Small-Angle Scattering
Atomic Structure Based Modeling

Alejandro Panjkovich

EMBL Hamburg

07.12.2017
From the forest to the particle accelerator
Outline - Combining SAS data and atomic models

- Introduction
- Comparison of atomic models to scattering data (CRYSOL/CRYSON)
- Graphical interface to modeling (SASpy)
- Rigid-body modeling (SASREF)
- Missing fragments (BUNCH & CORAL)
- Flexible refinement (SREFLEX)
Structural biology techniques

- Macromolecular crystallography (up to atomic resolution)
- Nuclear magnetic resonance (up to atomic resolution)
- Electron microscopy (resolution revolution, already beyond 3 Å)
- Small angle X-ray scattering (nominal resolution 10 Å)

High brilliance EMBL beamline P12 at DESY synchrotron, Hamburg
Small angle X-ray scattering (SAXS)

- homogeneous and monodisperse solution (no crystal)
- sample can be at room temperature
- ‘no limitation’ in terms of size or oligomeric state
- requisites: 1.0 mg purified material, concentration > 0.5 mg/ml
Shape
Size

25 nm$^3$
SAXS applications in structural biology

- *ab initio* shape determination
- Atomistic (hybrid) modeling
  - Validation
  - Rigid-body
  - Missing fragments
  - Refinement
  - Conformational transitions
- Mixtures
- Ensemble approach
ATSAS software package
http://www.embl-hamburg.de/biosaxs/

- Large collection of programs for SAS data analysis.
- Online and standalone versions (Windows, Mac & Linux).
- Multiple algorithms and modeling approaches:
  - *ab-initio* (simulated annealing)
  - Rigid-body (Monte Carlo)
  - Ensemble approach (genetic algorithm)
  - Reference-less superposition
  - kd-trees search, PCA, etc.
Computing SAS from atomic model

Scattering from the particles can be obtained by subtracting solvent scattering, yielding effective density distribution:

\[ \Delta \rho = \langle \rho(r) - \rho_s \rangle \]  

(1)

where \( \rho_s \) is the scattering density of the solvent.

Bound solvent density may differ from that of the bulk.
Scattering from a macromolecule in solution

\[ I(s) = \langle |A(s)|^2 \rangle_\Omega = \langle |A_a(s) - \rho_s A_s + \delta \rho_b A_b(s)|^2 \rangle_\Omega \tag{2} \]

- \( A_a(s) \): atomic scattering in vacuum
- \( A_s(s) \): scattering from the excluded volume
- \( A_b(s) \): scattering from the hydration shell

Programs:
- CRYSON (neutrons): Svergun et al. (1998) P.N.A.S USA 95, 2267
Multipole expansion

\[ I(s) = \langle |A(s)|^2 \rangle_\Omega = \langle |A_a(s) - \rho_s E(s) + \delta \rho_b B(s)|^2 \rangle_\Omega \]  \hspace{1cm} (3)

If the intensity of each contribution is represented using spherical harmonics

\[ I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |A_{lm}(s)|^2 \]  \hspace{1cm} (4)

the average is performed analytically:

\[ I(s) = 2\pi^2 \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm}(s) - \rho_0 E_{lm}(s) + \delta \rho B_{lm}(s)|^2 \]  \hspace{1cm} (5)

This approach permits to further use rapid algorithms for rigid body refinement.
CRYSOL (X-ray) and CRYSON (neutron) scattering from macromolecules

\[
I(s) = 2\pi^2 \sum_{l=0}^{L} \sum_{m=-l}^{l} |A_{lm}(s) - \rho_0 E_{lm}(s) + \delta \rho B_{lm}(s)|^2
\]  

(6)

The programs:

- either fit the experimental data by varying the density of the hydration layer \(\delta \rho\) (affects the third term) and the total excluded volume (affects the second term)
- or predict the scattering from the atomic structure using default parameters (theoretical excluded volume and bound solvent density of 1.1 g/cm\(^3\))
- provide output files (scattering amplitudes) for rigid body refinement routines
- compute particle envelope function \(F(\omega)\)
How does the atomic model fit the solution scattering profile?

\[
\chi^2 = \frac{1}{N} \sum_{i=1}^{N_p} \left( \frac{l_e(s_i) - cI(s_i)}{\sigma(s_i)} \right)^2
\]  

(7)
CRYSOL: how to run CRYSOL (+ other ATSAS programs)

- command line, batch mode
- command line, interactive mode
- ATSAS-Online
- GUI, through Primus
- GUI, through SASpy
SASpy: integrating ATSAS and PyMOL

- **SASpy** combines **ATSAS** and **PyMOL**
- **ATSAS** SAS analysis software package
  - Freely available for academics
  - Compiled for Win, Mac and Linux
  - From initial data reduction to advanced modeling
- **PyMOL** Molecular visualization software
  - Visualize and edit 3D molecular models
  - Ray-rendering for publication quality figures
  - Extensible through ‘plugins’
SASpy - ATSAS PyMOL plugin

Panjkovich A & Svergun DI (2016) *Bioinformatics* 32, 2062-64
SUPALM, superposition of high- and low-resolution models

Konarev PV, Svergun DI. *IUCrJ.* 2015 Apr 21;2(Pt 3):352-60
SAS modeling principle

1D scattering data (or multiple data sets)

\[ \chi^2 = \frac{1}{N-1} \sum_j \left[ \frac{I_{\text{exp}}(s_j) - cI(s_j)}{\sigma(s_j)} \right]^2 \]

Trial-and-error

3D search model
\[ \mathbf{X} = \{X\} = \{X_1 \ldots X_M\} \]
M parameters

Non-linear search
The idea of rigid body modeling

- The structures of two subunits in reference positions are known.
- Arbitrary complex can be constructed by moving and rotating the second subunit.
- This operation depends on three Euler rotation angles and three Cartesian shifts.
The idea of rigid body modeling

- The structures of two subunits in reference positions are known.
- Arbitrary complex can be constructed by moving and rotating the second subunit.
- This operation depends on three Euler rotation angles and three Cartesian shifts.
Equation for rigid body modeling

Using spherical harmonics, the amplitude(s) of arbitrarily rotated and displaced subunit(s) are analytically expressed via the initial amplitude and the six positional parameters: $C_{lm}(s) = C_{lm}(B_{lm\alpha, \beta, \gamma, x, y, z})$.

The scattering from the complex is then rapidly calculated as:

$$I(s) = I_A(s) + I_B(s) + 4\pi^2 \sum_{0}^{\infty} \sum_{-l}^{l} \text{Re}[A_{lm}(s)C_{lm}^*(s)]$$

Constraints for rigid body modeling

- interconnectivity
- absence of steric clashes
- symmetry
- intersubunit contacts (from chemical shifts by NMR or mutagenesis)
- Distances between residues (e.g. FRET)
- Relative orientations of subunits (RDC by NMR)
- Scattering data from subcomplexes

Global rigid body modeling (SASREF)

- Fits (multiple X-ray and neutron) scattering curve(s) from partial constructs or contrast variation using simulated annealing
- Requires models of subunits, builds interconnected models without steric clashes.
- Uses constrains: symmetry, distance, relative orientation if applicable.

BUNCH combines rigid body and *ab-initio* modelling to find the positions and orientations of rigid domains and probable conformations of flexible linkers represented as dummy residues chains.

Multiple experimental scattering data sets from partial constructs (e.g. deletion mutants) can be fitted simultaneously with the data of the full-length protein.

accounts for symmetry, allows one to fix some domains and to restrain the model by contacts between specific residues.
Addition of missing fragments (CORAL)

- A combination of SASREF and BUNCH to account for missing loops in multi-subunit biological macromolecules.
- Loops are modeled based on known high-resolution structures.
User example: hybrid rigid-body modeling of a protein complex
Rigid-body modeling - SAXS model (SASREF)
Crystallographic complex 1

\[ \chi^2 = 6.7 \]
Crystallographic complex 2

\[ \chi^2 = 3.0 \]
SAXS-model vs. crystallographic model 2
Flexible refinement
Mismatch between homology model and SAXS data (and \textit{ab-initio} model)
SREFLEX fits SAS data by exploring conformational changes
Conformational change and ligand binding
SAXS and conformational change

- Crystalline and solution conformation may differ
- SAXS can provide insight into conformational transition
SREFLEX: SAS REFinement through FLEXibility
Estimating protein flexibility: normal mode analysis (NMA)

SREFLEX: SAS REFinement through FLEXibility

Input:
- SAXS data
- PDB coordinates

Program stages:
1. Structure partition
2. Domain level refinement
3. Residue level refinement
Automatic domain assignment based on dynamics
SREFLEX web interface (ATSAS online)

www.embl-hamburg.de/biosaxs/atsas-online/sreflex.php

Also available in Primus/qt, SASpy and through the command line
SREFLEX advanced usage

- Exploit options and parameters (more details in the manual)
  - Provide domain definitions a priori
  - Use CONCOORD refinement (protein only)
  - Skip one of the refinement stages

- Current limitations:
  - Long linkers (better use BUNCH or CORAL)
  - Bead models
  - Future developments:
    - Very large assemblies
    - Symmetric complexes
    - Membrane proteins
Model quality depends on SAXS data quality

- monodispersity
- radiation damage
- aggregation
- concentration effects
- overall parameters
- signal-to-noise level (SHANUM)
- ambiguity (AMBIMETER)
Ambimeter - ambiguity measurement

Ambimeter - different shapes, same scattering profile
DARA: structural neighbours, dara.embl-hamburg.de

- searches structural space available at PDB (~150000 structures)
- fast query thanks to PCA and k-d trees.
- nearest neighbours suggest shape, Rg, MW, etc.

Resources and references

- ATSAS manuals: www.embl-hamburg.de/biosaxs/manuals/
- ATSAS online: www.embl-hamburg.de/biosaxs/atsas-online/
- SAXS forum: www.saxier.org/forum/
- email: atsas@embl-hamburg.de, apanjkovich@embl-hamburg.de
Summary on combining SAS data with high-res models

- Comparison of atomic models to scattering data (CRYSOL/CRYSON)
- Graphical interface to modeling (SASpy)
- Rigid-body modeling (SASREF)
- Missing fragments (BUNCH & CORAL)
- Flexible refinement (SREFLEX)
Acknowledgements

- Dmitri Svergun and the whole BioSAXS team at EMBL Hamburg.
  - BioStruct-X 283570 and iNext.
  - Marie Curie Actions and EMBL for postdoc fellowship (EIPOD).