Five reasons to pay attention to Small Angle X-ray Scattering Assisted CASP

Susan Tsutakawa, Greg Hura, and John Tainer

Advanced Light Source
Beamline 12.3.1 (SIBYLS)

sibyls.lbl.gov
www.bioisis.net
Why SAXS?
Five reasons to pay attention to Small Angle X-ray Scattering Assisted CASP

<table>
<thead>
<tr>
<th>1.</th>
<th>2.</th>
<th>3.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAXS data can be collected at a fraction of the cost, time, and labor of other structural techniques.</td>
<td>SAXS data provides distance information as experimental restraints for prediction algorithms.</td>
<td>SAXS data can filter starting models, provide shape, provide fold info, and orient domains and subunits.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.</th>
<th>5.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAXS data can experimentally validate prediction models.</td>
<td>IT WORKS! (with need for improvement)</td>
</tr>
</tbody>
</table>
SAXS is a distance method, measuring shape and all electron pair distances (including flexible and cross-subunit).

SANS is a complementary method for complexes, separating the signal from one subunit from the others.

The p(r) is the sum of all "distograms" within the protein plus hydration layer.
How good was the SAXS data provided for CASP13?

Note: ALL SAXS data was experimental. We have the throughput to do this.
The two SAXS data methods used are good for proteins that are flexible and/or multimerizing.

Technical challenges intrinsic to proteins in solution

1. Protein sample is low conc.
2. Protein is flexible
3. Protein is multimerizing
4. Stoichiometry is heterogeneous or protein is aggregated.

<table>
<thead>
<tr>
<th>High throughput (HT-SAXS)</th>
<th>Size-exclusion Chromatography (SEC-SAXS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good signal to noise</td>
<td>Homogenous stoichiometry</td>
</tr>
</tbody>
</table>

H0980 and T0999 collected by outside groups.
We provided metrics, SAXS curves, and our quality assessment, including flexibility metric.

### Sample Quality

**H0957 – Analysis 4/20/18**

- Small heterogeneity observed and corrected for in SEC-SAXS

### Challenge

- 1:1 Complex

### Sample: CASP Chicago209Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Error +/-</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rg</td>
<td>21.8</td>
<td></td>
<td>Angstroms</td>
</tr>
<tr>
<td>Porod Exponent</td>
<td>4.0</td>
<td></td>
<td>Scale (2-4)</td>
</tr>
<tr>
<td>Mass SAXS</td>
<td>32</td>
<td></td>
<td>kDa</td>
</tr>
<tr>
<td>Mass</td>
<td>4 gold</td>
<td>6 silver</td>
<td>1 bronze</td>
</tr>
<tr>
<td>Max Dimension</td>
<td>71</td>
<td>3</td>
<td>Angstroms</td>
</tr>
<tr>
<td>Radius of Cross Section</td>
<td>18.4</td>
<td>1</td>
<td>Angstroms</td>
</tr>
<tr>
<td>Volume</td>
<td>54545</td>
<td>5,000</td>
<td>Cubic Angstroms</td>
</tr>
<tr>
<td>Real Space Rg</td>
<td>21.58</td>
<td>3</td>
<td>Angstroms</td>
</tr>
</tbody>
</table>

Chain A:

```
SNSFEVSSLPDANGKHITAVKGDAKIPVDKIELYMRGKASGLDLSQAEYNSLKDARISSQKEFAKDPNNAKRMVELVLEKQHNIERSQDMA
RVLEQAGIVNTASNNSMIMDKLDDLSCAQGATSANRTSVVVSGPANNVRIATWTLPGTSDKRLSHTVGTFK
```

Chain B:

```
SNAMINVNSAKDIEGLESYLANYVEANSFNDPDEDDALECSLNVKDSRGLSFCKKILNNSNIDGVFKGSALNFLSLSEQWSYAFEYLTSS
NADNITLAELEKALFYFYCAKNETDPYPVPEGFKKLMKRYEELKNDPDAKYHLYDSTFSKAYPLNN
```

### Mass

- 4 gold
- 6 silver
- 1 bronze
What information about the targets is provided by SAXS (solution)? How does it compare to crystal (lattice/assessment)?
11 CASP13-SAXS targets include 4 monomers and 7 multimers, 14-340 kDa.

- **T0992**: 14 kDa
- **T0949**: 17 kDa
- **T0987**: 46 kDa
- **T0975**: 38 kDa
- **H0957**: 36 kDa
- **T0985**: 200 kDa
- **T0999**: homodimer 340 kDa
- **H0980**: 2:2 40 kDa
- **H0968**: 54 kDa 2:2
- **H0953**: 3:1 48 kDa
- **T0981**: homotrimer 228 kDa
Compared to 2016 CASP12-SAXS, the structures had much less disordered regions. Only 3 were missing more than 10% and those were maximum 18%.
6/11 Crystal Structures fit within the SAXS-derived shape. 5/11 stick out a bit.

CASP: Image redacted

not quite

not quite

not quite

not quite

not quite

not quite
4 Crystal Structures + disordered tails agreed with solution SAXS data based on comparison of Scattering Curve.

6-7 Will likely Require Domain movements.
4 Crystal Structures + disordered tails agreed with solution SAXS data based on comparison of Scattering Curve.

6-7 Will likely Require Domain movements.

6% 18% 2%

CASP: Image redacted
How did the predictors do? The Devil is in the details.

Two ways that SAXS data helps.
1. Improve overall shape (density using gmfit tool)
2. Improve fold (GDT-TS, QCS score)

Considerations
1. Some groups only submitted SAXS predictions. Cannot assess SAXS-based improvement.
2. Should see improvement on whole protein. GDT-TS & QCS scoring by domains or polypeptides.
3. SAXS data in solution does not necessarily match crystal.
Discrepancy between model sequence and SAXS sample suggests modelers might be mislead by fitting for something that is supposed to be there or not there.

6/11 models – AA sequence matched SAXS sample
Targets H0953, H0957, H0968, T0975, H0980 (-6 nt), T0999

2/11 models – prediction seq sometimes did not match SAXS sample
3/11 models – all model seq did not match SAXS sample.
Prediction models -102 to +32 AA.
Crystal sequence might not match SAXS sample

SAXS measures **ALL** electron pairs of protein & its hydration layer.
*Please double check that model sequence matches the SAXS sample*
Predictions showed improvement in overall shape.
Shape – Predictors improved their overall shape with SAXS data.

H0953 elongated

Regular (unassisted)

SAXS-assisted

Density, First by first, all targets, all groups

Gmfit – Dmytro Guzenko
For the top scoring GDT_TS model for H0953s2, can visibly see improved fold.
<table>
<thead>
<tr>
<th></th>
<th>D-Haven without SAXS</th>
<th>D-Haven with SAXS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDT_TS</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Density</td>
<td>0.625</td>
<td>0.625</td>
</tr>
<tr>
<td>TM</td>
<td>0.28</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Only 4 groups participated (D-Haven, kozakov-vajda, Grudinin, SBROD)—why? 

CASP: Image redacted
Some improvement on SAXS-assisted fold.
SAXS helps with fold and with ranking in top scoring SAXS-assisted GDT_TS models.

**SAXS Model 1 vs Reg Model 1**

**SAXS Best vs Best**
By group, Liwo consistently improves. Grudinin and Tomii see improvements with some targets, where they become one of the top scoring groups.

Delta, ordered by group, first vs first GDT_TS

Liwo and Grudinin are discussion participants in Workshop 1 (after this session)

Tomii is a discussion participants in Workshop 2 (3 pm today)
Individual Example
For S0968 s2 which showed GDT_TS improvement, see that edges improve. For biologists, edges are important – it’s where active sites and interfaces are.

196/Grudinin (#1 saxs)

329/D-Haven

GDT_TS score improves to 72 (top)
For S0957 s2 which showed GDT_TS improvement, see that biologically-important edges improves.

GDT_TS score improves to 60 (top).

<table>
<thead>
<tr>
<th>#</th>
<th>Model</th>
<th>GDT_TS</th>
<th>T0</th>
<th>S</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>_0957s2TS320_1</td>
<td>41.452</td>
<td>60.068</td>
<td>16.516</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>_0957s2TS196_1</td>
<td>40.000</td>
<td>56.935</td>
<td>16.935</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>_0957s2TS135_1</td>
<td>40.000</td>
<td>49.355</td>
<td>9.355</td>
<td></td>
</tr>
</tbody>
</table>
There was some improvement for all models, but most significant improvement in certain cases.
See most improvements in GDT_TS score in heteromeric complexes, medium size, AA correct.

- **T0992**: 14 kDa
- **T0949**: 17 kDa
- **T0987**: 46 kDa
- **T0975**: 38 kD (FE-S)
- **H0957**: 36 kD (1:1)
- **T0985**: 200 kDa (2:2)
- **T0981**: 228 kD (3:1)
- **H0980**: 54 kDa (2:2)
- **H0968**: 48 kD (3:1)
- **T0999**: homodimer 340 kD
- **H0953**: 54 kDa (2:2)
- **H0957**: 48 kD (1:1)
Improved models do not correlate with flexibility or fit in the SAXS envelope...

T0992 14 kDa
T0949 17 kDa
T0987 46 kDa
T0975 38 kDa
H0957 1:1 36 kDa
T0981 homotramer 228 kD
T0985 2 200 kDa

H0953 3:1 48 kD
H0980 2:2 40 kD
H0968 54 kDa 2:2

CASP: Image redacted
Structures that predictors overall had low scores were in the not improved.

- **small**
  - T0992
  - T0949
  - T0987
  - T0975
  - H0999
  - H0980
  - H0968
  - H0953
  - T0981

- **template**
  - T0985

<table>
<thead>
<tr>
<th>Structure</th>
<th>Mass (kDa)</th>
<th>CASP Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0992</td>
<td>14</td>
<td>&lt;80</td>
</tr>
<tr>
<td>T0949</td>
<td>17</td>
<td>&lt;25</td>
</tr>
<tr>
<td>T0987</td>
<td>46</td>
<td>&lt;40</td>
</tr>
<tr>
<td>T0975</td>
<td>38</td>
<td>&lt;40</td>
</tr>
<tr>
<td>H0999</td>
<td>340</td>
<td>&lt;20</td>
</tr>
<tr>
<td>H0980</td>
<td>2:2</td>
<td>40</td>
</tr>
<tr>
<td>H0968</td>
<td>2:2</td>
<td>54</td>
</tr>
<tr>
<td>H0953</td>
<td>3:1</td>
<td>48</td>
</tr>
<tr>
<td>T0981</td>
<td>200</td>
<td>&lt;60</td>
</tr>
</tbody>
</table>

CASP: Image redacted
Could the SAXS have helped to reach the crystal structure?
Predictors have room for improvement – crystal structure fits the exp data best.

ExpSAXS
Grudinin
Liwo

Crystal
FoxsMod

ExpSAXS
Grudinin
Liwo

P(r)

Target 0953 - Residuals for SAXS

Residuals

q (Å⁻¹)

0

0.1

0.2

0.3

0.4

10

100

1000

P(r)

Target 0953 - Residuals for SAXS

Residuals

q (Å⁻¹)

0

0.1

0.2

0.3

0.4

0.001

0.01

1

10

100

P(r)

ExpSAXS
Grudinin
Liwo

Crystal
ModFOXS
CASP13 Assessor Results for SAXS-assisted category.  Target variability makes unifying summary difficult

Improvement of models for some targets
SAXS provided accurate guidance in most cases
Further improvement is required (please work with our beamline)

Workshops 1 and 2
Improved metrics??
Improved integration with algorithms??
Simulated SAXS data for crystal lattice structures??
Higher resolution SAXS data??
Novel methods to reduce signal from disorder regions??
Dynamic predictions??