H. D., & Svergun, D. I. (2017). Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Archives of Biochemistry and Biophysics*, *628*, 33–41.

Miake-Lye, R. C., Doniach, S., & Hodgson, K. O. (1983). Anomalous x-ray scattering from terbium-labeled parvalbumin in solution. *Biophysical Journal*, *41*, 287–292. https://doi.org/10.1016/S0006-3495(83)84440-3.

Molodenskiy, D. S., Svergun, D. I., & Mertens, H. D. T. (2021). MPBuilder: A PyMOL plugin for building and refinement of solubilized membrane proteins against small angle X-ray scattering data. *Journal of Molecular Biology*, *433*, 166888. https://doi.org/10.1016/j.jmb.2021.166888.

Mylonas, E., & Svergun, D. I. (2007). Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *Journal of Applied Crystallography*, *40*, s245–s249.

Orthaber, D., Bergmann, A., & Glatter, O. (2000). SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *Journal of Applied Crystallography*, *33*, 218–225. https://doi.org/10.1107/S0021889899015216.

Pabst, G., Rappolt, M., Amenitsch, H., & Laggner, P. (2000). Structural information from multilamellar liposomes at full hydration: Full q-range fitting with high quality x-ray data. *Physical Review E*, *62*, 4000.

Panjkovich, A., & Svergun, D. I. (2016). Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Physical Chemistry Chemical Physics*, *18*, 5707–5719. https://doi.org/10.1039/C5CP04540A.

Pencer, J., & Hallett, F. R. (2000). Small-angle neutron scattering from large unilamellar vesicles: An improved method for membrane thickness determination. *Physical Review E*, *61*, 3003–3008. https://doi.org/10.1103/PhysRevE.61.3003.

Pencer, J., Krueger, S., Adams, C. P., & Katsaras, J. (2006). Method of separated form factors for polydisperse vesicles. *Journal of Applied Crystallography*, *39*, 293–303. https://doi.org/10.1107/S0021889806005255.

Pérez, J., & Koutsioubas, A. (2015). Memprot: A program to model the detergent corona around a membrane protein based on SEC–SAXS data. *Acta Crystallographica. Section D, Biological Crystallography*, *71*, 86–93. https://doi.org/10.1107/S1399004714016678.

Perkins, S. J., Wright, D. W., Zhang, H., Brookes, E. H., Chen, J., Irving, T. C., et al. (2016). Atomistic modelling of scattering data in the collaborative computational project for small angle scattering (CCP-SAS). *Journal of Applied Crystallography*, *49*, 1861–1875. https://doi.org/10.1107/S160057671601517X.

Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., et al. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, *45*, 342–350. https://doi.org/10.1107/S0021889812007662.

Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., & Svergun, D. I. (2007). ATSAS 2.1—Towards automated and web-supported small-angle scattering data analysis. *Journal of Applied Crystallography*, *40*, s223–s228.

Petoukhov, M. V., & Svergun, D. I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical Journal*, *89*, 1237–1250. https://doi.org/10.1529/biophysj.105.064154.

Petukhov, M. V., Konarev, P. V., Dadinova, L. A., Fedorova, N. V., Volynsky, P. E., Svergun, D. I., et al. (2020). Quasi-atomistic approach to modeling of liposomes. *Crystallography Reports*, *65*, 258–263. https://doi.org/10.1134/S1063774520020182.

Piiadov, V., Ares de Araújo, E., Oliveira Neto, M., Craievich, A. F., & Polikarpov, I. (2019). SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. *Protein Science*, *28*, 454–463. https://doi.org/10.1002/pro.3528.

Porod, G. (1951). Die Röntgenkleinwinkelstreuung von dichtgepackten kolloiden Systemen. *Kolloid Zeitschrift*, *124*, 83–114. https://doi.org/10.1007/BF01512792.

Rambo, R. P., & Tainer, J. A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, *496*, 477–481. https://doi.org/10.1038/nature12070.

Schneidman-Duhovny, D., Hammel, M., & Sali, A. (2010). FoXS: A web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Research*, *38*, W540–W544. https://doi.org/10.1093/nar/gkq461.

Seddon, A. M., Curnow, P., & Booth, P. J. (2004). Membrane proteins, lipids and detergents: Not just a soap opera. *Biochimica et Biophysica Acta (BBA)—Biomembranes, Lipid-Protein Interactions*, *1666*, 105–117. https://doi.org/10.1016/j.bbamem.2004.04.011.

Semenyuk, A. V., & Svergun, D. I. (1991). GNOM—A program package for small-angle scattering data processing. *Journal of Applied Crystallography*, *24*, 537–540. https://doi.org/10.1107/S002188989100081X.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Illinois: Urbana EEUU.

Stuhrmann, H. B. (1970). Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle scattering function. *Acta Crystallogr. Sect. A*, *26*, 297–306. https://doi.org/10.1107/S0567739470000748.

Stuhrmann, H. B., & Notbohm, H. (1981). Configuration of the four iron atoms in dissolved human hemoglobin as studied by anomalous dispersion. *Proceedings of the National Academy of Sciences*, *78*, 6216–6220.

Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, *25*, 495–503. https://doi.org/10.1107/S0021889892001663.

Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal*, *76*, 2879–2886.
Svergun, D., Barberato, C., & Koch, M. H. J. (1995). CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography*, *28*, 768–773.
Svergun, D. I., Koch, M. H., Timmins, P. A., & May, R. P. (2013). *Small angle X-ray and neutron scattering from solutions of biological macromolecules*. Oxford University Press.
Svergun, D. I., & Nierhaus, K. H. (2000). A map of protein-rRNA distribution in the 70 SEscherichia coli ribosome. *The Journal of Biological Chemistry*, *275*, 14432–14439.
Svergun, D. I., Petoukhov, M. V., & Koch, M. H. J. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal*, *80*, 2946–2953. https://doi.org/10.1016/S0006-3495(01)76260-1.
Tama, F., & Sanejouand, Y.-H. (2001). Conformational change of proteins arising from normal mode calculations. *Protein Engineering, Design & Selection*, *14*, 1–6. https://doi.org/10.1093/protein/14.1.1.
Tanford, C., & Reynolds, J. A. (1976). Characterization of membrane proteins in detergent solutions. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes*, *457*, 133–170. https://doi.org/10.1016/0304-4157(76)90009-5.
Tobi, D., & Bahar, I. (2005). Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proceedings of the National Academy of Sciences*, *102*, 18908–18913. https://doi.org/10.1073/pnas.0507603102.
Trewhella, J. (2022). Recent advances in small-angle scattering and its expanding impact in structural biology. *Structure*, *30*, 15–23. https://doi.org/10.1016/j.str.2021.09.008.
Tria, G., Mertens, H. D., Kachala, M., & Svergun, D. I. (2015). Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*, *2*, 207–217.
Wako, H., & Endo, S. (2017). Normal mode analysis as a method to derive protein dynamics information from the protein data Bank. *Biophysical Reviews*, *9*, 877–893. https://doi.org/10.1007/s12551-017-0330-2.
Zhang, R., Suter, R. M., & Nagle, J. F. (1994). Theory of the structure factor of lipid bilayers. *Physical Review E*, *50*, 5047–5060. https://doi.org/10.1103/PhysRevE.50.5047.