SREFLEX: SAXS REFinement through FLEXibility

Alejandro Panjkovich
EMBL Hamburg

20.10.2016
Outline

- Motivation for flexible refinement
- SREFLEX method
- Example from user
- Running the program
  - Standard run
  - Advanced domain definition
- Interpreting results
SAXS and conformational change

- Crystalline and solution conformation may differ
- SAXS can provide insight into conformational transition
SAS modeling principle

1D scattering data (or multiple data sets)

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

Trial-and-error

3D search model \( \mathbf{X} = \{X\} = \{X_1 \ldots X_M\} \)
M parameters

Non-linear search
Estimating protein flexibility: normal mode analysis (NMA)


Cα trace → elastic network model
Time for some movement
SREFLEX: SAXS REFinement through FLEXibility

A. Panjkovich (EMBL)

A. native
B. restrained
C. complete

D. log I(s) relative

s, Å⁻¹

target
initial
restrained
complete

20.10.2016 7 / 27
SREFLEX: SAXS 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
Assignment comparison, SCOP vs. auto
Experimental example: Josephin domain of ataxin-3

Experimental example: MurA

SREFLEX: examples from benchmark

Figure 4A: DNA-binding domain, modeling the unbound-to-bound transition. SAXS profiles show the better consistency to the target profile (dots) of the SREFLEX scop model over the SREFLEX auto model. Structures shown correspond to: (A) the SREFLEX scop model in red superimposed to the target structure in green. (B) Pseudo-domains as defined by SREFLEX auto. (C) SCOP domains used by SREFLEX scop, vectors show the modeled movement.

Figure 5: Limitations of the approach. Two known structures of calmodulin in different conformation are shown superimposed, even though they differ considerably (RMSD is 10.2 Å). The corresponding theoretical SAXS profiles are very similar, meaning that in this case SREFLEX will not be able to model the conformational change as explained in the text.
Themed issue: Exploring the conformational heterogeneity of biomolecules

A. Panjkovich (EMBL)
SAXS + NMA
SREFLEX, user example
SREFLEX, *ab-initio* model mismatch
SREFLEX, conformational change
Conformational change fits ligand binding
SREFLEX, web interface (ATSAS online)

SREFLEX online

Project ID

SAXS data
Browse... No file selected.

Structure (.pdb or .zip)
Browse... No file selected.

Reset SUBMIT

About SREFLEX

© BioSAXS group 2015
SREFLEX, command line interface

Usage: sreflex [OPTIONS] <SAXS FILE> <COORD FILE>

Arguments:
<SAXS FILE> SAXS curve
<COORD FILE> PDB coordinates

Options:
-p, --prefix output directory prefix/name
-Q, --quiet do not print log to stdout
-t, --threads=<INT> # of threads/CPU's to use
-a, --clashes=<INT> max clashes, default 5, set -1 to skip
-b, --breaks=<D/A/L> structural breaks mode: Default, Allow or Limit
-r, --ratio=<FLOAT> set convergence ratio, default 0.7, range [0.5:0.9]
-f, --first=<INT> first SAXS data point to consider, default 1
-n, --nntop=<INT> top normal mode to consider, default 16, range [9:64]
-s, --skip=<STAGE> skip RESTRAINED or UNRESTRAINED refinement stage
-v, --version print version and exit
-h, --help print this help and exit

Additionally, the following arguments can be forwarded to CRYSOL:
--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 (default 1)
--dns Solvent density (default 0.334 e/A**3)
--dro Contrast of hydration shell (0.03 e/A**3)
--cst Subtract constant (default no)
--eh Account for explicit hydrogens

Report bugs to atsas@embl-hamburg.de
Divide et impera - split your structure into pieces

- The new ATSAS binary partistr may help
- Check Pfam or SCOP domains
- Use your intuition (particularly for opening structures)
- Feed the structure as
  - d1.pdb,d2.pdb,d3.pdb (command line)
  - Or as a .zip file to the ATSAS-online version
Interpretation of SREFLEX results

- **Restrained** models (rc), higher $\chi^2$ but better stereochemistry
- **Unrestrained** models (uc), lower $\chi^2$, more structural distortion
- Prioritize lower RMSD against initial structure
- Check degree and relevance of clashes and breaks
- Make an informed decision on which model is the better choice

---

### SREFLEX REPORT

- **VERSION:** ATSAS 2.8.0 (r8073:8152)
- **DATE AND TIME of run:** 2016-08-11 22:43:16
- **INPUT SAXS data filename:** closed1akeA.dat
- **INPUT PDB coordinates:** open4akeA.pdb
- **Initial Chi2:** 101.35
- **Initial clashes:** 0

### RESULTS SUMMARY:

<table>
<thead>
<tr>
<th>model</th>
<th>Chi2</th>
<th>RMSD</th>
<th>breaks</th>
<th>clashes</th>
<th>modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>rc01</td>
<td>1.13</td>
<td>5.40</td>
<td>0.39</td>
<td>0.00</td>
<td>n7s-3xn8s1</td>
</tr>
<tr>
<td>rc02</td>
<td>2.35</td>
<td>3.04</td>
<td>0.90</td>
<td>0.19</td>
<td>n7s-3xn8s1xn9s2</td>
</tr>
<tr>
<td>uc01</td>
<td>1.02</td>
<td>4.79</td>
<td>0.42</td>
<td>0.00</td>
<td>n7s-3xn8s1_X_n7s-1</td>
</tr>
<tr>
<td>uc02</td>
<td>1.02</td>
<td>2.74</td>
<td>0.46</td>
<td>0.09</td>
<td>n7s-3xn8s1_X_n7s-1xn9s-1</td>
</tr>
</tbody>
</table>

---

---

A. Panjkovich (EMBL)
Use the proper tool for your question

- MW or oligomeric state are way off (better SASREF)
- There are long linkers or unstructured regions (better use EOM)
- Unrealistic expectations (folding problem not yet solved)
Summary

- Motivation
- Understanding the approach
- Running the program
  - Standard run
  - Custom domain definition
- Interpreting results
Acknowledgements

- Dmitri Svergun and the rest of the SAXS team at EMBL Hamburg (https://www.embl-hamburg.de/biosaxs/)
- Funding:
  - EMBL and Marie Curie Actions for postdoc fellowship (EIPOD)
SAS in structural biology

- \textit{ab-initio} shape determination
- missing fragments
- rigid-body hybrid modeling
- mixtures
- ensemble approach