Rigid body refinement (basics)

D.Svergun, EMBL-Hamburg
Shapes from recent projects at EMBL-HH

- Transcription factor heterodimer
  - Matiasen et al
  - FEBS J (2016)

- Folded RTX Domain of CyaA
  - Bumba et al
  - Mol Cell (2016)

- Bivalent binding to BET bromodomains
  - Waring et al
  - Hajizadeh et al
  - Structure (2018)

- Bcr-Abl Tyrosine kinase PH domain
  - Reckel et al

- Chromatin remodeling enzyme Chd1
  - Sundaramoorthy et al
  - eLife (2017)

- Transcription factor heterodimer
  - Marcianò et al

- Folded RTX Domain of CyaA
  - Luo et al

- Bivalent binding to BET bromodomains
  - Multi-PDZ domain protein PDZK1
  - Hajizadeh et al
  - Structure (2018)

In some cases, high resolution models are drawn inside the shapes.
Using SAXS with MX/EM/NMR: ‘hybrid’ modelling

Model building where high resolution portions are positioned to fit the low resolution SAXS data
Monodisperse systems

Shape and conformational changes of macromolecules and complexes

Validation of high resolution models and oligomeric organization

Rigid body models of complexes using high resolution structures

Addition of missing fragments to high resolution models
To obtain scattering from the particles, solvent scattering must be subtracted to yield effective density distribution \( \Delta \rho = \langle \rho(r) \rangle - \rho_s \), where \( \rho_s \) is the scattering density of the solvent.

Further, the bound solvent density may differ from that of the bulk.
Scattering from a macromolecule in solution

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

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

**CRYSON (neutrons)**: Svergun et al. (1998) *P.N.A.S. USA*, **95**, 2267
The use of multipole expansion

If the intensity of each contribution is represented using spherical harmonics

\[ I(s) = \left< |A(s)|^2 \right>_\Omega = \left< |A_a(s) - \rho_s E(s) + \delta \rho B(s)|^2 \right>_\Omega \]

the average is performed analytically:

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

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

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 $F(\omega)$

$$I(s) = 2\pi^2 \sum_{l=0}^{L} \sum_{m=-l}^{l} \left| A_{lm}(s) - \rho_0 E_{lm}(s) + \delta \rho B_{lm}(s) \right|^2$$
Scattering components (lysozyme)

1) Atomic
2) Shape
3) Border
4) Difference
Effect of the hydration shell, X-rays

Experimental data

- Lysozyme
- Hexokinase
- EPT
- PPase

Fit with shell

Fit without shell

\[ s, \text{nm}^{-1} \]

\[ \lg I, \text{relative} \]
Denser shell or floppy chains: X-rays versus neutrons

- For X-rays: both lead to similar effect (particle appears larger)
- Floppy chains would in all cases increase the apparent particle size
- Neutrons in H$_2$O (shell): particle would appear nearly unchanged
- Neutrons in D$_2$O (shell): particle would appear smaller than the atomic model
Lysozyme: appears larger for X-rays and smaller for neutrons in D$_2$O

Thioredoxine reductase: CRYSOL and CRYSON fits with denser shell
Other approaches/programs I


Other approaches/programs II

- The ‘cube method’ (Luzzati et al, 1972; Fedorov and Pavlov, 1983; Müller, 1983) ensures uniform filling of the excluded volume. Could/should/must be superior over the effective atomic factors method at higher angles.

Further CRYSOL developments:
CRYDAM/CRYCUB

- Represent hydration shell by dummy water atoms
- Handle proteins, carbohydrates, nucleic acids and their complexes
- Represent excluded volume either by dummy atoms or by cubes
- Are applicable for wide angle scattering range

Petoukhov, M. & Svergun, D.I. planned as CRYSOL 3.0 in ATSAS 3.0
DARA, a database for rapid characterization of proteins

http://dara.embl-hamburg.de/

About 20000 atomic models of biologically active molecules are generated from the entries in Protein Data Bank and the scattering patterns computed by CRYSOL.

Rapidly identifies proteins with similar shape (from low resolution data) and neighbors in structural organization (from higher resolution data).

Recent developments: recalculation of the curves, new interface, new search (A.Kikhney, A.Panjkovich)

New DARA version: over 150,000 SAXS patterns, accelerated search new interface

Validation of high resolution models

Crystallographic packing forces are comparable with the intersubunit interactions. The solution structures of multisubunit macromolecules could be significantly different from those in the crystal.

Packing forces in the crystal restrict the allosteric transition in aspartate transcarbamylase.

Validation of high resolution models

NMR models of the Josephin domain of ataxin-3: red curve and chain: 1yzb, Nicastro et al. (2005) PNAS USA 102, 10493; blue curve and chain: 2aga, Mao et al. (2005) PNAS USA 102, 12700.

Domain Closure in 3-Phosphoglycerate Kinase

Closure of the two domains of PGK upon substrate binding is essential for the enzyme function. Numerous crystal structures do not yield conclusive answer, which conditions are required for the closure.

Identification of biologically active oligomers: Catalytic core of E2 multienzyme complex

The E2 cores of the dihydrolipoyl acetyltransferase (E2) enzyme family form either octahedral (24-mer) or icosahedral (60-mer) assemblies. The E2 core from Thermoplasma acidophilum assembles into a unique 42-meric oblate spheroid. SAXS proves that this catalytically active 1.08 MDa unusually irregular protein shell does exist in this form in solution.

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

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

Constraints for rigid body modelling

- Interconnectivity
- Absence of steric clashes
- Symmetry
- Intersubunit contacts (from chemical shifts by NMR or mutagenesis)
- Distances between residues (FRET or mutagenesis)
- Relative orientation of subunits (RDC by NMR)
- Scattering data from subcomplexes

Interactive and local refinement

♦ ASSA (SUN/SGI/DEC)

♦ MASSHA (Win9x/NT/2000)
Interactive and local refinement

SASpy (universal)

Global rigid body modelling (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 constraints: symmetry, distance (FRET or mutagenesis) relative orientation (RDC from NMR), if applicable

A global refinement run with distance constraints

A tyrosine kinase MET (118 kDa) consisting of five domains

Program SASREF

Single curve fitting with distance constraints: C to N termini contacts

**Quaternary structure of tetanus toxin**

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

Rigid body modelling of the Xpot ternary complex

Eleven X-ray and neutron curves

Atomic and homology models

Distance restraints from tRNA footprinting (Arts et al. (1998) EMBO J. 17, 7430)

Fukuhara et al. (unpublished)
Addition of missing fragments

- Flexible loops or domains are often not resolved in high resolution models or genetically removed to facilitate crystallization.
- Tentative configuration of such fragments are reconstructed by fixing the known portion and adding the missing parts to fit the scattering from the full-length macromolecule.
Building native-like folds of missing fragments

- Using DR-type models and protein-specific penalty functions

Primary sequence

Secondary structure

Excluded volume

Neighbors distribution

Knowledge-based potentials

Bond angles & dihedrals distribution

Addition of missing fragments: BUNCH

- 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.

- BUNCH accounts for symmetry, allows one to fix some domains and to restrain the model by contacts between specific residues.

Structure of sensor histidine-kinase PrrB

The dimeric sensor histidine-kinase PrrB from *Mycobacterium tuberculosis* contains ATP binding and dimerization domains and a 59 aas long (flexible) HAMP linker.

Structure of sensor histidine-kinase PrrB

The dimeric sensor histidine-kinase PrrB from *Mycobacterium tuberculosis* contains ATP binding and dimerization domains and a 59 aas long (flexible) HAMP linker.


A merger of SASREF and BUNCH: advanced methods to account for missing loops in multi-subunit protein structures (RANLOGS, CORAL)

A truncated construct WbdD\textsuperscript{1-459} is monomeric. For the construct WbdD\textsuperscript{1-556} MX yields an active trimer but AAs 505-556 are not seen in the crystal.

SAXS \textit{ab initio} shape reveals that the C-terminal is further extended. A rigid body model was constructed using coiled-coil C-terminal and refining the position of the catalytic domains.

\textit{In vivo} analysis of insertions and deletions in the coiled-coil region revealed that polymer size is controlled by varying the length of the coiled-coil domain.

A softer refinement with NMA

Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis


SREFLEX online

Project description

The first 8 characters in the description will be used to generate the project identifier.

SAXS data

Browse... No file selected.

Structure (.pdb or .zip)

Browse... No file selected.

Reset SUBMIT
Some words of caution

Or Always remember about ambiguity!
**Sampling formalism**

**Shannon sampling theorem**: the scattering intensity from a particle with the maximum size $D$ is defined by its values on a grid $s_k = k\pi/D$ (Shannon channels):

$$sI(s) = \sum_{k=1}^{\infty} s_k a_k \left[ \frac{\sin D(s - s_k)}{D(s - s_k)} - \frac{\sin D(s + s_k)}{D(s + s_k)} \right]$$

Shannon sampling was utilized by many authors (e.g. Moore, 1980). An estimate of the number of channels in the experimental data range ($N_s = s_{\text{max}} D/\pi$) is often used to assess the information content in the measured data.
Shape determination: $M \approx 10^3$ variables (e.g. 0 or 1 bead assignments in DAMMIN)

Rigid body methods: $M \approx 10^1$ variables (positional and rotational parameters of the subunits)

From the informational point of view, rigid body modeling should provide unique or at least much less ambiguous models than shape determination

**NO WAY**

As all the problems are non-linear, the number of Shannon channels does not give you exact number of parameters, which is possible to extract from the scattering data (depending on accuracy, this number varies between zero and infinity).

Further, uniqueness of reconstruction depends largely on the complexity of the function $f(x)$ to be minimized
Ambiguity of rigid body analysis

A synthetic example: two different orientations of tRNA in a dimeric complex with aspartyl-tRNA synthetase obtained by rigid body modelling and compatible with X-ray and contrast variation neutron scattering data.

Sampling formalism appears to be a good tool to determine the useful data range. Given a (noisy, especially at high angles) experimental data set, which part of this set provides useful information for the data interpretation?

A usual practice is to cut the data beyond a certain signal-to-noise ratio but
• there is no objective estimation of the threshold
• this cut-off does not take into account the degree of oversampling
Determination of the useful data range

The useful range is defined by the number of meaningful Shannon channels $N_M$, which can be determined from the data set. An algorithm is developed to determine this range based on fitting Shannon representations with increasing number of channels. Note: depending on errors and oversampling, $N_M$ may be smaller or even larger than $N_S$.

Constraints and restraints used in global modelling procedures

- Information about contacting residues from other experiments (spin labelling, site-directed mutagenesis, FRET, chemical shifts etc)
- Information about symmetry
- Avoiding steric clashes
- For missing loops and linkers: contiguous chain conditions, excluded volume, Ramachandran plot for $C_\alpha$, knowledge-based potentials etc

AND STILL, one must always cross-validate SAS models against all available biochemical/biophysical information
Architecture of nuclear receptor heterodimers on DNA direct repeat elements

Nuclear hormone receptors (NHRs) control numerous physiological processes through the regulation of gene expression. SAXS, SANS and FRET were employed to determine the solution structures of NHR complexes, RXR–RAR, PPAR–RXR and RXR–VDR, free and in complex with the target DNA.

Ab initio and rigid body models of NHRs complexed with direct repeat elements

Architecture of nuclear receptor heterodimers on DNA direct repeat elements

NHR-DNA complexes show extended asymmetric shape and reveal conserved position of the ligand-binding domains at the 5′ ends of the target DNAs. Further, the binding of only one coactivator molecule per heterodimer, to RXR’s partner, is observed.

The models and the polarity of RXR–RAR–DR5 and RAR–RXR–DR1 were validated using neutron scattering and FRET.

Hybrid user projects at P12 beamline

- Tungstate transporter protein TupA
- Shigella Virulence Factor IcsA
- Autoinhibited Ubiquitin Ligase Prp19
  - de Moura et al, Mol Cell (2018)
- Membrane-proximal domains of epsin
- Mechanosensitive ion channel protein T2
- PDH complex E2 core
- ABC Transporter MsbA in stealth nanodisk
- Ammonium sensor histidine kinase
- de Moura et al, Mol Cell (2018)
- Flayhan et al, Structure (2018)
By the way, can X-ray scattering yield the fold?

- Lysozyme and its near-native scattering mates
And now, let us awake for the hands-on practical