How to manage a flexible system with SAXS

Dr Haydyn Mertens
EMBL Hamburg
SAXS in structural biology

Complementary techniques

<table>
<thead>
<tr>
<th>MS</th>
<th>NMR</th>
<th>EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MX</td>
<td>Bio-informatics chemistry</td>
<td>FRET</td>
</tr>
</tbody>
</table>

Additional information

- Homology models
- Atomic models
- Distances
- Orientations
- Interfaces
- Sequence

Data analysis
- Shape determination
- Rigid body modelling
- Missing fragments
- Oligomeric mixtures
- Hierarchical systems
- Flexible systems

October 20, 2016
Haydyn Mertens, EMBO 2016 (Hamburg)
Solution VS crystals

- Crystals not required
- Absence of crystal packing forces
- SAXS not limited by molecular mass, and measurement under near physiological conditions
- SAXS in solution can more easily follow reactions/responses to change in conditions
- SAXS facilitates quantitative analysis of complex systems and processes
- Flexible macromolecules!
No function without structure
But flexibility is so cool!
.... and important!

- Hub proteins in interaction networks
- Interaction specialists

Scattering from dilute macromolecular solutions (monodisperse systems)

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

\[ I(s) = 4\pi \int_{0}^{D} p(r) \frac{\sin sr}{sr} dr \]
Polydispersity

• Modelling from SAXS data usually assumes:
  • Monodispersity
  • Absence of interparticle interactions (dilute)
  • Knowledge of sample identity
Scattering from a mixture

- Size polydispersity
- Total scattering is a weighted sum

\[ I(s) = \sum_k v_k I_k(s) \]
Scattering from a mixture

- Shape polydispersity
- eg. monomer-dimer equilibrium

\[ I(s) = \sum_k v_k I_k(s) \]
Scattering from a mixture

- Conformational polydispersity

\[ I(s) = \sum_k v_k I_k(s) \]
Scattering from a mixture

- Both?

\[ I(s) = \sum_k v_k I_k(s) \]
Scattering from a mixture

- Size/Shape polydispersity (eg. distributions, oligomers)
  - If component structure unknown requires additional parameters
- Conformational polydispersity (eg. IDPs)
  - Almost infinite range of conformations
  - Cannot really identify all possible $v_k$ and $I_k(s)$
  - Requires a more indirect approach
Flexibility characterisation directly from data

- Kratky plot, (Kratky 1982)
  - $I^*s^2$ vs $s$
- Dimensionless Kratky, (Durand, 2010)
  - $(I/I_0)(sR_g)^2$ vs $sR_g$

![Graphs showing flexibility characterisation](image)
Flexibility characterisation directly from data

- “Featureless” curves $\rightarrow$ flexible
- Clear features $\rightarrow$ “rigid”

ATSAS integrated workflow

**Data collection**
- PRIMUS
- Ambimeter
- SHANUM
- GNOM

**MONODISPERSE**
- No *apriori* information
  - Ab *initio* modeling
    - DAMMIN
    - GASBOR
    - SUPALM/DAMAVER/SASRES

- Partial hi-res model available
  - Hybrid modeling
    - "rigid" system
      - EOM
    - flexible system
      - SREFLEX
      - CRYSON

- Hi-res model available
  - SVD
  - CRYSOL
  - CRYSON

**SVD**
- Rapid search of structural neighbours using solution SAXS data

**EMBL**
- EMBO 2016 (Hamburg)
- Haydyn Mertens, EMBO 2016 (Hamburg)

October 20, 2016
OLIGOMER (Konarev et al, 2003)

- eg. monomer-dimer equilibrium

\[ I(s) = \sum_k v_k I_k(s) \]
OLIGOMER (Konarev et al, 2003)

- eg. conformational equilibrium

\[ I(s) = \sum_k v_k I_k(s) \]
Conformational polydispersity

- Ensemble based approaches
  - When many structures are required to describe the data
  - Flexible systems (eg. IDPs)
  - Chemically denatured proteins
  - Flexible multi-domain proteins

Mertens & Svergun, JSB, 2010, 172(1), 128-141
EOM (Bernado et al. 2007, Tria et al. 2014)

- Ensemble Optimisation Method
EOM (Bernado et al. 2007, Tria et al. 2014)

- Ensemble Optimisation Method ... more detail

EOM (Bernado et al. 2007, Tria et al. 2014)

• Required input

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

Experimental data (*.dat)

Sequence (*.txt)

Models (*.pdb) (rigid bodies)

Symmetry

EOM

\( \chi^2 \) (fit)

\( R_g \) dist.

\( D_{max} \) dist.
LRCMQCKTNDCRVEECALGQDLCRTTIVRLWEEGEELVEKS
CTCSEKTNRALSRTGLKITSLTEVVCGLDLCNQGNSGRAVTY
RSRYLECISCGSDMCERGRHQSLQCRSPEEQCLDVVTWHIE
GEEGRPKDDRHLRGCGYLPGPSNGFHNNDTFHFLKCCNTTKC
NEGPILELENLPQNGRQCYSCKGNSTHGCSEETFIDCRGPMN
QCLVATGTHEPKNQSYMVRGCATASMCQHAHLGDAMSCHIDVS
CCTKSCGNHLPDLDVQYRS

Rigid body 1 (PDB)
Rigid body 2 (PDB)
Rigid body 3 (PDB)
EOM

LRCMQCKTNGDCRVEECALGQDLCRTTIIVRLWEEGELELVEKS
CTCSEKTNRTLSYRTGLKITSLTEVVCGLDLCNQGNSGRAVTYS
RSRYLECISCSDMASCERGRHQSLQCRSPEEQCLDVVTWHIQE
GEEGRPKDDRHLRGCYLGCPGNSNGFHNNDTFHLKCCNTTKC
NEGPILELENLPQNGRQCYSCGNSHTHSSEETFIDCGRPMN
QCLVATGTHEPKNQSYMVRGCATASMQHAILGDFSMCHIDVS
CCTKSGCNHPDLDVQRSG

Rigid body 1 (PDB)
Rigid body 2 (PDB)
Rigid body 3 (PDB)
• Symmetry

- Symmetric core
- Symmetric linkers/termini
- Asymmetric linkers/termini
Example: Tau protein structure?

- IDP even when bound to microtubules
- Tau repeat is source of residual secondary structure

Mylonas et al. (2008), *Biochemistry* 47:10345-10353
Example: uPAR

- Flexibility driving function: uPAR
- Urokinase plasminogen activation receptor

- uPAR is a receptor involved in cell-adhesion and plasminogen activation
- Receptor flexible (SAXS)
- Therapeutics based on ligand
- Decreased flexibility upon drug binding → and metastasis???

Mertens et al., JBC, 2012, 287(41), 34304-34315
Example: uPAR

- Flexibility driving function: uPAR
- Urokinase plasminogen activation receptor

- uPAR is a receptor involved in cell-adhesion and plasminogen activation
- Receptor flexible (SAXS)
- Therapeutics based on ligand
- Decreased flexibility upon drug binding → and metastasis???

Mertens et al., JBC, 2012, 287(41), 34304-34315
Example: uPAR

- Flexibility driving function: uPAR
- Urokinase plasminogen activation receptor

Mertens et al., JBC, 2012, 287(41), 34304-34315
Example: uPAR

- Flexibility driving function: uPAR
- Urokinase plasminogen activation receptor

Mertens et al., JBC, 2012, 287(41), 34304-34315
Example: uPAR

- Flexibility driving function: uPAR
- Urokinase plasminogen activation receptor

- uPAR is a receptor involved in cell-adhesion and plasminogen activation
- Receptor flexible (SAXS)
- Therapeutics based on ligand
- Decreased flexibility upon drug binding → and metastasis???
EOM

- Analysis procedures (EOM v2.0)

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

0% \( \leftarrow \) \( R_{\text{flex}} \rightarrow \) 100%

Rigid \hspace{0.5cm} Flexible

(high uncertainty)

\[ R_{\sigma} = \frac{\sigma_s}{\sigma_p} \]

0 \( \leftarrow \) \( R_{\sigma} \rightarrow \) 1

Rigid \hspace{0.5cm} Flexible

Tria et al., 2014

\[ H_b(S) = -\sum p(x_i) \log_b(p(x_i)) \]
EOM

• Analysis procedures (EOM v2.0)

IUCr
Tria et al., 2014

B

Metric

$R_{\text{flex}} (R_g)$

$R_g$

$D_{\text{max}}$

randomness (84.7%)

Value

82.1% (1.03)

45.3% (0.10)

2.8 (nm)

2.2 (nm)

9.5 (nm)

7.2 (nm)
EOM

• Analysis procedures (EOM v2.0)

Tria et al., 2014

\[EOM\]

\[\text{Analysis procedures (EOM v2.0)}\]
Crystal structures of substrate-bound chitinase from *Moritella marina* and its structure in solution

- Chitinases break down glycosidic bonds in chitin and only few crystal structures are reported because of the flexibility of these enzymes.
- Dimeric crystal structure of MmChi60 contains four domains: catalytic, two Ig-like, and chitin-binding (ChBD).
- SAXS demonstrates that MmChi60 is monomeric and flexible in solution. The flexibly hinged Ig-like domains may thus allow the catalytic domain to probe the surface of chitin.

Ensemble Analysis Summary

- Ensemble Optimization Method (EOM)
  - Size distributions
  - Useful metrics based on entropy
    - \( R_{\text{flex}} = -H_b(S) \)
    - \( R_{\text{sigma}} = \sigma_S / \sigma_p \)

\[ H_b(S) = \sum p(x) \log_b p(x) \]

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_{\text{flex}} )</td>
<td>82.1% (1.0)</td>
</tr>
<tr>
<td>( R_{\text{sigma}} )</td>
<td>45.3% (0.10)</td>
</tr>
<tr>
<td>( R_g )</td>
<td>2.8 (pm)</td>
</tr>
<tr>
<td>( D_{\text{max}} )</td>
<td>9.5 (pm)</td>
</tr>
</tbody>
</table>

Low: “rigid” → Uncertainty → high: flexible

October 20, 2016

Haydyn Mertens, EMBO 2016 (Hamburg)
SREFLEX

- Elastic network model
- Normal modes
- Automatic “domain” definition
- Structure deformed to best fit data

October 20, 2016

Haydyn Mertens, EMBO 2016 (Hamburg)
Summary

- Polydisperse systems
  - Oligomeric equilibria
  - Conformational equilibria
- Useful approaches
  - SAXS profiles (features vs featureless)
  - Kratky representations
  - Ensemble methods
- Software
  - OLIGOMER (Konarev et al. 2003)
  - EOM (Bernado et al., 2007; Tria et al., 2014)
  - SREFLEX (Panjkovic & Svergun, 2015)
... don’t be afraid of flexible systems!
Acknowledgments

- BIOSAXS team EMBL
- Dmitri Svergun
- SAXS/SANS user community