The Bio-SAXS beamline

Solution Scattering from Biological Macromolecules
What information can we obtain?

Contents

- What can be obtained from Bio-SAXS
  - Measurable parameters
  - Modelling strategies

- How to collect Solution SAXS data at ID14-3
  - Procedure
    - Data collection tests
    - Data Verification and quality control
    - Modelling and analysis
  - Practical demonstration at the Bio-SAXS beamline
Solution SAS Experiment

Black Box
Neutron/X-ray beamline
Home source

Solution Scattering Data from Protein of Interest

Radius of Gyration and Zero Angle Intensity
The size of your protein
**Radius of Gyration and Zero Angle Intensity**

**The size of your protein**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Radius of Gyration (nm)</th>
<th>Forward scattering $I_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSA standard</td>
<td>3.07</td>
<td>185 units</td>
</tr>
<tr>
<td>Sample protein</td>
<td>5.58</td>
<td>867 units</td>
</tr>
</tbody>
</table>

$\text{MW}_{\text{protein}} = \frac{66\text{kDa}}{185} \times 867 = 307\text{kDa}$

Thanks to M. Roessle for this slide.
Porod Analysis in Primus gives excluded volume

**Ab-initio Modelling**

A sphere with diameter $D_{\text{max}}$ is filled by densely packed beads of radius $r_0 \ll D_{\text{max}}$. A configuration vector $X$ indicates whether the $j$-th atom belongs to the particle or to the solvent.

Vector of model parameters:

$\text{Position } (j) = x(j) = \begin{cases} 1 & \text{if particle} \\ 0 & \text{if solvent} \end{cases}$

The number of model parameters $M = (D_{\text{max}}/r_0)^3 \approx 10^3$ is too large for conventional minimization methods.

A Monte-Carlo type search starting from a random $X$ can be employed to find a configuration that yields the calculated scattering curve fitting the experimental data.


Thanks to D. Svergun for this slide
**Ab-initio Modelling**

DAMMIN modelling penalties

Using simulated annealing, finds a compact dummy atoms configuration $X$ that fits the scattering data by minimizing

$$f(X) = \chi^2[I_{\text{exp}}(s), I(s, X)] + \alpha P(X)$$

where $\chi$ is the discrepancy between the experimental and calculated curves, $P(X)$ is the penalty to ensure compactness and connectivity, $\alpha > 0$ its weight.

---

**Ab-initio Can it be Trusted?**

Ab initio bead models compared to high resolution X-ray structures
Refinement of Rigid Domains

Rigid Body Refinement: Moving protein sub-parts (called domains) as rigid bodies to fit the scattering data.

Example: Structural changes upon ligand binding

PX-structure with ligand

SAXS Shape obtained by GASBOR - unliganded state

Rigid Body refinement using MASSHA

Adding Missing Linkers

Remodeling of proteins from high resolution fragments/constructs
Program BUNCH (M. Petoukhov; Biophysical J.)

High resolution protein fragments from X-ray crystallography
+ sequence data (TrEMBL/Swissprot)

SAXS Data of constructs
AB
BC
ABC

Model for the entire protein including not resolved linker components
Experimental X-ray scattering of the PYR1 protein in solution in the presence of 1mM (+) ABA.

Scattering curves for possible ensembles were calculated.

Only the curve for ensembles AB/CD produced a good fit to the experimental data ($\chi^2=0.72$)

SAXS demonstrated that the AB ensemble corresponds to the biologically relevant form found under physiological conditions.

**Summary**

**What can we learn from solution SAXS**

- **Model independent parameters:** Size! Rg, Dmax, Volume and MM estimates. Basic shape (Extended or Globular)

- **Behaviour in different buffer conditions:** To assess if interparticle effects or flexibility could be preventing crystallisation (find optimum conditions)

- **Complete high resolution structure known:** Validation of crystal structure in solution under physiological conditions

- **High resolution structure of domains/subunits known:** Quaternary structure using docking/rigid body refinement

- **Incomplete high resolution structure known:** probable configuration of missing portions

- **Nothing known:** ab-initio low resolution shape

- **Dynamic investigations under physiological conditions:** Conformational changes with temperature, pH, binding etc.
Part 2
Data collection at the ESRF Bio-SAXS beamline ID14-3

Solution scattering data collection

Thanks to A. Kikhney for this slide
Experimental Procedure

Clean
Water
Detergent
Water
Dry

Load New
Sample/Buffer

Interlock
Measure

Automation: Sample loading and cleaning
Automation: Sample loading and cleaning

Primary Data Processing
PRIMUS
Quality Control Tests
Multiple time frames used to check for radiation damage!

Thanks to A. Kikhney for this slide

Merging Data

Low and High Concentration

Thanks to A. Kikhney for this slide
Merging Data

Low and High Concentration

Log I(s)

$c = 2 \text{ mg/ml}$

$c = 5 \text{ mg/ml}$

Thanks to A. Kikhney for this slide.
Rg and I(0)
Radius of Gyration and Zero Angle Intensity

Porod Analysis in Primus gives excluded volume
Distance Distribution $P(r)$ Function
Calculated by GNOM give the Dmax and the input file required for Ab-initio modeling

Indirect Fourier Transform!
Distance Distribution \( P(r) \) Function

Calculated by GNOM give the Dmax and the input file required for \textit{Ab-initio} modeling.
Distance Distribution P(r) Function
Calculated by GNOM give the Dmax and the input file required for Ab-initio modeling

Summary
what should be done while at the beamline

• Data collection
• Radial averaging → 1D
• Normalization
• Background subtraction
• Checks for effects of
  – Radiation
  – Concentration
• Log plot
• Guinier plot (Rg, MM)
• Porod plot (volume)
• P(r) plot
• Lo-res 3D model
Automation: Data Processing

Processing Routines Developed at EMBL Hamburg
Results To be Stored in ESRF IspyB Database
(Currently being modified for bio-SAXS)

Automation: Data Processing

| Log started: 2009-05-23 10:47 |

<table>
<thead>
<tr>
<th>Run #</th>
<th>File</th>
<th>Conc. mg/mL</th>
<th>Description</th>
<th>Rg, nm</th>
<th>I(O)</th>
<th>Guinier points</th>
<th>D_max, nm</th>
<th>MM, KDa</th>
<th>Volume, nm$^3$</th>
<th>Quality</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>hsa_003.dat</td>
<td>2.0</td>
<td>hsa</td>
<td>3.05 ± 0.0</td>
<td>111.50</td>
<td>71 - 134 (64)</td>
<td>10.5</td>
<td>66</td>
<td>143</td>
<td>81</td>
<td>12/03</td>
</tr>
<tr>
<td>6</td>
<td>oxg_004.dat</td>
<td>1.0</td>
<td>oxg-2</td>
<td>5.01 ± 0.0</td>
<td>87.87</td>
<td>27 - 71 (40)</td>
<td>16.0</td>
<td>46</td>
<td>225</td>
<td>88</td>
<td>11/03</td>
</tr>
<tr>
<td>8</td>
<td>oxg_008.dat</td>
<td>2.0</td>
<td>oxg-2</td>
<td>5.40 ± 0.0</td>
<td>132.10</td>
<td>21 - 142 (43)</td>
<td>18.0</td>
<td>70</td>
<td>388</td>
<td>91</td>
<td>11/03</td>
</tr>
<tr>
<td>10</td>
<td>oxg_010.dat</td>
<td>6.3</td>
<td>oxg-2</td>
<td>6.64 ± 0.0</td>
<td>265.80</td>
<td>35 - 182 (36)</td>
<td>24.0</td>
<td>140</td>
<td>544</td>
<td>91</td>
<td>12/03</td>
</tr>
<tr>
<td>12</td>
<td>oxg_012.dat</td>
<td>6.3</td>
<td>oxg-2</td>
<td>5.41 ± 0.0</td>
<td>75.43</td>
<td>38 - 64 (27)</td>
<td>130</td>
<td>135</td>
<td>84</td>
<td>12/17</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>oxg_014.dat</td>
<td>1.0</td>
<td>oxg-1</td>
<td>1.98 ± 0.0</td>
<td>21.06</td>
<td>63 - 118 (19)</td>
<td>6.5</td>
<td>11</td>
<td>22</td>
<td>85</td>
<td>12/07</td>
</tr>
<tr>
<td>16</td>
<td>oxg_016.dat</td>
<td>4.0</td>
<td>oxg-6</td>
<td>2.01 ± 0.0</td>
<td>23.28</td>
<td>66 - 217 (152)</td>
<td>7.0</td>
<td>12</td>
<td>19</td>
<td>86</td>
<td>12/01</td>
</tr>
<tr>
<td>18</td>
<td>oxg_018.dat</td>
<td>10.0</td>
<td>oxg-6</td>
<td>2.08 ± 0.0</td>
<td>23.44</td>
<td>88 - 207 (120)</td>
<td>7.3</td>
<td>12</td>
<td>47</td>
<td>79</td>
<td>13/11 Aggregated.</td>
</tr>
<tr>
<td>21</td>
<td>mgp_021.dat</td>
<td>1.0</td>
<td>mgp1</td>
<td>2.61 ± 0.0</td>
<td>202.40</td>
<td>53 - 141 (129)</td>
<td>8.0</td>
<td>107</td>
<td>55</td>
<td>87</td>
<td>12/04</td>
</tr>
<tr>
<td>23</td>
<td>mgp_023.dat</td>
<td>1.0</td>
<td>mgp2</td>
<td>2.57 ± 0.0</td>
<td>197.20</td>
<td>83 - 156 (82)</td>
<td>8.4</td>
<td>104</td>
<td>45</td>
<td>78</td>
<td>14/01</td>
</tr>
<tr>
<td>25</td>
<td>mgp_025.dat</td>
<td>1.0</td>
<td>mgp1</td>
<td>2.48 ± 0.0</td>
<td>110.50</td>
<td>69 - 163 (59)</td>
<td>8.1</td>
<td>56</td>
<td>32</td>
<td>83</td>
<td>14/13</td>
</tr>
<tr>
<td>27</td>
<td>mgp_027.dat</td>
<td>4.0</td>
<td>mgp2</td>
<td>2.39 ± 0.0</td>
<td>95.07</td>
<td>75 - 177 (102)</td>
<td>7.6</td>
<td>50</td>
<td>34</td>
<td>81</td>
<td>14/30</td>
</tr>
<tr>
<td>29</td>
<td>indz_029.dat</td>
<td>1.0</td>
<td>indz</td>
<td>2.73 ± 0.0</td>
<td>334.80</td>
<td>58 - 155 (94)</td>
<td>9.9</td>
<td>177</td>
<td>105</td>
<td>85</td>
<td>14/45</td>
</tr>
<tr>
<td>31</td>
<td>indz_031.dat</td>
<td>1.0</td>
<td>indz</td>
<td>3.24 ± 0.0</td>
<td>132.80</td>
<td>74 - 124 (35)</td>
<td>10.9</td>
<td>70</td>
<td>100</td>
<td>06</td>
<td>14/50</td>
</tr>
<tr>
<td>34</td>
<td>zipkvt_034.dat</td>
<td>1.0</td>
<td>zipkineid</td>
<td>3.37 ± 0.0</td>
<td>139.30</td>
<td>59 - 118 (50)</td>
<td>11.8</td>
<td>74</td>
<td>45</td>
<td>82</td>
<td>12/05</td>
</tr>
<tr>
<td>36</td>
<td>zpk180_036.dat</td>
<td>1.0</td>
<td>zpkineid</td>
<td>3.55 ± 0.0</td>
<td>54.25</td>
<td>50 - 111 (82)</td>
<td>12.0</td>
<td>29</td>
<td>133</td>
<td>68</td>
<td>13/04</td>
</tr>
<tr>
<td>41</td>
<td>zpk250_041.dat</td>
<td>1.0</td>
<td>zpkineid</td>
<td>3.51 ± 0.0</td>
<td>74.51</td>
<td>50 - 112 (46)</td>
<td>11.2</td>
<td>39</td>
<td>45</td>
<td>65</td>
<td>14/40</td>
</tr>
<tr>
<td>43</td>
<td>dpk1a_043.dat</td>
<td>2.0</td>
<td>dpkineid</td>
<td>5.46 ± 0.0</td>
<td>277.50</td>
<td>42 - 62 (32)</td>
<td>17.0</td>
<td>147</td>
<td>306</td>
<td>65</td>
<td>14/35</td>
</tr>
<tr>
<td>45</td>
<td>dpk3a_033.dat</td>
<td>1.0</td>
<td>dpkineid</td>
<td>5.44 ± 0.0</td>
<td>277.50</td>
<td>80 - 162 (36)</td>
<td>12.4</td>
<td>482</td>
<td>180</td>
<td>80</td>
<td>17/10</td>
</tr>
<tr>
<td>47</td>
<td>dpk1tep_047.dat</td>
<td>2.0</td>
<td>dpkineid</td>
<td>3.34 ± 0.0</td>
<td>177.00</td>
<td>54 - 111 (79)</td>
<td>10.9</td>
<td>93</td>
<td>32</td>
<td>49</td>
<td>17/24 Low quality.</td>
</tr>
</tbody>
</table>

Developed at EMBL Hamburg

Image courtesy of A. Kikhney
Part 3
Preparation for Bio-SAXS Data collection

Sample Preparation
In solution SAXS we observe the **Average**

**Monodisperse**

**Mixture**

**Average**
Sample Preparation

If your samples are pure the scattering represents the individual subunit

With this assumption we have one single shape that represents the scattering data and we can calculate models of the low resolution shape of unknown proteins

Contamination will affect the average make the assumption of an individual shape invalid and although you will get models they will be increasingly different from the true shape the more contamination you have.

Sample Preparation

A good experiment requires:

**MONODISPERSE!**

samples in solution

Pure protein (>90%)!
In only 1 oligomeric state!
NO aggregation!
Free from antiparticle effects!

Check before you go to the beamline

MALS/DLS
Analytical ultra centrifugation
Sample Preparation

Know your protein!

In solution samples can degrade, aggregate, react!

Know the best buffer conditions for stability!

Know what storage temperature is required!
Take adequate precautions when shipping samples Extra ice! It may take longer than you expect!

If you can’t purify and store stock solution for shipping you can request access to preparation facilities to do a final purification immediately prior to measurements
Summary

- Bio-SAXS gives:
  - Size! Rg, Dmax, Volume and MM estimates.
  - Basic shape (Extended or Globular)
  - Information on why crystallisation isn't happening (find optimum conditions)
  - Validation of crystal structure in solution under physiological conditions
    - ab-initio low resolution shape
  - Quaternary structure using docking/rigid body refinement
  - Dynamic investigations under physiological conditions

- ESRF Bip-SAXS Beamline is easy to use:
  - High level of automation for
    - Data collection (collaboration between ESRF and EMBL GR and HH outstations)
    - Data processing (ESRF BLISS)
    - Data Analysis (EMBL-Hamburg data processing pipeline)

- Sample preparation and experimental design are important:
  - Monodispersity
  - Interparticle effects

- Information from complimentary techniques is very helpful:
  - MX, NMR, Electron Microscopy

Acknowledgments

ESRF
Sine Larson
Serge Perez
Sean McSweeney
Gordon Leonard
Matthew Bowler
Petra Pernot
Sasha Popov
Daniele de Sanctis
Christoph Müller-Dieckmann
Joanna Timmins
Didier Nurizzo
David Flot
Elspeth Gordon
Debby Davison
Pascal Théveneau
Menhard Menyhert Kocsis
John Surr
Thierry Giraud
Ricardo Nogueira Fernandes
Darren Spruce

And many more people who gave advise and help with testing hardware and software at ID14-3

EMBL

Grenoble
Stephen Cusack
Florent Cipriani
Andrew McCarthy
Sandor Brockhauser
Franck Felisaz
Lukas Födinger
Alexandre Gobbo
Julien Huet
Bernard Lavault
Cyril Villard

Hamburg
Dmitri Svergun
Manfred Roessle
Stefan Fiedler
Uve Ristau
Clement Blanchet
Daniel Franke
Michal Jan Gajda
Christian Gorba

Data Examples
Julia Santiago
Marcello Clerici
Florence Dupeux
Regina Antoni
Sang-Youl Park
Sean R. Cutler
Pedro Luis Rodriguez
José Antonio Márquez
Daniel Panne
Teemu Ikonen
Alexey Kikhney
Haydyn Mertens
Maxim Petoukhov
Weifeng Shang
Aliaksandr Shkumatau
Alexey Zozulya