Data recording, reduction and processing

Manfred Roessle
EMBL Hamburg
Outline

Recording of Small Angle Scattering (SAS) data
   SAXS Variety: Laboratory Sources and Synchrotron based SAXS Beamlines
   What makes SAXS Beamlines special?
   X-rays in and X-rays out: From the source to the detector (via the sample?)
   Seeking and finding: What detects what?

Primary Data Reduction
   1-dim is 1-dim
   Everything has to be normalized!
   Subtracting nothing from something a bit more than nothing

The beauty of SAXS (raw) data
   What the hell does this all mean?
   Initial parameters to judge the success of the experiment

Automated SAXS data recording and analysis
   Let’s have a beer, we are automated!
SAXS Synchrotron based

At all running synchrotron source are SAXS beamlines available!

At all new constructed synchrotrons SAXS beamlines are planned!

All SAXS beamlines are highly oversubscribed and “working horses” on their facilities!

The scientific applications ranging from soft condensed matter, nano-science and fiber diffraction on ordered biological systems up to structural biology in solution.
SAXS Lab sources

Bruker NanoStar
Anton Paar SAXSess
Rigaku S-Max3000
Hecus SAXSeye
The Electromagnetic Spectrum
Radiation from Synchrotron Storage Rings
Production of X-rays

Bending magnet

Necessary part of a synchrotron. Many of these dipole magnets form the synchrotron storage ring. The electrons (positrons) are deflected and accelerated in the magnetic field. This acceleration generates the synchrotron light. The light is emitted tangential to the electron beam.

Wiggler

Stack of magnetic dipoles. Put into the straight sections of the storage ring. Wigglers produces more light than bending magnets in a smaller source size.

Undulator

Most powerfull insertion device! A stack of magnetic dipoles generate a high flux of photons in a very small source size. The specific arrangement of the dipoles \((d=n^*\lambda)\) produces a discrete spectrum with coherent properties.
Radiation from Synchrotron Storage Rings

Dipole bending magnet (APS)
Properties of X-rays from Synchrotron Sources

1. High flux of photons
2. Tangentially emitted with a central cone such like a spotlight
3. Polychromatic light from infrared to X-rays

X-ray optics for:
1. Beam defining
2. Focusing
3. Monochromatization

High heat load and radiation damage!
Beamsize increases with distance from source point!
Scattering/Diffraction experiments need monochromatic light!
X-ray optics
Beam defining

Beam defining by slit pairs:

Problem:
Every edge cutting the beam produces parasitic scattering! Refractive streaks at low angles impedes SAXS experiments!

Solution:
Successive slit systems. First slits cutting the beam and second system cuts out the undesired parasitic scattering.

For slits exposed to high head load, cooling is necessary.
X-ray optics

Focusing

The glancing angle for X-rays on surfaces are very small. Only under grazing conditions X-ray mirrors can be used for focusing.

Bending of the highly polished mirror surface permits focusing on the parabolic mirror profile.

In order to use the full beam size X-ray mirrors on synchrotrons are typically in the range of 30cm to 1m. The grazing angle can be increased by using high Z-elements (e.g. Rhodium, Platin, Nickel) as reflecting surface. These mirrors are acting as well as high energy filter; important for monochromatization.
X-ray Optics
Monochromator

Monochromatization of the X-ray is achieved by using single crystals in Bragg diffraction geometry

\[ n\lambda = 2d \sin(\Theta) \]

For instance a Si 111 crystal with d-spacing (=distance between the crystal lattice) of d=3.14 Å deflects the beam for a wavelength of \( \lambda = 1 \) Å to \( \Theta \sim 10° \) (9.16°).

Problem:
The Bragg condition is also fulfilled for multiple orders of the wavelength \( n \). These higher order have to be filtered by the X-ray mirror.
X-ray optics

X33 mirror
1m Rh coated on Zerodur substrate

X33 monochromator
Si 111 single crystal in asymmetric cut
This device is also used for horizontal focusing.
Detectors for SAXS
Seeking and finding

Fiber optic tapered CCD cameras for X-rays.
+ fast readout (10Hz framing possible)
+ large area possible

- low dynamic range
- sensitive to overexposing
- spatial correction due to the fiber optic taper
- intrinsic background of CCD chip
Detectors for SAXS
Seeking and finding

Novel pixel detectors

PILATUS system Swiss light source

+ high dynamic range ($10^{20}$ phot/s!!!)
+ no intrinsic background
+ fast framing

Ideal detector for solution scattering!

500k model was installed successfully at X33 and upgraded to a 1M prototype
X33-Beamline Schematic

Not to scale

Experimental Hutch

Optics Hutch

PS – Primary Slit

SS – Secondary Slits

BM – Beam Monitor

EMBO Workshop on solution scattering EMBL Hamburg 25.10 to 01.11.2010
Schematic X33 SAXS setup

Beamstop with diode for measurement of transmitted beam

Beamshutter with diode for measurement of incident beam (prior to exposure)

Pilatus WAXS detector up to 13°
q ~ 10 nm\(^{-1}\) d ~ 6 Å

Sample cell

s: 0.1 nm\(^{-1}\) to 4.3 nm\(^{-1}\)
d: 65 nm to 15 nm

1400 mm

1000 mm
EMBL’s X33 Beamline
Data Reduction
From 2-dim to 1-dim

Radial (azimuthal) averaging:
The 2-dim detector intensities are stored as a images with three values:
Intensity counts; pixel X and pixel Y
I(x,y)

The x=0 and y=0 position can be determined by the concentric diffraction cycles of Silver Behenate powder.
Radial Averaging

Normalization against:
- data collection time,
- concentration,
- transmitted sample intensity.
Data Reduction
Assigning the s-axis

\[ s = \frac{2\pi}{\lambda} \sin(2\Theta) = \frac{1}{d} \]

<table>
<thead>
<tr>
<th>Peak order</th>
<th>d-spacing</th>
<th>s-value</th>
<th>d-Channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st order</td>
<td>5.834 nm</td>
<td>1.076 nm(^{-1})</td>
<td>495</td>
</tr>
<tr>
<td>2nd order</td>
<td>2.917 nm</td>
<td>2.153 nm(^{-1})</td>
<td>995</td>
</tr>
<tr>
<td>3rd order</td>
<td>1.944 nm</td>
<td>3.231 nm(^{-1})</td>
<td>1400</td>
</tr>
<tr>
<td>4th order</td>
<td>1.458 nm</td>
<td>4.304 nm(^{-1})</td>
<td>1990</td>
</tr>
</tbody>
</table>
Background subtraction

Standard protocol:
1. Measurement: Buffer
2. Measurement: Protein
3. Measurement: Buffer

We are looking for protein signals of less than 0.5% above the background level!
Primary Data Analysis
Judging the data quality

First experiment
BSA standard solution

This standard is used for calibration and has to be freshly prepared and measured.
First test
Quality check on intensity data

Ideal solution of particles
Repulsive particle interactions
Attractive particle interactions
Primary Data Analysis
The Guinier Fit

Comparison of two protein samples to the BSA standard.

The difference at low angles $< 1 \text{ nm}^{-1}$ because of the different molecular weights (MW).

The scattering at low angles is proportional to the product of $c_{\text{protein}} \cdot \text{MW}$.

At known protein concentration and a proper BSA standard the molecular weight of the sample can be estimated using the Guinier-Fit.
Primary Data Analysis
The Guinier Fit

\[ I(s) \approx I_0 e^{-\frac{R_g^2}{3}s^2} \]
\[ \ln I(s) \approx \ln I_0 - \frac{R_g^2}{3}s^2 \]

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

\[ MW_{\text{protein}} = \frac{66kDa}{185} \cdot 867 \]
\[ MW_{\text{protein}} = 307kDa \]
Protein folding state

Kratky plot

The Kratky plot is typically used to analyze the conformation of proteins, but can be used to analyze the random walk model of polymers.

A Kratky plot can be made by plotting:

$$I(s) s^2$$ versus s.
Porod Volume

Porod approximation:

\[ I(s) = K s^{-4} \]

\[ \frac{K}{Q} \propto \frac{S}{V} \]

\[ Q = \int_{0}^{\infty} s^2 I(s) \, ds \]

\[ Q = \text{Porod invariant} \]

\[ \frac{S}{V} = \text{shape to volume ratio} \]

Rule of thumb: Porod volume is approximately two times molecular weight

Here BSA: MW 66 kDa; Porod volume 115
Automated SAXS data treatment “Pipeline”
BioSAXS sample changer evaluation setup

Integrated in experiment control and analysis pipeline

Joint project with EMBL Grenoble (project leader and construction of the device) and ESRF

In Hamburg:
Fully automated operation implemented and integrated in the analysis pipeline

Filling faster than with the Fraunhofer changer

In user operation since 2010

35 µl sample volume (before 80 µl)
Automation Results: what the users see

<table>
<thead>
<tr>
<th>Run #</th>
<th>File</th>
<th>Conc., mg/ml</th>
<th>Description</th>
<th>R_g, nm</th>
<th>I(0)</th>
<th>Guinier points</th>
<th>D_max, nm</th>
<th>MM, kDa</th>
<th>Volume, nm³</th>
<th>Quality, %</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>bsa_003.dat</td>
<td>2.0</td>
<td>bsa</td>
<td>3.05±0%</td>
<td>111.50</td>
<td>71 - 134 (64)</td>
<td>10.5</td>
<td>59</td>
<td>143</td>
<td>81</td>
<td>11:23</td>
</tr>
<tr>
<td>6</td>
<td>cnq_006.dat</td>
<td>1.0</td>
<td>cnq 2-1 1</td>
<td>5.01±0%</td>
<td>87.87</td>
<td>27 - 71 (43)</td>
<td>16.0</td>
<td>46</td>
<td>225</td>
<td>88</td>
<td>11:33</td>
</tr>
<tr>
<td>8</td>
<td>cnq_008.dat</td>
<td>2.0</td>
<td>cnq 2-1 2</td>
<td>5.40±0%</td>
<td>132.10</td>
<td>21 - 62 (42)</td>
<td>18.9</td>
<td>70</td>
<td>388</td>
<td>91</td>
<td>11:40</td>
</tr>
<tr>
<td>10</td>
<td>cnq_010.dat</td>
<td>6.3</td>
<td>cnq 2-1 6.3</td>
<td>6.64±0%</td>
<td>265.80</td>
<td>15 - 45 (31)</td>
<td>21.2</td>
<td>140</td>
<td>544</td>
<td>91</td>
<td>12:03</td>
</tr>
<tr>
<td>12</td>
<td>cnq_012.dat</td>
<td>2.4</td>
<td>cnq 4-3 2.4</td>
<td>5.41±0%</td>
<td>75.43</td>
<td>38 - 64 (27)</td>
<td>18.0</td>
<td>135</td>
<td>84</td>
<td>86</td>
<td>12:17</td>
</tr>
<tr>
<td>14</td>
<td>cnq_014.dat</td>
<td>1.0</td>
<td>cnq 6-1 1</td>
<td>1.98±0%</td>
<td>21.06</td>
<td>63 - 210 (152)</td>
<td>6.5</td>
<td>11</td>
<td>22</td>
<td>86</td>
<td>12:37</td>
</tr>
<tr>
<td>15</td>
<td>cnq_016.dat</td>
<td>4.0</td>
<td>cnq 6-1 4</td>
<td>2.01±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>12:51</td>
</tr>
<tr>
<td>18</td>
<td>cnq_018.dat</td>
<td>10.0</td>
<td>cnq 6-1 10</td>
<td>2.08±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</td>
</tr>
<tr>
<td>21</td>
<td>mg1_021.dat</td>
<td>1.0</td>
<td>mg1</td>
<td>2.61±0%</td>
<td>202.40</td>
<td>53 - 161 (105)</td>
<td>8.5</td>
<td>107</td>
<td>85</td>
<td>87</td>
<td>13:44</td>
</tr>
<tr>
<td>23</td>
<td>mg2_023.dat</td>
<td>1.0</td>
<td>mg2</td>
<td>2.57±0%</td>
<td>197.20</td>
<td>83 - 164 (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>mgp1_025.dat</td>
<td>1.0</td>
<td>mgp1</td>
<td>2.48±0%</td>
<td>110.50</td>
<td>69 - 163 (95)</td>
<td>8.1</td>
<td>58</td>
<td>32</td>
<td>83</td>
<td>14:15</td>
</tr>
<tr>
<td>27</td>
<td>mgp2_027.dat</td>
<td>4.0</td>
<td>mgp2</td>
<td>2.39±0%</td>
<td>95.07</td>
<td>75 - 177 (103)</td>
<td>7.6</td>
<td>50</td>
<td>34</td>
<td>81</td>
<td>14:50</td>
</tr>
<tr>
<td>29</td>
<td>ind1_029.dat</td>
<td>1.0</td>
<td>indica</td>
<td>2.73±0%</td>
<td>334.80</td>
<td>58 - 151 (94)</td>
<td>9.3</td>
<td>177</td>
<td>105</td>
<td>85</td>
<td>14:45</td>
</tr>
<tr>
<td>31</td>
<td>ind2_031.dat</td>
<td>1.0</td>
<td>indica</td>
<td>3.24±0%</td>
<td>132.80</td>
<td>74 - 124 (51)</td>
<td>10.9</td>
<td>70</td>
<td>106</td>
<td>56</td>
<td>14:59</td>
</tr>
<tr>
<td>34</td>
<td>zipkw_034.dat</td>
<td>1.0</td>
<td>zip kinase wild</td>
<td>3.37±0%</td>
<td>139.30</td>
<td>59 - 118 (60)</td>
<td>11.8</td>
<td>74</td>
<td>45</td>
<td>82</td>
<td>15:29</td>
</tr>
<tr>
<td>36</td>
<td>zipk180_036.dat</td>
<td>1.0</td>
<td>zip kinase mutant</td>
<td>3.56±0%</td>
<td>54.25</td>
<td>50 - 111 (62)</td>
<td>12.5</td>
<td>29</td>
<td>133</td>
<td>68</td>
<td>15:44</td>
</tr>
<tr>
<td>41</td>
<td>zipk265_041.dat</td>
<td>1.0</td>
<td>zip kinase mutant</td>
<td>3.51±0%</td>
<td>74.51</td>
<td>50 - 113 (64)</td>
<td>11.2</td>
<td>39</td>
<td>45</td>
<td>65</td>
<td>16:00</td>
</tr>
<tr>
<td>43</td>
<td>drp1k_043.dat</td>
<td>2.0</td>
<td>drp kinase full</td>
<td>5.46±0%</td>
<td>277.50</td>
<td>42 - 63 (22)</td>
<td>17.9</td>
<td>147</td>
<td>306</td>
<td>65</td>
<td>16:53</td>
</tr>
<tr>
<td>45</td>
<td>drp1k2_045.dat</td>
<td>2.0</td>
<td>drp kinase full</td>
<td>3.54±0%</td>
<td>912.90</td>
<td>60 - 98 (33)</td>
<td>12.4</td>
<td>482</td>
<td>180</td>
<td>80</td>
<td>17:10</td>
</tr>
<tr>
<td>47</td>
<td>drp1atp_047.dat</td>
<td>2.0</td>
<td>drp 1 kinase</td>
<td>3.34±0%</td>
<td>177.00</td>
<td>34 - 111 (78)</td>
<td>10.9</td>
<td>93</td>
<td>32</td>
<td>49</td>
<td>17:24 Low quality.</td>
</tr>
</tbody>
</table>
Automated Data Analysis: AutoRg
Estimation of radius of gyration \( R_g \) and molecular weight

1. Automated (no help from a user)
2. Implemented into processing pipeline
3. Input used for Autognom

Solution scattering data

Program autorg.exe

\( R_g \) – radius of gyration
\( \sigma \) – accuracy
\( I_0 \) – zero angle intensity

Used range

Quality

Estimation of $D_{\text{max}}$ with AUTOGNOM

$D_{\text{max}}$ underestimated  

$D_{\text{max}}$ overestimated  

$D_{\text{max}}$ ok.
Automation
Easy going and money is for nothing?

Ab initio model with closed shutter!!
Conclusions

1. SAXS beamlines are special equipped for making excellent solution scattering experiments

2. Trust the beamline automation, it is working!

3. Don’t trust the beamline automation – if you find something strange going on

4. Validate your results!

5. Ask the beamline staff in the case of troubles – they will help you