The First NIH Workshop on Small Angle X-ray Scattering and Application in Biomolecular Studies

Open Remarks: ***Ad Bax (NIDDK, NIH)***
Introduction: ***Yun-Xing Wang (NCI-Frederick, NIH)***
Lectures:
Xiaobing Zuo, *Ph.D. (NCI-Frederick, NIH)*
Alexander Grishaev, *Ph.D. (NIDDK, NIH)*
Jinbu Wang, Ph.D. *(NCI-Frederick, NIH)*
Organizer: Yun-Xing Wang *(NCI-Frederick, NIH)*

Place: NCI-Frederick campus
Time and Date: 8:30am-5:00pm, Oct. 22, 2009

Suggested reading

Books:

Glatter, O., Kratky, O. (1982) Small angle X-ray Scattering. Academic Press.
Feigin, L., Svergun, D. (1987) Structure Analysis by Small-angle X-ray and Neutron Scattering. Plenum Press.

Review Articles:

Svergun, D., Koch, M. (2003) Small-angle scattering studies of biological macromolecules in solution. Rep. Prog. Phys. 66, 1735-1782.
Koch, M., et al. (2003) Small-angle scattering : a view on the properties, structures and structural changes of biological macromolecules in solution. Quart. Rev. Biophys. 36, 147-227.
Putnam, D., et al. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Quart. Rev. Biophys. 40, 191-285.

Software

Primus: 1D SAS data processing
Gnom: Fourier transform of the I(q) data to the P(r) profiles, desmearing
Crysol, Cryson: fits of the SAXS and SANS data to atomic coordinates
EOM: fit of the ensemble of structural models to SAXS data for disordered/flexible proteins
Dammin, **Gasbor**: *Ab initio* low-resolution structure reconstruction for SAXS/SANS data
All can be obtained from
http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html
MarDetector: 2D image processing
Igor: 1D scattering data processing and manipulation
SolX: scattering profile calculation from atomic coordinates
Xplor/CNS: high-resolution structure refinement
GASR: http://ccr.cancer.gov/staff/links.asp?profileid=5546

Part One

Solution Small Angle X-ray Scattering: Basic Principles and Experimental Aspects

Xiaobing Zuo (NCI-Frederick)

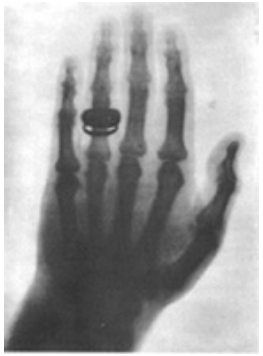
Alex Grishaev (NIDDK)

1. General Aspects

- History
- Study scope
- X-ray scattering vs neutron and light scattering

History of X-ray and scattering

- X-ray discovered in 1895

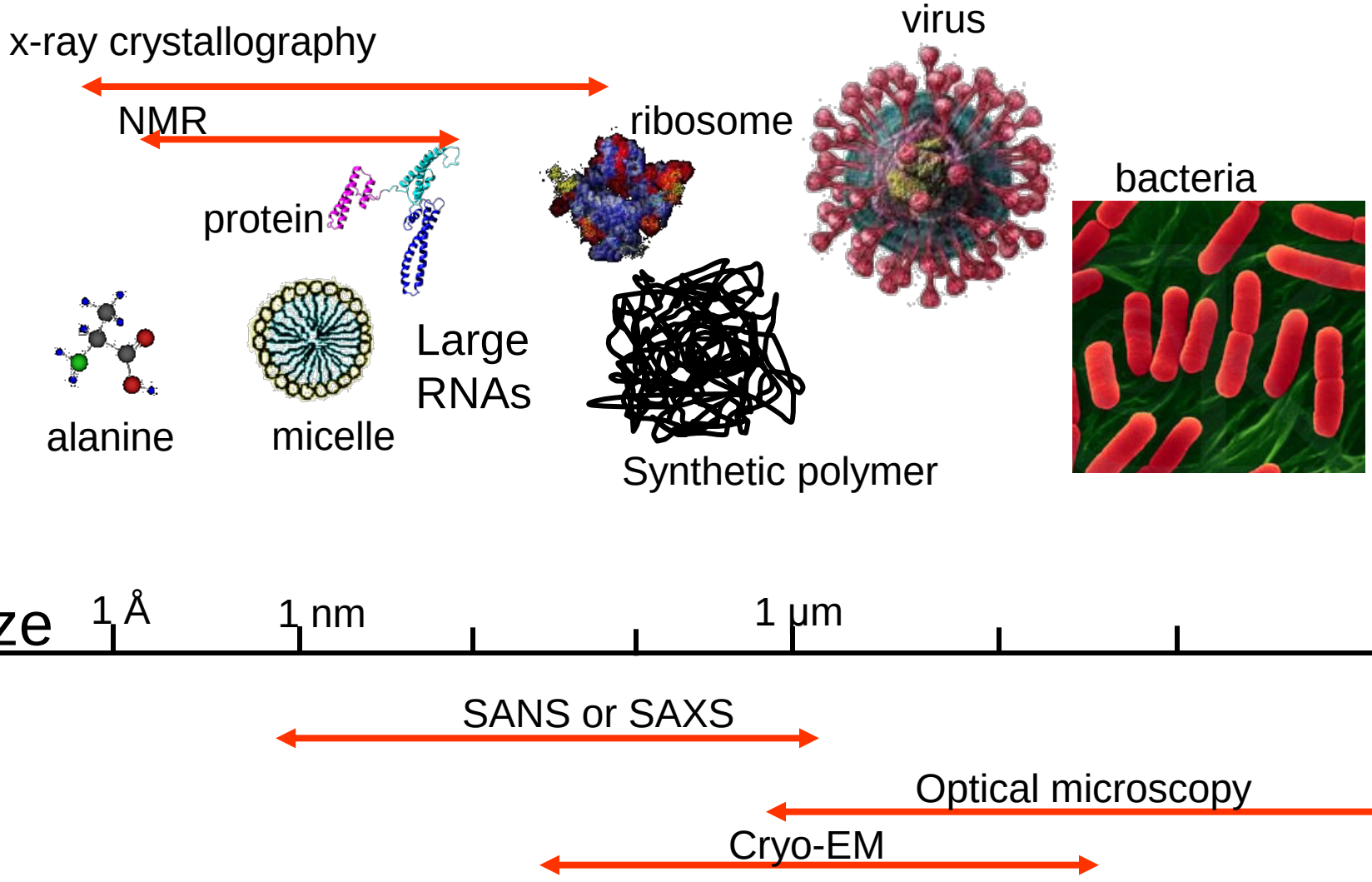


Wilhelm Conrad Röntgen 1845-1923

- First solution X-ray scattering performed in late 1930s
  - Measure molecular weight, size

- For a long time, applications had been limited by available X-ray source, detector and data analysis methods

Scales of various methods



The scope of small angle X-ray scattering in terms of spatial dimension covers ~1nm to ~1μm ranges, perfectly suitable for biomolecular structrual study.

X-ray scattering vs. neutron, light scattering

- Solution scattering techniques
  - Small-angle X-ray scattering (SAXS)
  - Small-angle neutron scattering (SANS)
  - Small-angle light scattering (LS)

- Interaction with objects
  - X-ray and light intact with electrons
  - Neutron scattering: nuclear spins

- Wavelength / probe scale
  - Light scattering: λ: 2000 – 8000 Å
  - Neutron scattering: 5 – 20 Å
  - X-ray scattering: λ: 0.5 – 2.0 Å

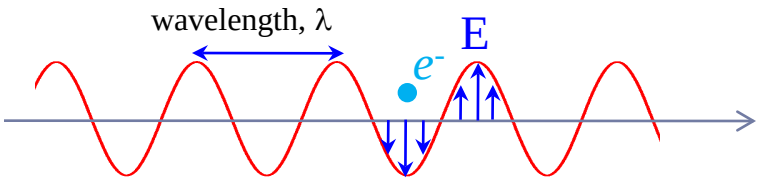
- Provided information
  - Sizes,
  - Shapes
  - Low resolution structure
  - higher resolution for shorter wavelength

2. Physics of X-ray scattering

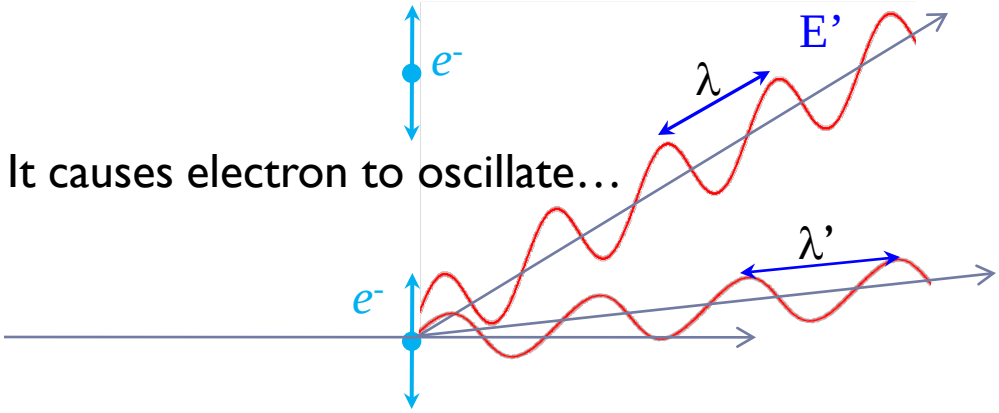
- Scattering phenomenon and interference
- From crystal and fiber diffraction to solution scattering
- X-ray contrasting
- Form factor and object shapes
- Theory on solution scattering calculations

Scattering phenomenon and interference

X-ray is traveling wave…



X-ray electric field hits electron…



It causes electron to oscillate…

And make it radiate secondary X-ray!
This is scattering!

X-ray Wave function:

$$E(t) = E_0 e^{i(2\pi\nu t + \phi_0)}$$

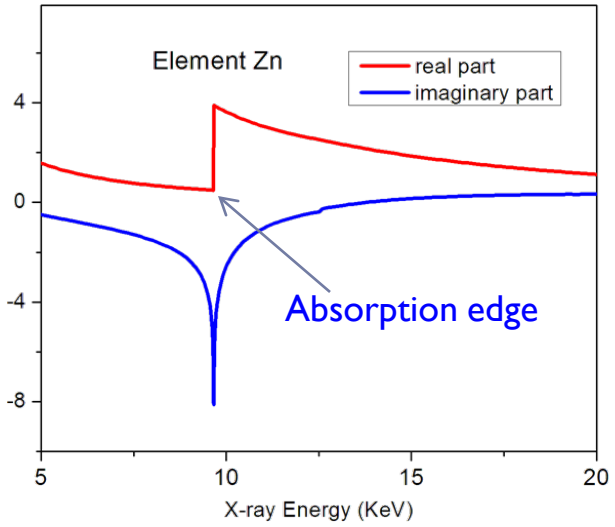
frequency: $\nu = c/\lambda$
Initial phase: ϕ_0
Phase change: $\Delta\phi = 2\pi d/\lambda$ (radian)
after travelling distance d.

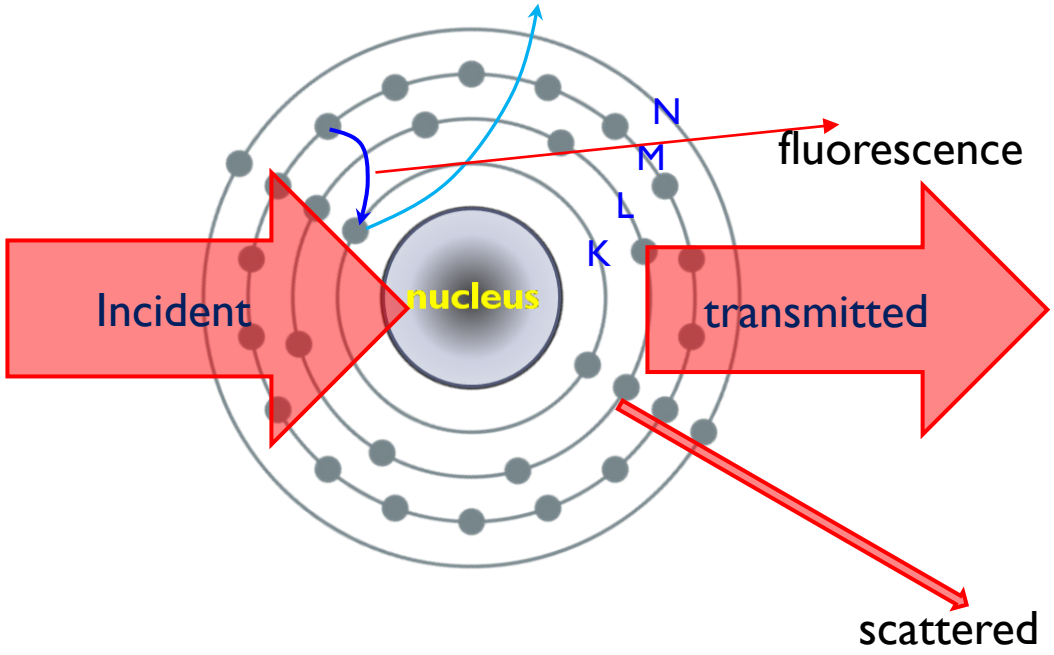
Elastic scattering: λ → λ not change
Inelastic (Compton) scattering: λ → λ' changed

If the scattered X-ray has the same wavelength of the original X-ray, it is called elastic/coherent scattering (no energy loss). If the wavelength of the scattering is changed (longer), it is inelastic/incoherent/Compton scattering (there is an energy loss)

X-ray interacts with atoms

Inelastic scattering is very weak and often ignorable for bound electrons, for example electron in atoms. Besides elastic and inelastic scattering, atoms absorb X-ray and emit fluorescence X-ray, especially at the absorption edges.





- Fluorescence X-ray has lower energy or longer wavelength due to energy loss.
- Both Inelastic scattering and fluorescence X-ray only add background to scattering data because they don't interfere with elastically scattering x-ray due to different wavelength!
- However, stay away the absorption edges when choose X-ray radiation energy for scattering to avoid problematic fluorescence X-ray background.

X-ray measures distance/structure by interference

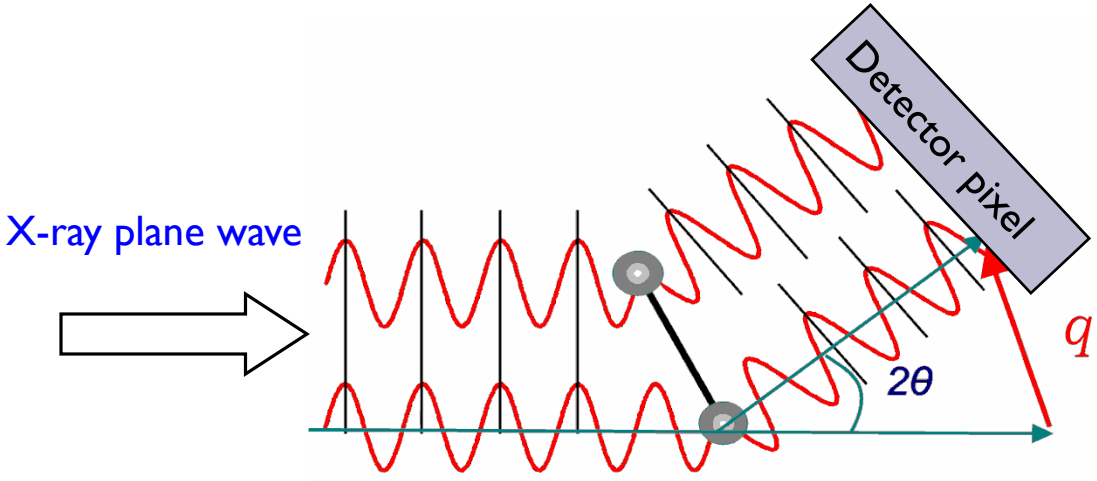
- Interference in pond:

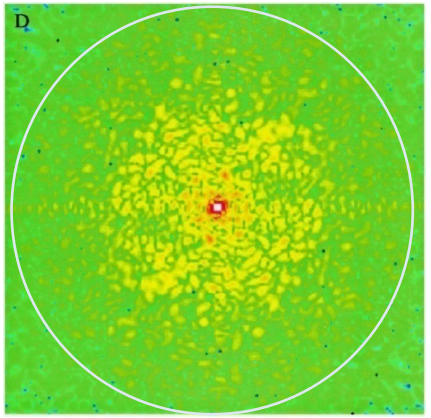
  Interference pattern of water waves codes distance between sources

- Elastic scattering Interference:

  Elastic X-ray scattering interferes.

  Before hit the scatterers, the X-ray plane waves travel with same phase (in phase). When hit the scatterers, X-ray waves are scattered. When two scattered X-ray waves arrive at a detector pixel with same phase (they are in phase), they enhance each other (constructive interference). If they arrive a pixel with opposite phases (out of phase), they cancel each other (deconstructive interference). The interference pattern on detector encodes the distances among scatterers.

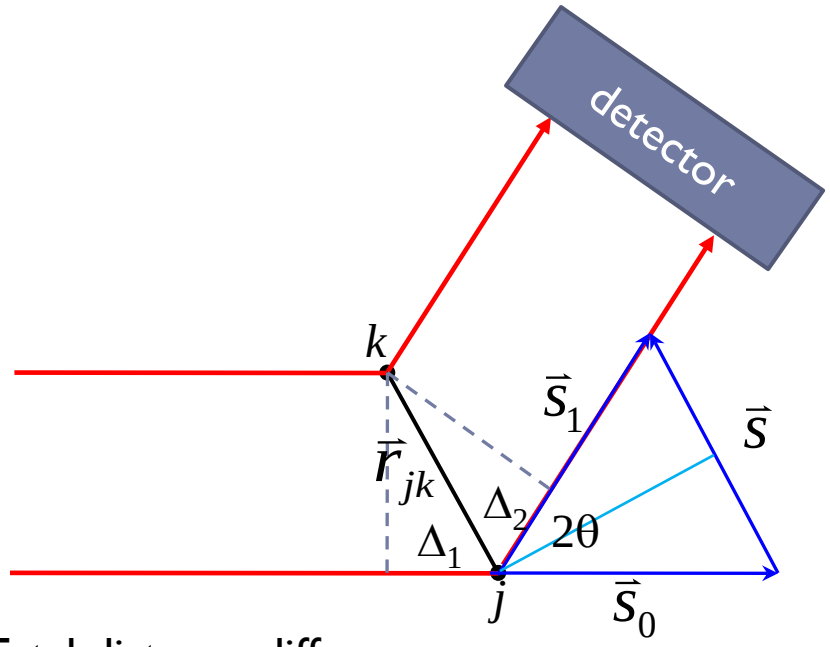




X-ray wave Interference pattern

Phase change and Momentum transfer

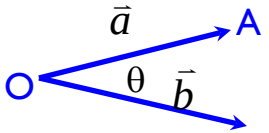
The phase difference between two scattered X-ray beams is determined by the difference of distance they traveled before arrive the detector pixel.



A vector $\vec{a}$ has its length $a = |\vec{a}|$ and direction (O→A).

Vector dot product:

$$\vec{a} \bullet \vec{b} = a * b \cos(\theta)$$



unit incident wavevector $\vec{s}_0$

unit scattering wavevector $\vec{s}_1$

wave vector change: $\vec{s}$

$$\vec{s} = \vec{s}_1 - \vec{s}_0$$

$$|\vec{s}| = 2\sin\theta$$

Total distance difference:

$$\Delta = \Delta_1 + \Delta_2 = -\vec{s}_0 \bullet \vec{r}_{jk} + \boxed{\vec{s}_1 \bullet \vec{r}_{jk}}$$

Projection of $\vec{r}_{jk}$ on $\vec{s}_1$

$$= (\vec{s}_1 - \vec{s}_0) \bullet \vec{r}_{jk} = \vec{s} \bullet \vec{r}_{jk}$$

Definition: momentum transfer

$$\vec{q} = (2\pi/\lambda)\vec{s}$$

$$q = |\vec{q}| = (4\pi/\lambda)\sin\theta$$

Phase change: (radian)

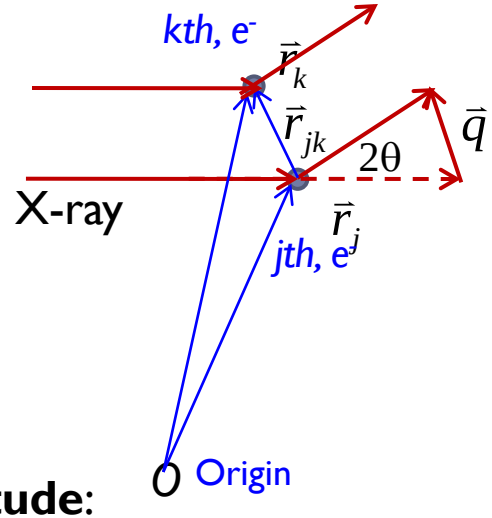
$$\Delta\phi = 2\pi * \frac{\Delta}{\lambda} = \vec{q} \bullet \vec{r}_{jk}$$

Scattering by an electron pair

Amplitude at q of scattered X-ray by electron j at position $\vec{r}_j$:

Wave function $E'(t) = E_0' e^{i(2\pi\nu t+\phi_0)} e^{i\vec{q}\bullet\vec{r}_j} \Longrightarrow A_j(\vec{q}) = f_e \exp(i\vec{q}\bullet\vec{r}_j)$

additional phase change



Total amplitude at q of scattered X-ray by electrons j & k:

$$A_{tot}(\vec{q}) = A_j(\vec{q}) + A_k(\vec{q}) = f_e \exp(i\vec{q}\bullet\vec{r}_j) + f_e \exp(i\vec{q}\bullet\vec{r}_k)$$
$$= f_e \exp(i\vec{q}\bullet\vec{r}_j)\{1+\exp[i\vec{q}\bullet(\vec{r}_k-\vec{r}_j)]\}$$
$$= f_e \exp(i\vec{q}\bullet\vec{r}_j)\{1+\underbrace{\cos(\vec{q}\bullet\vec{r}_{jk})}_{\text{real}}+\underbrace{i\sin(\vec{q}\bullet\vec{r}_{jk})}_{\text{imaginary}}\}$$

Intensity registered on detector is the **square of the amplitude**:

$$I(\vec{q}) = \left|A_{tot}(\vec{q})\right|^2 = A_{tot}(\vec{q})A_{tot}^*(\vec{q})$$
$$= f_e^2\left|\exp(i\vec{q}\bullet\vec{r}_j)\right|^2 \times \left|1+\cos(\vec{q}\bullet\vec{r}_{jk})+i\sin(\vec{q}\bullet\vec{r}_{jk})\right|^2$$
$$= f_e^2 \times 1 \times \left[\underbrace{1+1}_{\text{from individuals}}+\underbrace{2\cos(\vec{q}\bullet\vec{r}_{jk})}_{\text{cross/ interference term}}\right]$$

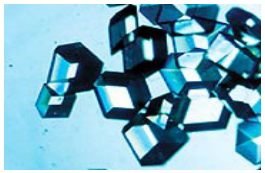
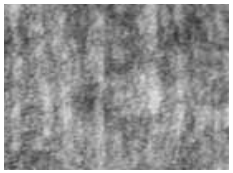
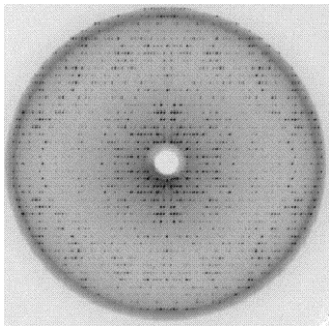
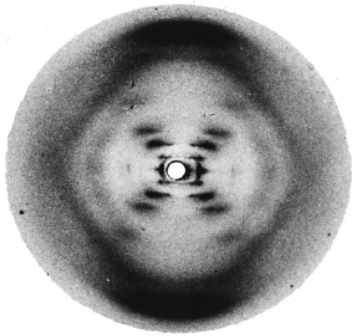
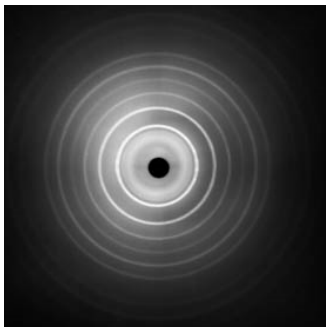
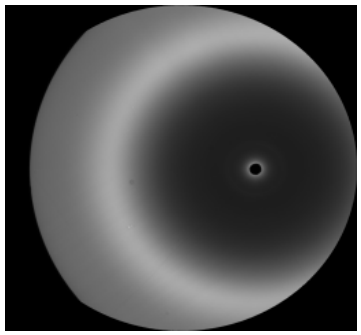
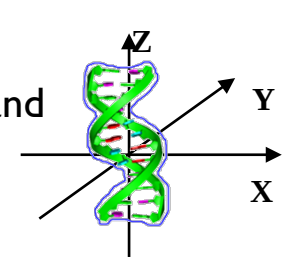
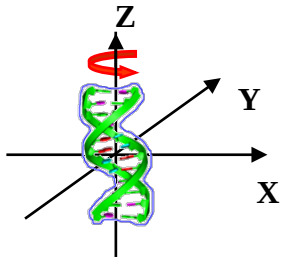
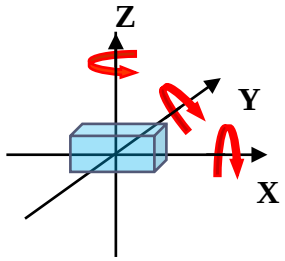
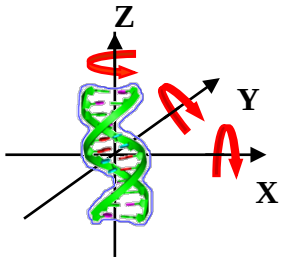
- Scattering intensity does not remember/care where the Origin we choose, but is a function of scatterer pair distance(s), which is the structural information.
- Loss of phase information during measuring
- The interference pattern comprises the contribution from individual electrons (scatterers), and more importantly the cross term from each and every scatterer pair

In principle, interference pattern of an object with N e-:

$$I(\vec{q}) = \left|\sum_{j=1}^{N} A_j(\vec{q})\right|^2 = \left|\sum_{j=1}^{N} f_e \exp(i\vec{q}\bullet\vec{r}_j)\right|^2$$

From crystal and fiber diffraction to solution scattering

Interference patterns of objects vary along with the samples' nature, including the symmetry of matrix of molecules embedded and the freedom of molecules in the matrix.

	Single Crystal	Fiber/ Membrane	Powder/ Micro-crystals	Solution
Sample states				
Interference pattern				
matrix symmetry and molecular freedom				

Matrix symmetry: high ← low

Molecular freedom: low → high

X-ray scattering form factor/scattering length

$$I(\vec{q}) = \left| \sum_{j=1}^{N} A_j(\vec{q}) \right|^2 = \left| \sum_{j=1}^{N} f_e \exp(i\vec{q} \bullet \vec{r}_j) \right|^2$$

- In reality, it is impossible to find the positions of each electrons in an object, then calculate the scattering / interference pattern. We can divide the object into small relative independent units (e.g. atoms) and calculate the pattern from those small units.
- X-ray scattering length/form factor represent the capability of an object scatters X-ray. The scattering length of an electron is f_e=2.8179x10-13cm^{-1}. For the sake of simplicity, f_e set as a unit: $f_e \equiv 1$.

For an object with an ensemble of N electrons:

Define: $F(\vec{q}) = \sum_{j=1}^{N} f_e \exp(i\vec{q} \bullet \vec{r}_j)$

$$I(\vec{q}) = \left| \sum_{j} A_j(\vec{q}) \right|^2 = \left| \sum_{j} f_e \exp(i\vec{q} \bullet \vec{r}_j) \right|^2 \equiv F^2(\vec{q})$$

Scattering form factor/length

$$\left| \frac{F(\vec{q})}{f_e} \right|^2 = \left| \frac{I(\vec{q})}{I_e} \right|$$

If a scatterer can take all the orientation:

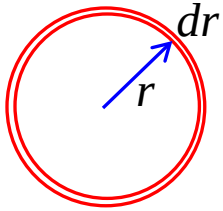
$$I(q) = \left\langle I(\vec{q}) \right\rangle_\Omega \qquad \left\langle \exp(i\vec{q} \bullet \vec{r}_j) \right\rangle_\Omega = \frac{\sin(qr)}{qr}$$

average over all orientation

Loss of the direction the momentum transfer/ angular term

For an object with radialy uniform electron density $\rho(r)$:

$$F(q) = \sum_j \underbrace{(4\pi r^2 dr)\rho(r)}_{\text{Total electrons between two spherical surfaces}} \left\langle \exp(i\vec{q} \bullet \vec{r}_j) \right\rangle_\Omega = 4\pi \int \rho(r) r^2 \frac{\sin(qr)}{qr} dr$$



form factor for sphere

The form factors of some objects with simple shapes have analytical formula expression, for example, sphere. Sphere is a widely used model in characterizing the size or size distribution of globular particles in structural biology and nanoscale material science.

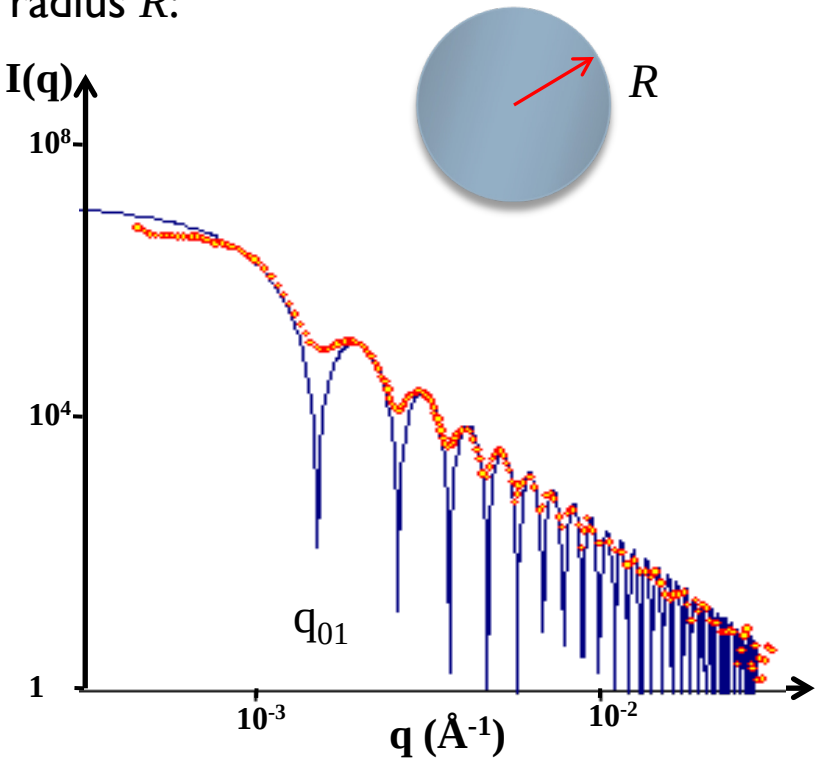
Sphere with homogenous electron density and radius R:

$$F(q) = \frac{3(\sin(qR) - qR\cos(qR))}{(qR)^3}$$

$$I(q) = F^2(q) = \left(\frac{3(\sin(qR) - qR\cos(qR))}{(qR)^3} \right)^2$$

From the scattering curve, we can estimate the radius of the sphere:

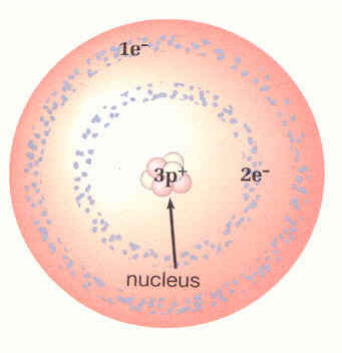
$$R \approx \frac{4.493}{q_{01}}$$



- Scattering profile of silica spheres (red) and simulation based on perfect sphere (blue)
- Discrepancy of silica scattering from sphere model due to size polydispersity, imperfect spherical shape, etc.

Figure modified from: Lambard, Lessieur and Zemb, J. Phys. I France, 1992, (2) 1191

Atomic form factor



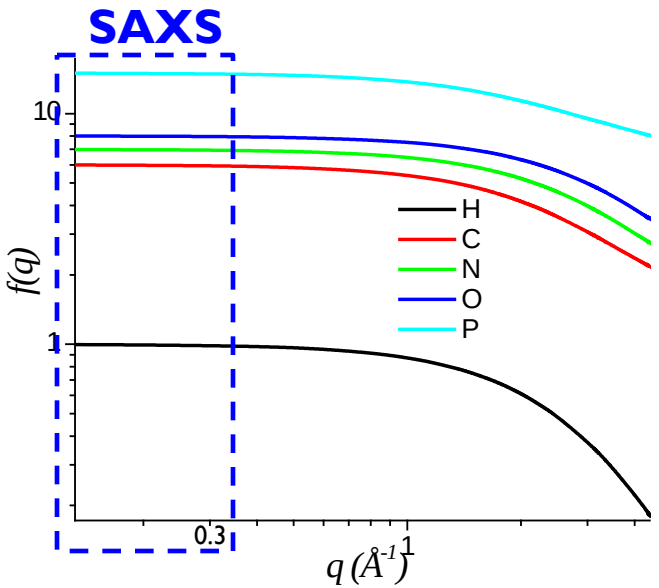
Atomic electron cloud

Atomic form factors are fundamental parameters in X-ray techniques.

Electron cloud in atoms has radial density distribution $\rho(r)$

$$f(q) = 4\pi \int \rho(r) r^2 \frac{\sin(qr)}{qr} dr$$

ρ(r) were obtained from quantum chemical calculations. Atomic form factors for all elements and important ions were tabulated in International Tables for Crystallography and other handbooks.



Data taken from International Tables for Crystallography, Vol. C, Table 6.1.1.1

- $f(0)=Z$: the total electron of the atom.
- Atoms with higher Z will scatter stronger.
- $f(q)$ decreases slowly along q in region (SAXS) of q close to 0
- In low-resolution model reconstruction from SAXS data using ***f(q)*=const**

Solution X-ray scattering measures the contrast / electron density difference

In vacuum, x-ray scattering directly measures Z number.

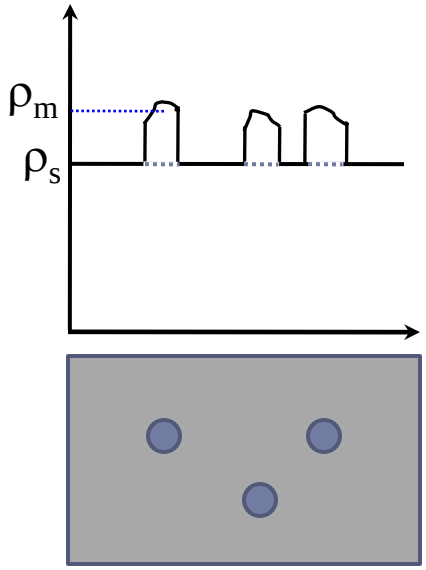
Solution sample scattering:

$$I_{molecule} = I_{solution} - I_{solvent}$$

What X-ray scattering measures:

$$\Delta\rho(\vec{r}) = \rho_m(\vec{r}) - \rho_s(\vec{r})$$

excess electron density/scattering length against solvent/buffer

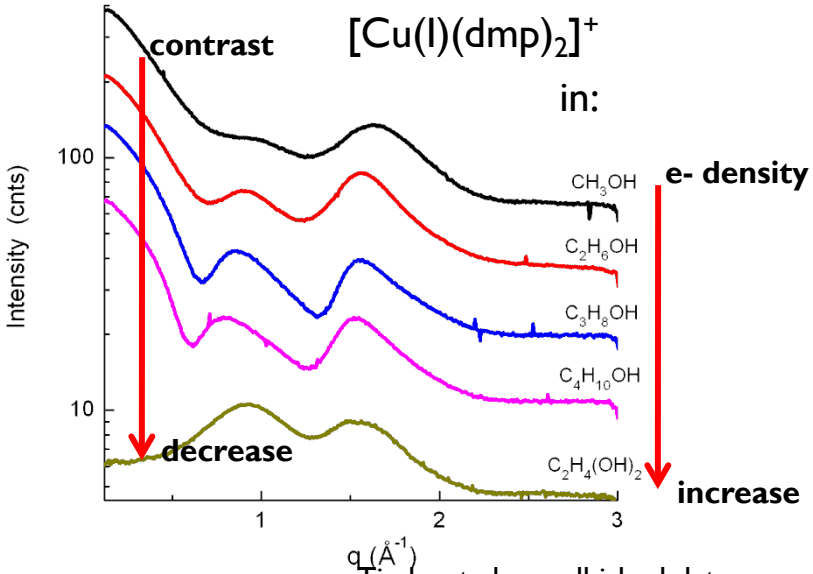


In vacuum: $A(\vec{r}) = \rho_m(\vec{r})e^{i\vec{q}\bullet\vec{r}}$

In solution: $A(\vec{r}) = \Delta\rho(\vec{r})e^{i\vec{q}\bullet\vec{r}}$

Solution X-ray scattering contrast matching:

Increasing solvent electron density, X-ray scattering contrast match occurs at average spatial scale [I(q~0) → 0], but electron density difference still exist locally / at high spatial resolution scale.



Tiede, et al, unpulbished data

Theory on molecular solution X-ray scattering calculations

Atomic apparent form factor / contrast :

$$A_j(q) = f_j(q) - g_j(q)$$

atomic form factor in vacuum

form factor of excluded solvent

X-ray scattering total amplitude:

$$\sum_j A_j \exp(i\vec{q} \bullet \vec{r}_j)$$

In solution, X-ray beam sees all orientations of molecules :

$$I(q) = \langle I(\vec{q}) \rangle_\Omega = \left\{ \left\langle \sum_j A_j \exp(i\vec{q} \bullet \vec{r}_j) \right\rangle_\Omega \right\}^2$$

loss of phase information

loss of angular direction

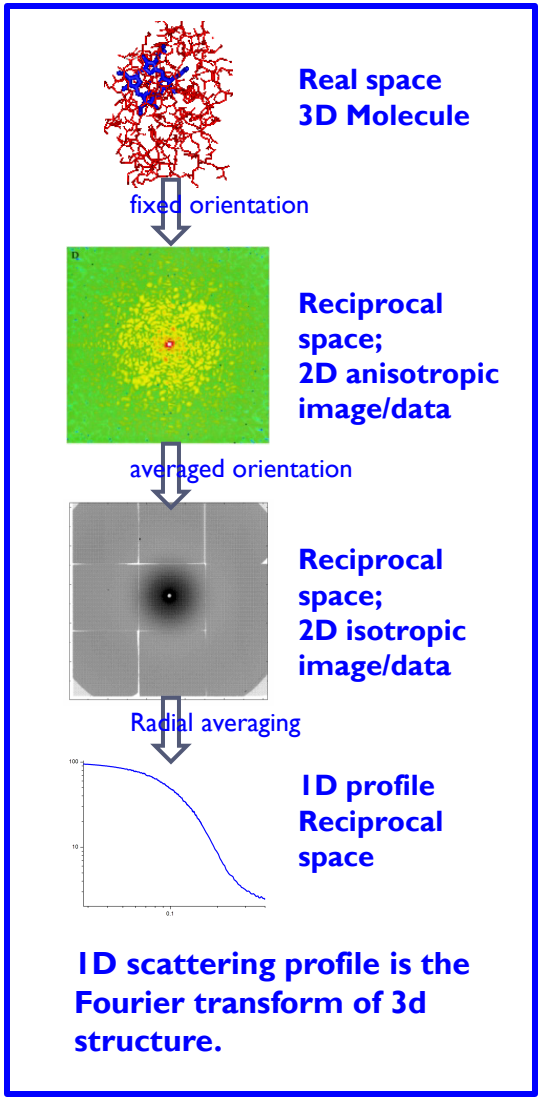
$$= \sum_j \sum_k A_j A_k \frac{\sin(qr_{jk})}{qr_{jk}} = \sum_j A_j^2 + 2\sum_j \sum_{k>j} A_j A_k \frac{\sin(qr_{jk})}{qr_{jk}}$$

Atom pair distance / Structural information

Individual contribution

interference

Solution x-ray scattering is a 1D profile which encodes molecular structural information.



3. Experimental Aspects of Scattering

a. Instruments
- X-ray generator (bench-top, synchrotron)
- Detector
- Synchrotron-based setups

b. Data acquisition
- Synchrotron based SAXS experiments
- Bench-top X-ray source based SAXS experiments
- Neutron scattering

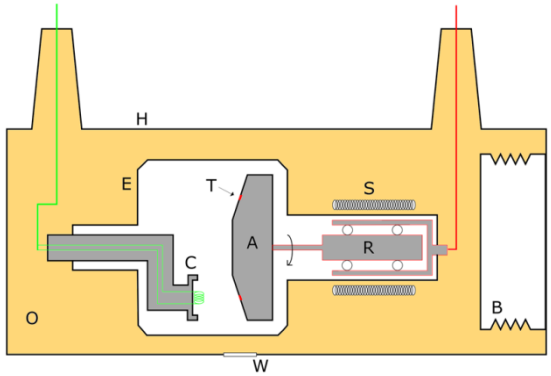
c. Scattering sample preparations

3a. Instruments

- X-ray generator (synchrotron & bench-top)
- Detector
- Synchrotron-based setups

Where do the lab X-rays come from?

Rotating anode source:
Flux ~10^{10} sec^{-1} for microfocused sources
Anode material: Cu – 8keV X-rays
Manufacturers: Rigaku, Bruker
Pricey to operate (at least $15K/yr),
Always at ~6000 RPM,
Filaments need to be replaced often.
In the scheme, c is cathode, a anode,
w – xray window



Sealed tube source:
No moving parts,
Cheap to operate ($2K/yr),
Last for up to 3 years.
Manufactures: Rigaku, Bruker, PANalytical, Hamamatsu

Line shaped sealed tube sources provide flux comparable
to non-microfocused rotating anodes
at the expense of data smearing.

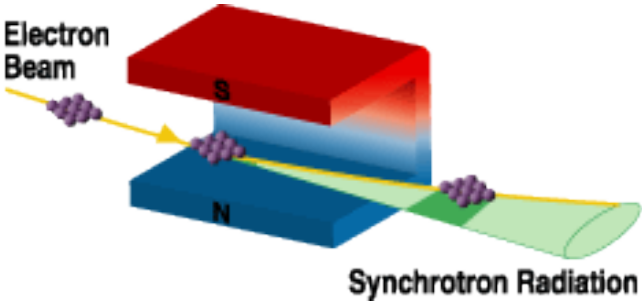
Comparison of lab sources (Source: Bruker)

	Relative Brightness[1]	Maintenance Requirements	Typical Relative Cost	Installation Requirements	Suitable Applications
Conventional sealed tube	1	Low[2]	1	3-phase power, water cooling	Strong diffractors
Microfocus sealed tube	3-5X	Low[2]	2-3X	None	Strong-moderate diffractors
Conventional rotating anode (5 kW, 300 µm)	3-10X	High[3]	3X	3-phase power, water cooling	Strong-moderate diffractors
Microfocus rotating anode (2 kW, 70 µm)	20-50X	Medium[4]	3-4X	3-phase power, water cooling	All

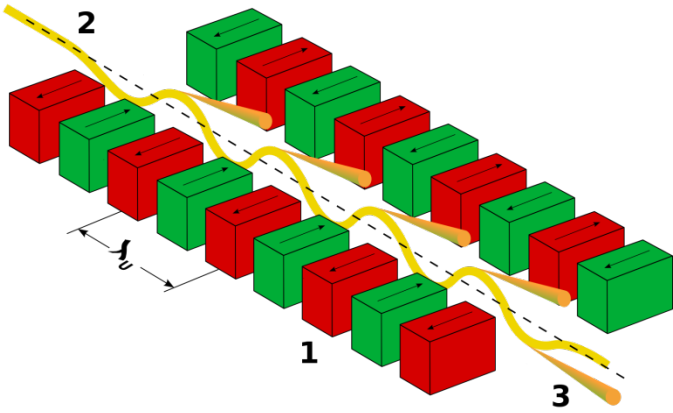
[1] Copper Kα
[2] Typically replace tube every 1-3 years
[3] Typically replace filament 4-6 times per year, refurbish anode annually
[4] Typically replace filament 2-3 times per year, refurbish anode annually

Where do the synchrotron X-rays come from?

Bending magnet: a single dipole magnet.
Has the lowest radiation flux, energy spectrum is broad.
Used at Beam Line X33 at EMBL-Hamburg.



Undulator: a periodic structure of dipole magnets that create an alternating magnetic field with a wavelength λ_u. Electron acceleration creates an intense radiation within a narrow energy band (due do a small e- motion amplitude and the resulting interference patterns), collimated on the electron orbit plane. For N periods, radiation intensity is up to a factor of N^2 higher that of the bending magnet. Used at APS Beam Lines 18ID and 12ID.



Wiggler: an undulator operating in a regime with a large e- motion amplitude, creating independent contributions from each period and resulting in a broad energy spectrum.
Used at SSRL Beam Line 4-2.

Synchrotron radiation characteristics:
Intense, focused, coherent, polarized, tunable wavelength, delivered in short bursts.

X-ray detector types

Multi-wire gas filled proportional detector (Gabriel type, 1D or 2D) detects individual ionization events by X-rays of the filler gas (CO_2/Ar). They have the lowest noise of all types, wide dynamic range limited by high local count rate, their main limitation (max ~50000/sec global or ~100/sec local). Their spatial resolution is limited (100-500 μm). They work best for lower flux sources (lab-based and synchrotrons such as SSRL).

Imaging plates are suitable for lab-based sources. They exhibit very high linear dynamic range (~5 orders of magnitude) and good spatial resolution (50 um). Not suitable for synchrotron sources.

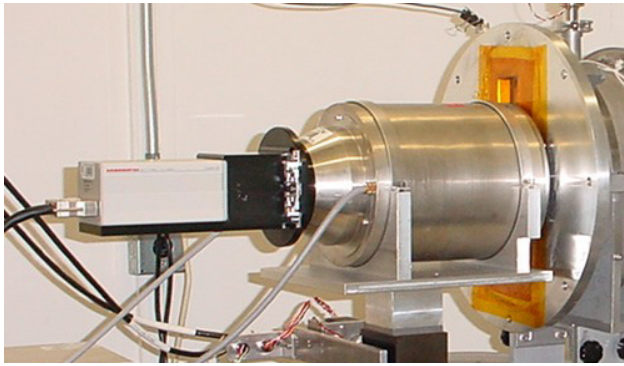
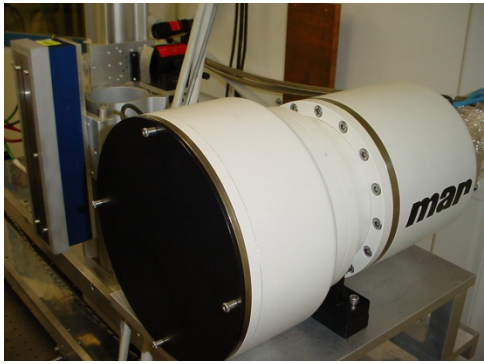


Image intensified CCD detectors (Hamamatsu): high sensitivity, large area, very rapid data acquisition. Since their count rate is unlimited, they are suitable for all synchrotron sources.



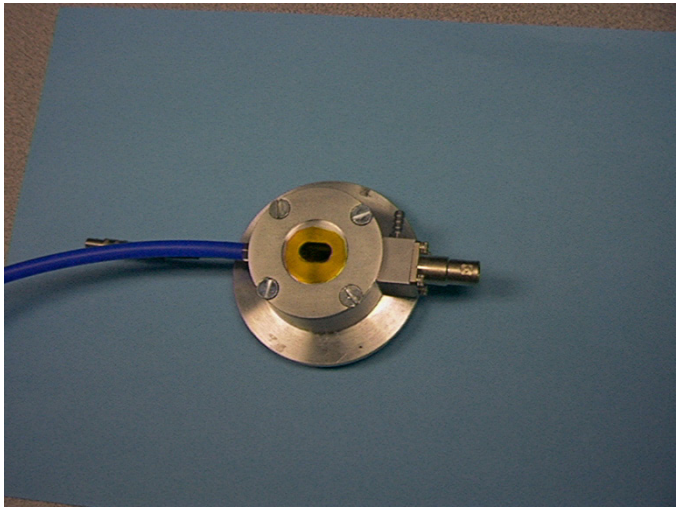
Fiber optic tape CCD detector (Mar): unlimited count rate, small pixel size (50-80 mm), low image distortion, rapid data collection. Suitable for all synchrotron sources

Detectors are periodically calibrated by long data collections of known profiles. This can be done with either ^{55}Fe radioactive source or with scattering by a glassy carbon standard samples previously calibrated using neutron scattering.

Count rate monitors

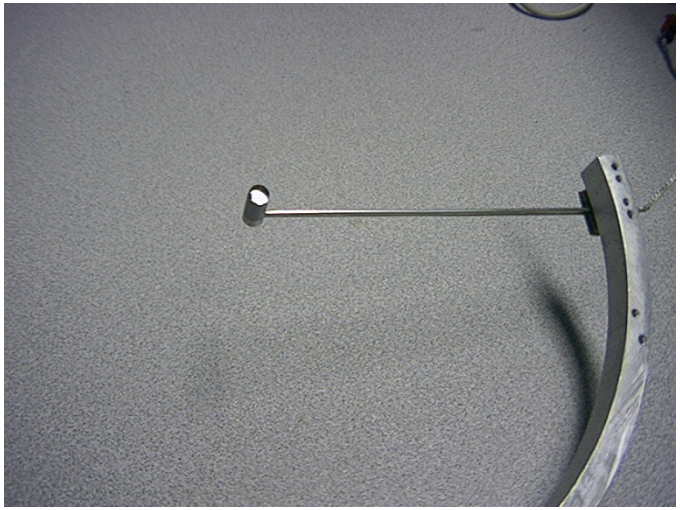
Accurate subtraction for the buffer scattering from the sample requires calibrations by the photon fluxes during the two measurements. This is accomplished by measuring beam intensity upstream and downstream from the sample.

The **incident intensity** (upstream) measurement is typically done with either ionization chambers or NaI scintillation detectors pointing towards a transparent scattering material (such as Kapton foil) in the beam. Ionization chamber measurements are limited by the drift of the dark current and NaI by the radiation damage/variable energy deposition to Kapton.



The **transmitted intensity** (downstream) measurement is done with a beamstop-mounted photodiode. Its accuracy is limited by the drift of the photodiode's dark current and the lateral drift of the beam during data collection. The transmitted intensity is affected by any changes in sample material including radiation damage and bubble formation.

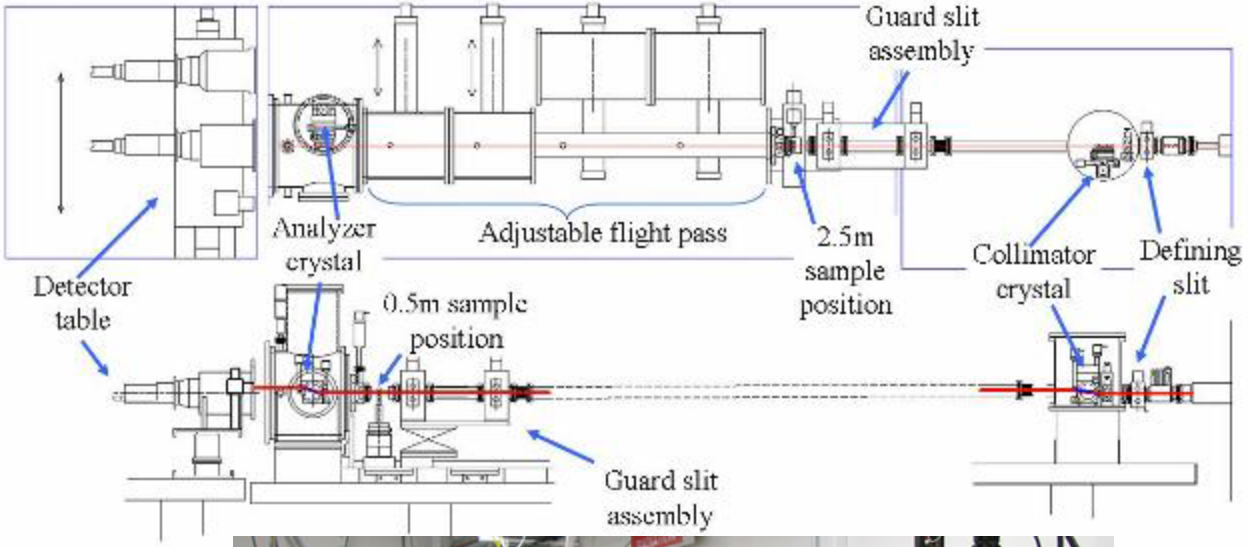
The transmitted intensity measurement can also be done using a semi-transparent beam stop to get the transmitted beam intensity to a level comparable to the scattering



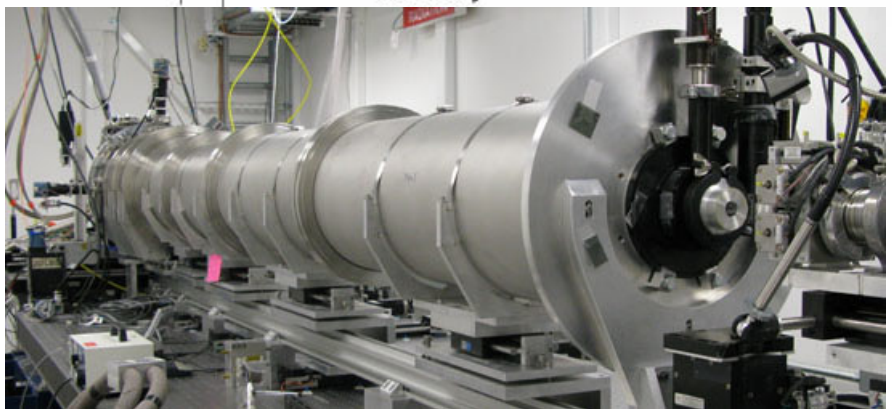
SSRL Beam Line 4-2

Dedicated for bio-SAXS and diffraction. Source is a 20-pole wiggler magnet.

energy range: 6-14 keV
q-range: 0.003 - 3 A^{-1}
flux: ~5 10^{11} sec^{-1}

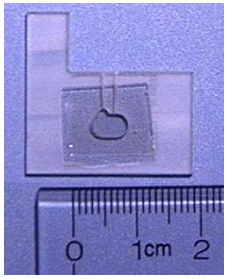


Detectors: linear gas proportional, quadrant gas proportional, Hamamatsu image-intensified CCD, Mar fiber optic taper coupled CCD, Rayonix 225HE back illuminated taper CCD

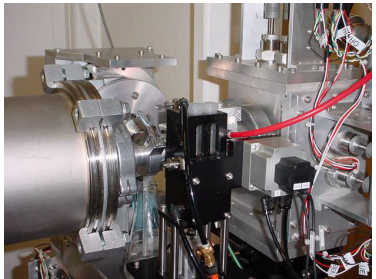


9 mo/yr operation, 3 proposal cycles

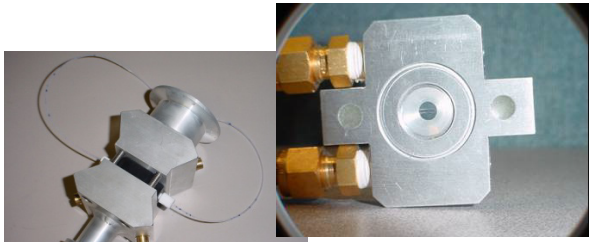
sample cells: 15 μL, static



stropped flow mixer



in-vacuum flow cell

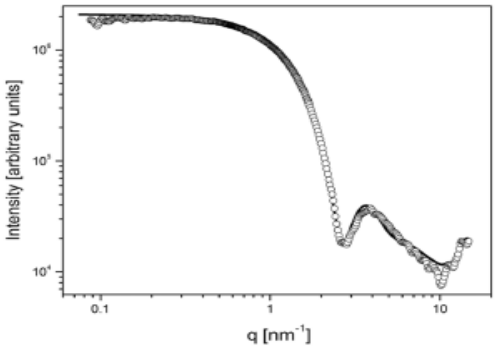


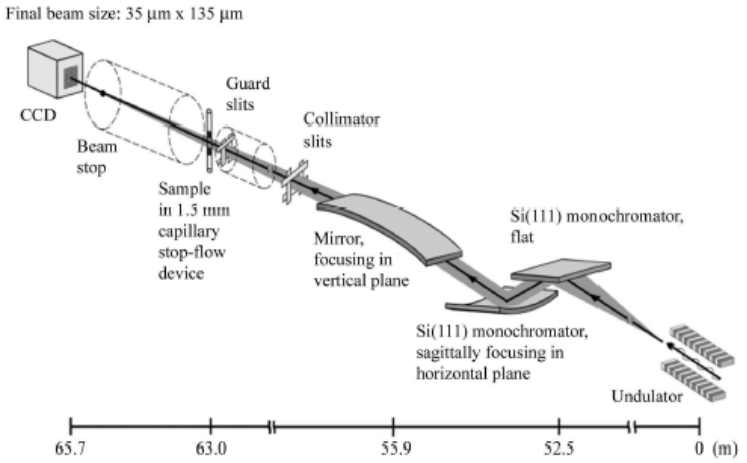
SAXS Beam Lines at APS

18ID (BioCAT): dedicated for bio-SAXS and fibre diffraction.

source: APS Undulator A
energy range: 3-13 keV (fundamental), 10-40 keV (3rd harmonic)
q-range: 0.003 – 2.5 A^{-1}
flux: ~10^{13} sec^{-1}

~70% of operation is user time

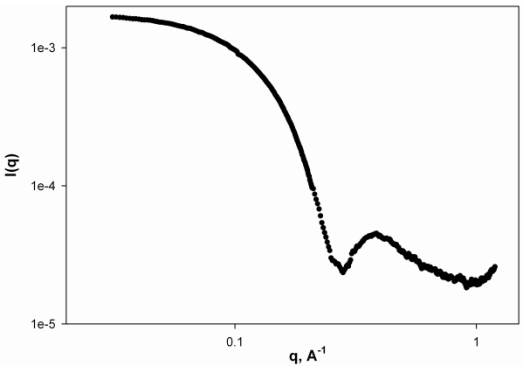


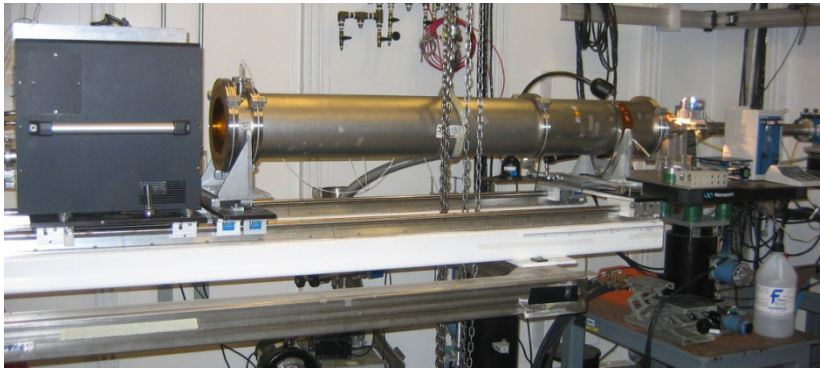


sample data above: 2.5 mg/mL cytochrome c, 1s data collection at 2m and 0.3 m

Fiscetti, F et al. (2004) The BioCAT undulator beamline 18ID: a facility for biological non-crystalline diffraction and X-ray absorption spectroscopy at the Advanced Photon Source. J. Synchrotron Rad. 11, 399-504

12ID (BESSRC): material science, chemistry, biomolecular.
source: APS Undulator A
energy range: 3-27 keV
q-range: 0.003 – 2.5 A-1
flux: ~10^{13} sec-1





sample data above: 2.5 mg/mL cytochrome c, 1s data collection at 4m and 0.3 m
~70% of operation is user time (not all is for SAXS)

3b. Data Acquisition

- Synchrotron-based SAXS experiments
- Bench-top X-ray source-based SAXS experiments
- Neutron scattering

Synchrotron data collection overview

Measure empty capillary, capillary with buffer and capillary with sample along with transmissions (if possible). Best practice is to subtract empty capillary form both buffer and sample then to subtract buffer from sample. Watch for radiation damage – monitor P(r) and Guinier during data collection, attenuate beam and decrease exposure when necessary.

Measure in 2 configurations (SAXS and WAXS whenever possible), then merge. Use SAXS to evaluate low-q behavior (inter-particle interference, aggregation, radiation damage) and WAXS to evaluate sample/buffer matching. Maximum exposure time with SAXS is going to be limited by low-q scattering near the beam stop. For WAXS it will be limited by the water peak intensity.

Optimize exposure times to prevent both radiation damage and detector saturation. Higher solute concentrations (5-20 mg/mL) can be used for WAXS - with q_{min} ~0.1 A^{-1} structure factor should not be an issue. Concentrations as low as ~0.1 mg/mL protein can be measured.

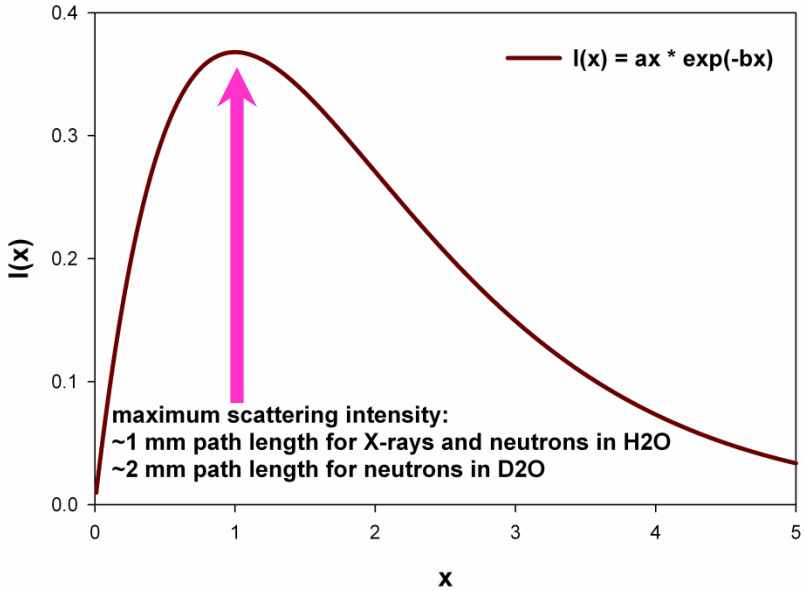
Data collection: sample handling

Sample volumes are 15-30 uL for a static cell and 50-150 uL for a flow-cell.

It is best to measure buffer before and after each sample measurement to detect protein deposition on the cell walls.

Cleaning the cell between measurements is crucial. A sequence of using water / bleach / isopropanol / water works well with quartz capillaries. The final wash before any load (whether sample or buffer) should be the matching buffer.

If the sample is oscillated during data collection, dilution with droplets of buffer left by the previous wash can and will occur. The consecutive individual data frames might thus exhibit a systematic lowering of scattering intensities. This should not be confused with radiation damage. On the other hand, oscillation of the sample, typically done to decrease radiation damage, does not eliminate the laminary nature of the flow. Therefore, protein molecules close to the capillary surface are not exchanged as fast as the interior of the capillary and might still experience a radiation dose that leads to radiation damage.



Optimal path length depends on solvent absorptivity and probability of multiple scattering

Lab-based data collection overview

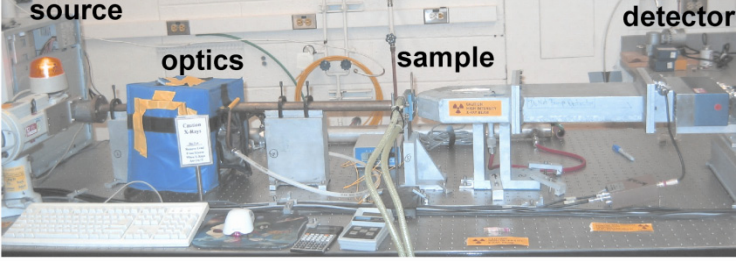
Radiation damage is not an issue, but s/n is. Long data collections (hours) for capillary with buffer and sample are necessary.

Room temperature stability and bubble formation in the cell during these long data collections is the main issue determining data match quality.

If line source is used, desmearing with be necessary.

Data scaling is done relative to the transmitted intensity measurements using beamstop mounted pin diodes or semi-transparent beam stops.

Minimum concentration is ~ 1 mg/mL protein.



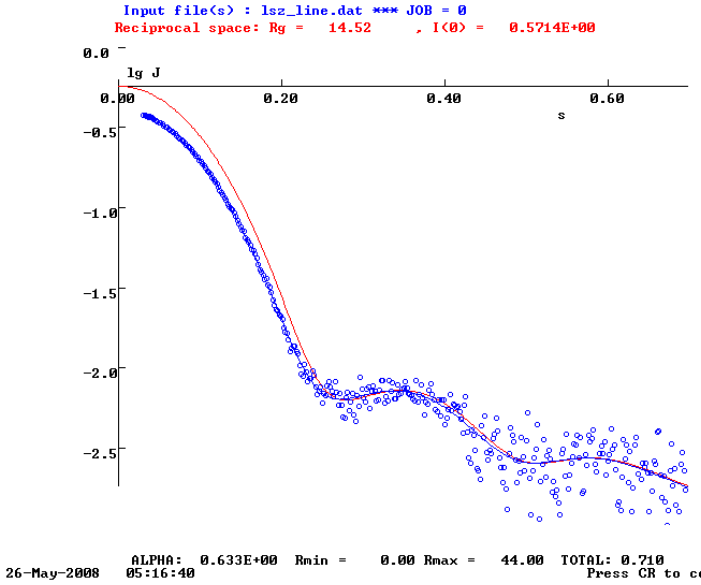
Line beam lab sources

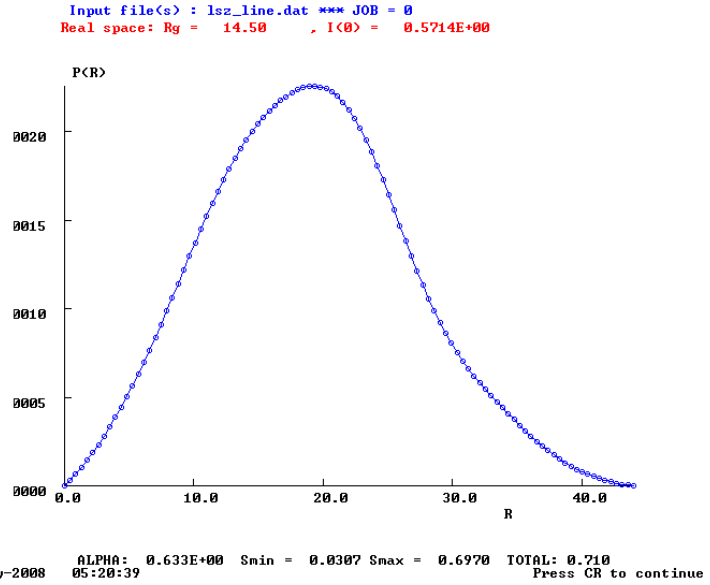
Advantage: high flux for a lab source.

Disadvantage: the need to remove the effect of the beam profile from the data (desmearing). Primarily low-q data are affected.

Beam shape measurement is needed for desmearing.

$$I_{0S}(h) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_z(z)\,W_y(y)\,I_0[((h-y)^2+z^2)^{\frac{1}{2}}]\,dydz$$





Feigin, L., Svergun, D. (1987) Structure Analysis by Small-angle X-ray and Neutron Scattering (Chapter 9). Plenum Press.

Lake, J. (1966) An iterative method for slit-correcting small-angle X-ray data. Acta Cryst. 23, 191-194.

SANS data collection overview

• Select mean and r.m.s. neutron wavelength, exposure times, sample/detector distances and collimation parameters to cover the desired q-range with an acceptable signal/noise level. Radiation damage is not an issue for SANS!

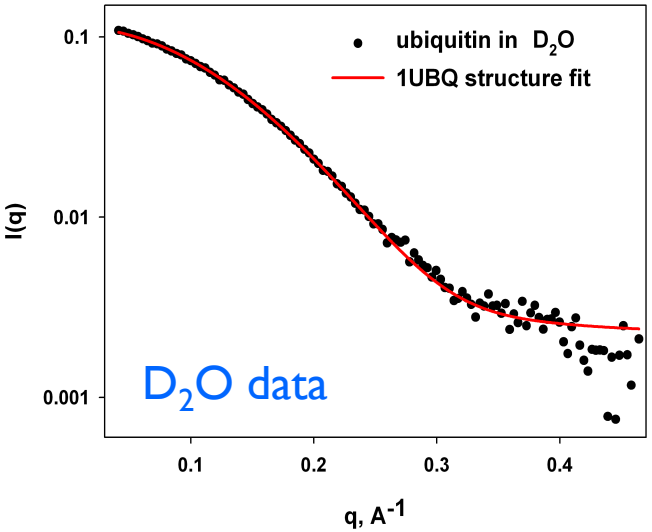
• Measure blocked beam, open beam, empty cell, cell with buffer and cell with sample along with the transmissions.

• Blocked beam and empty cell are scaled by their transmissions and subtracted from all data.

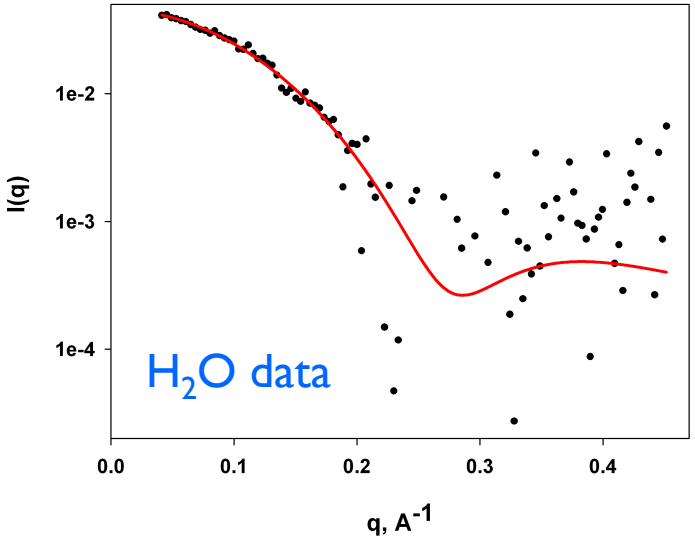
• Bubble formation and buffer evaporation during these long (hrs) measurements are main issues.

• Minimum sample concentrations are ~2 mg/mL protein. Low q-measurements with sample/detector distances exceeding ~5m require data collection times of several hrs.

• Buffers containing more than ~50% H_2O also require significantly longer data collection times. For buffers containing more than 50% D_2O, 2mm path length cell will provide significantly improved data statistics but will require ~600 μL sample instead of ~300 μL for a 1mm path length cell.



1 hr data collection on NG3 instrument for 10 mg/mL protonated protein samples



3c. Scattering Sample Preparation

Sample preparation procedure

Preparation of an exactly matching buffer is the most crucial step. Do not use the buffer the protein was dissolved in. Long (16-48 hr) dialysis works best. Hampton dialysis buttons (~ 50 uL) or Harvard Apparatus units for micro dialysis (up to 150 uL) both work very well. In cases when dialysis cannot be done (limited sample stability or limited time for prep), passing buffer by centrifugation though a proper MW-cutoff filter comes close. When doing this, be aware that the fresh membrane can have organics on it, so pass a volume of buffer to remove them first. Several cycles of buffer loading will be needed to ensure the exact match.

In order to decrease the tendency for bubble formation during data collection, the dialyzate buffer should be degassed under lab vacuum and/or sonicated.

Prior to loading the sample for dialysis, it is good to pass it through a 0.22 μm filter.

Sample concentration should be measured after the dialysis (OD_{280} for most proteins, OD_{260} for DNA).

When transporting the sample for data collection the sample should be keep air-tight to prevent solvent evaporation especially if sample volume is small.

When preparing the buffer for a synchrotron data collection, at least 50 mL of buffer will be needed since washing the flow cell (if it is going to be used) requires a lot of volume.

Composition of the buffer

Salts are often useful to suppress long-range electrostatic interactions between solutes (structure factor). They also increase background and decrease solute/solvent contrast but these effects are often negligible up to ~500 mM salt. High-Z elements should be avoided in the buffer. They decrease contrast and promote radiation damage by increasing photo-electron production. Due to RNA/DNA's higher surface charge relative to proteins, structure factor suppression by salts is usually weaker than for proteins at a given salt concentration.

Free radical scavengers should be included in the buffer when preparing for a synchrotron data collection as they help to minimize the radiation damage. Common choices are DTT (2-10 mM), TCEP (1-2 mM), or glycerol (~5%). In cases when these cannot be used, organic buffers containing TRIS or HEPES can also act as radical scavengers. When using DTT, keep in mind that it is oxidized by air. Degassing the dialysis buffer and keeping both sample and buffer under nitrogen throughout the entire preparation up till data collection will suppress this.

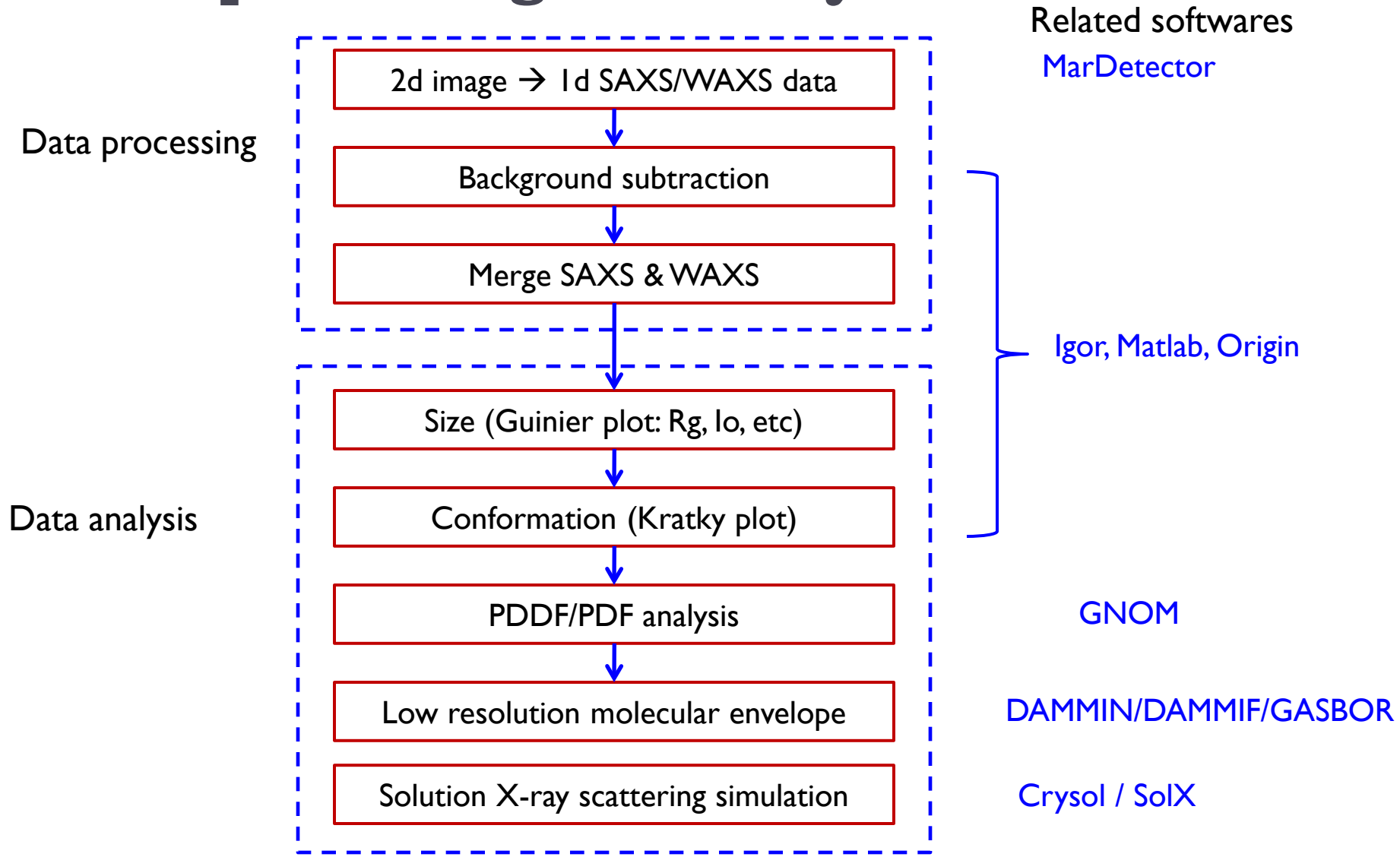
Detergents are best avoided unless absolutely necessary (membrane proteins). Their signal can be comparable to or exceeding the protein signal complicating data interpretation.

Sample preparation issues

Aggregation is the most common problem that can render data uninterpretable. Up to 5% of dimer relative to the monomer in question might not affect the data too much but higher levels will. DLS or analytical ultracentrifugation can be used to detect aggregation/polydispersity and native gel (single band is required), gel filtration, or centrifugation through high-MW cutoff membrane can help to remove it. Otherwise, sample conditions will have to be optimized. Fresh preps kept at low concentration till data collection (when they can be concentrated) work best in difficult cases. Freezing/thawing cycles can promote aggregation and are best avoided. Cryo-protectants (5-10% glycerol) can help in cases when freezing the samples is unavoidable. A fraction of very large aggregates (relative to the particle that is being studied) is better than a dimer/trimer/tetramer/etc continuum. In the former case, the scattering from aggregates can be removed by removing lowest-q data. In the latter case, the entire scattering curve is affected and data become truly uninterpretable.

Solute concentration: On one hand, higher concentration = better signal. On the other, it can bring increased aggregation or pronounced inter-particle repulsion (or both). For proteins with MW<200 kDa 1-10 mg/mL is a suitable range. For larger proteins concentrations below 2 mg/mL should be used. A 3-6 point concentration series (such as 1, 2, 5,10 mg/mL) should be acquired to test concentration dependence of the signal. Analysis of I(0)/c should indicate whether concentration effects are present. In those cases, zero-concentration extrapolation should be performed. For proteins in 150 mM salt buffer, structure factor is often quite small up to 8-10 mg/mL. Since RNA scattering is stronger, concentration can be lowered by a factor of ~5 relative to proteins for a comparable s/n. When using 150 mM salt buffer, structure factor often becomes close to negligible at 1 mg/mL DNA/RNA or lower. Higher concentrations are very likely to have noticeable structure factor effects at these conditions. OD_{280} is the preferred method for measuring the concentration for proteins. When doing so, be aware that oxidized DTT will also absorb at this wavelength.

Data processing and analysis flow chart



4. Data processing

a. 2D → 1D data conversion
b. Background subtraction
c. Data quality evaluation

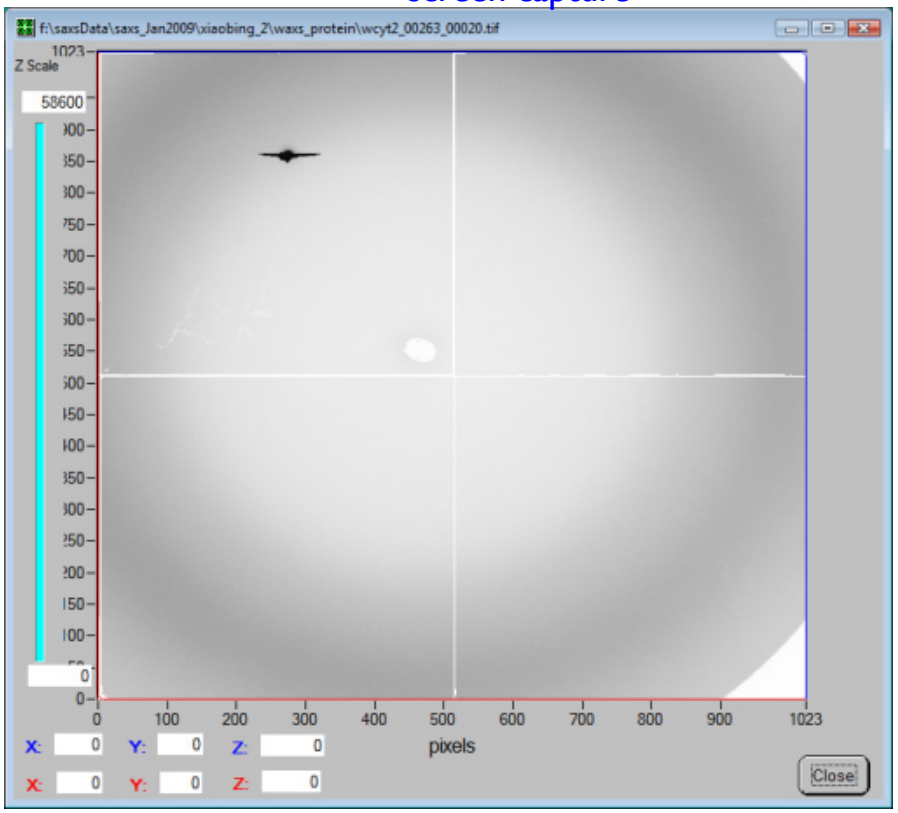
a. 2D image → 1D data conversion

1. Find beam center
2. q calibration/mapping
3. Image mask
4. Conversion

The above procedures for both SAXS & WAXS

done at synchrotron beamline station during data aqusition

Software: Mardetector v4.9 screen capture



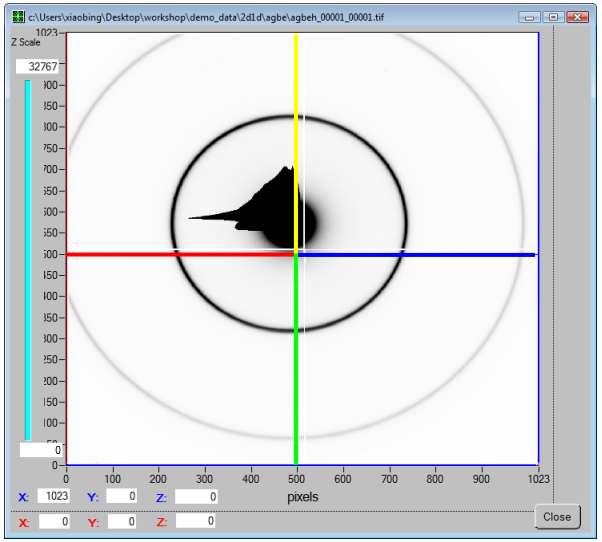
A 2D scattering image for a solution sample recorded on a home-made 2x2 CCD chips detector (gold detector) at 12-ID at Argonne

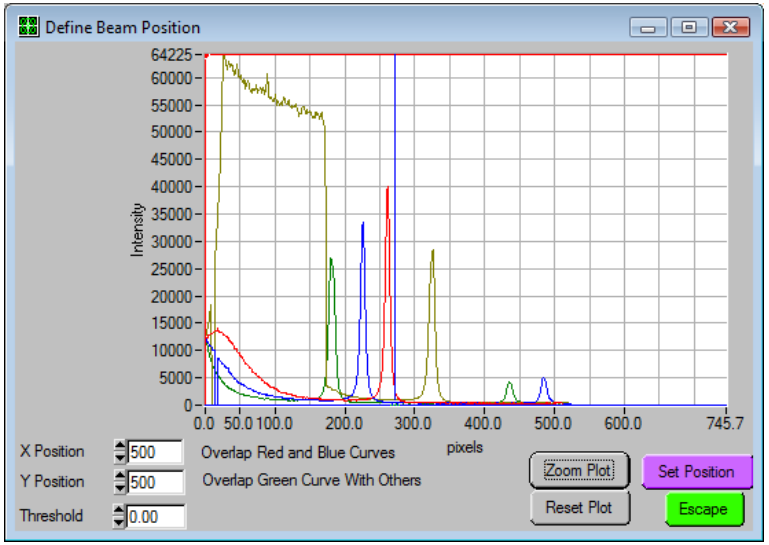
Find beam center

Software: Mardetector v4.9
D. Tiede (Argonne) & X. Zuo

Initial: off beam center

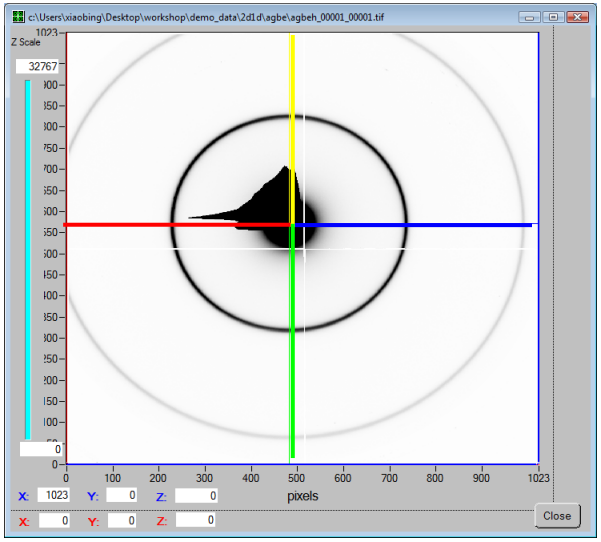
Along the four directions, peak positions do not overlap well



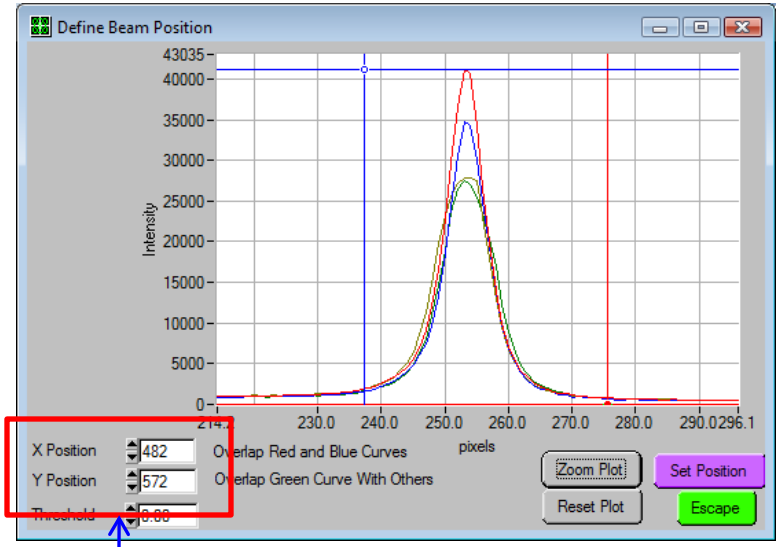


At beam center

Peak positions overlap well



Scattering image of silver behenate



position of beam center on detector

q-calibrants and sample-to-detector distance

Primary q-calibrant standards in X-ray scattering:

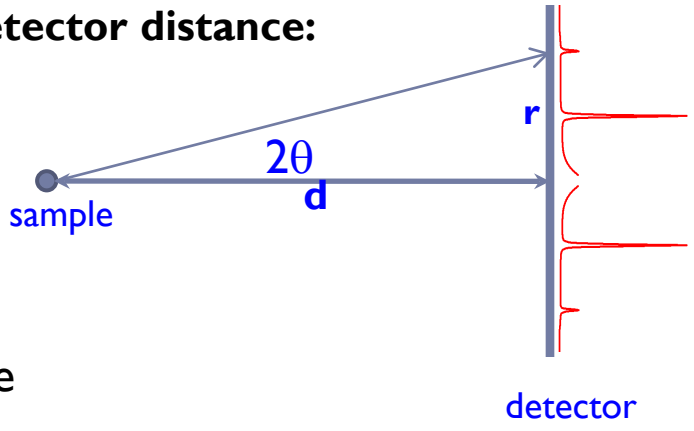
standards	d-spacing (Å)	First harmonic peak ($Å^{-1}$)
Silver behenate (IUCr standard, most popular)	58.380	0.1076 0.1076 x N (N≤10, Nth peak)
Cholesterol Myristate	51.1	0.1230
Wet Rat Tail Collagen	650-670	~0.01
Anatase (TiO_2)	3.51	1.790

Using q-calibrant to determine sample-to-detector distance:

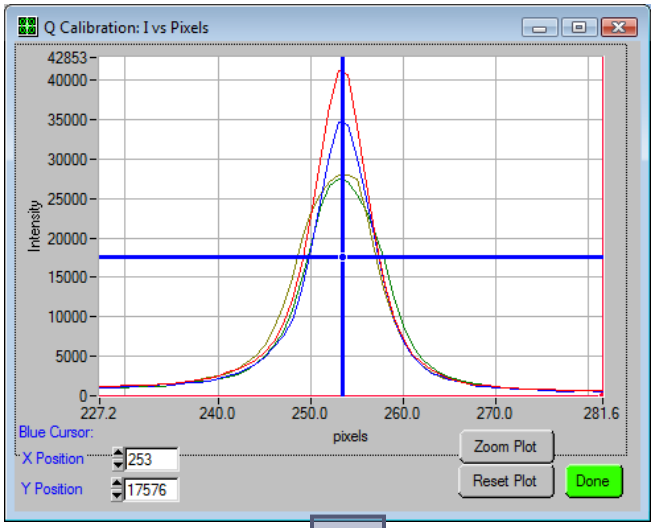
$$q = (4\pi / \lambda)\sin\theta$$

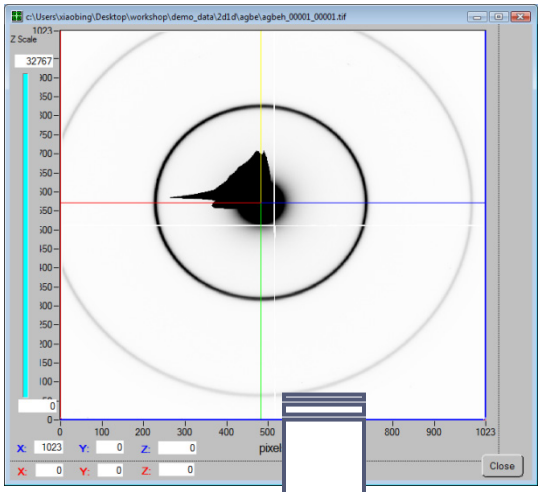
$$2\theta = \mathrm{atan}(r / d)$$

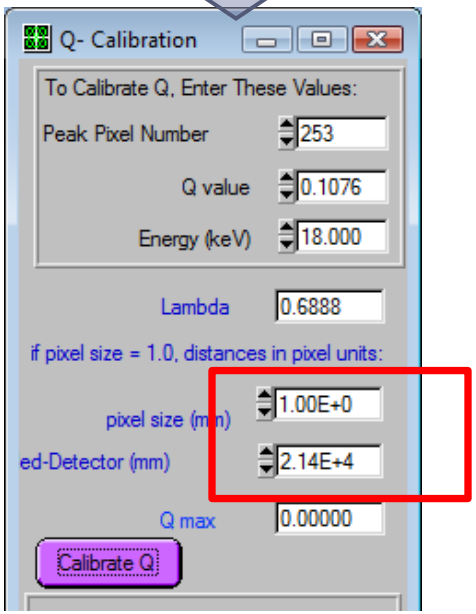
d: sample-to-detector distance



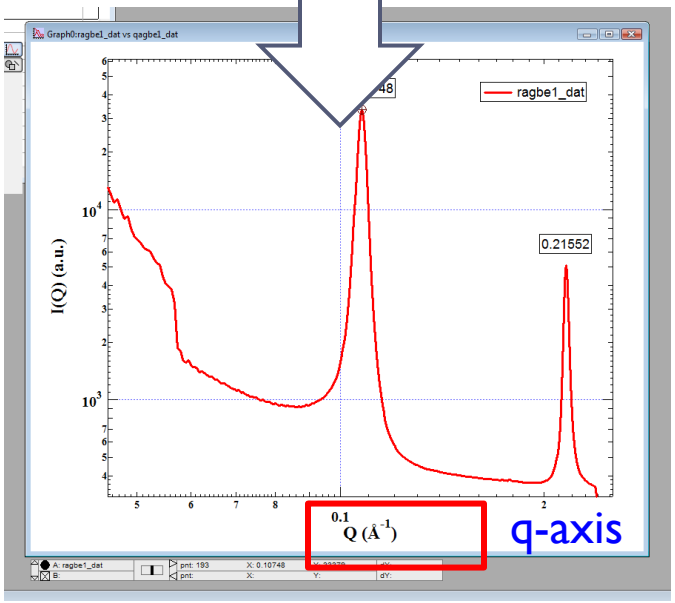
q value calibration/mapping







Sample-to-detector distance

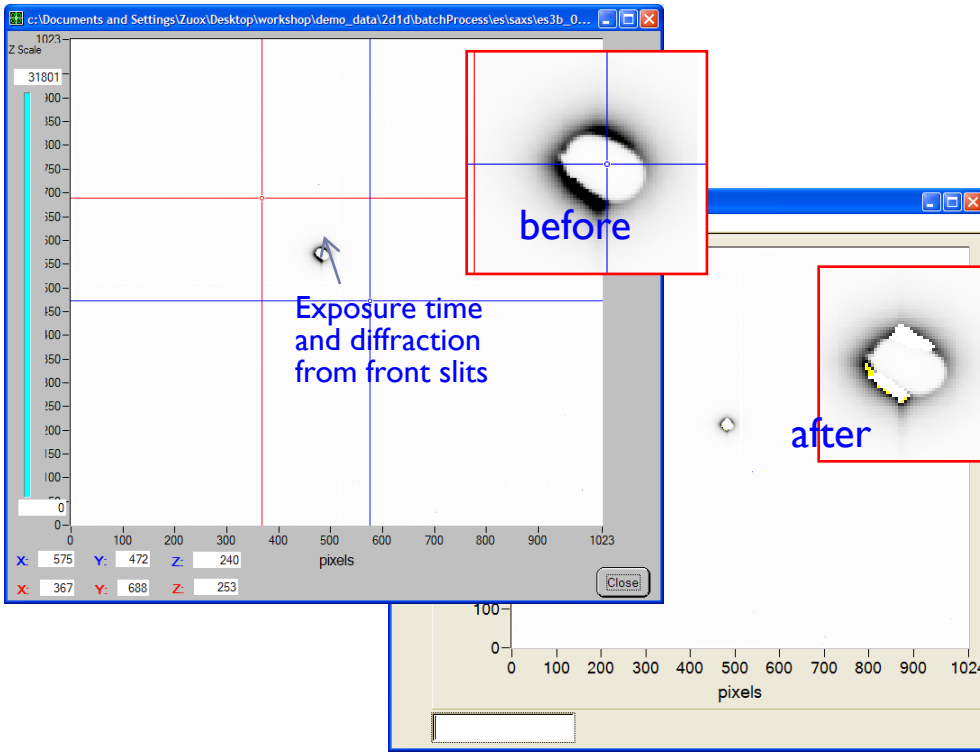


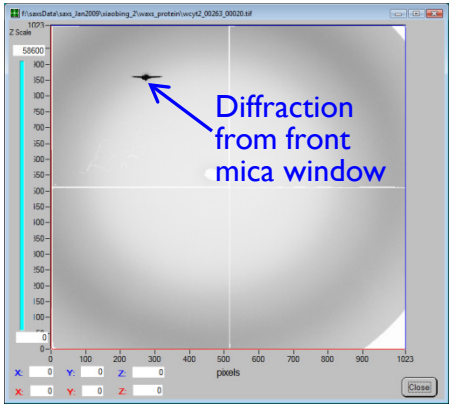
Every time sample-to-detector distance is changed, q-mapping need to redo.

Image Mask and zinger removal

Mask:

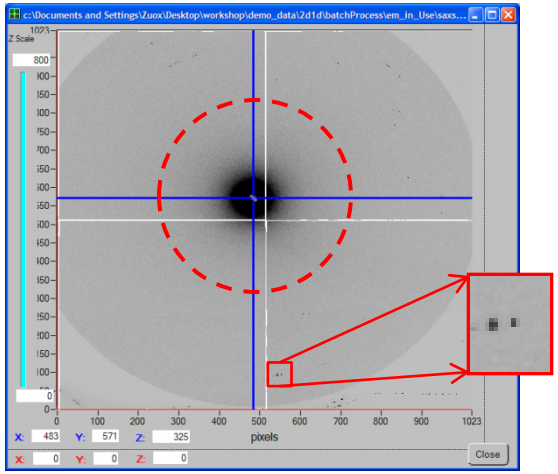
- detector responses nonlinearly beyond a threshold,
  - Gold detector at 12-ID: 44k
  - Marccd165: 65k
- High intensity at low q could be due to too long exposure time, diffraction from slits, fluorescence from beam stop, etc
- High intensity spots at higher q could come form diffraction of front mica window
- Mask out nonlinear pixels



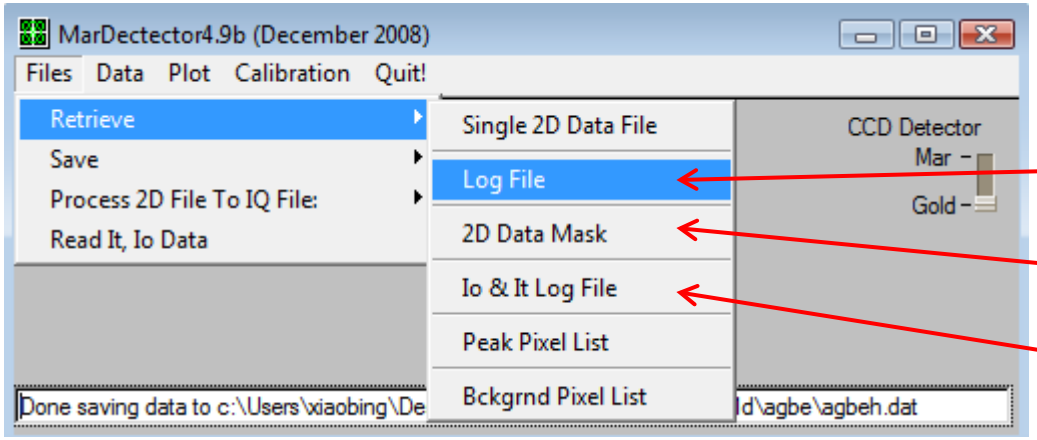


Zinger removal:

- Zinger: spots with intensity much higher than those of neighbor pixels
- Comes from diffractions, electronic noises, etc
- ignored if $|I(x,y)-I_{ave}|>6\sigma$



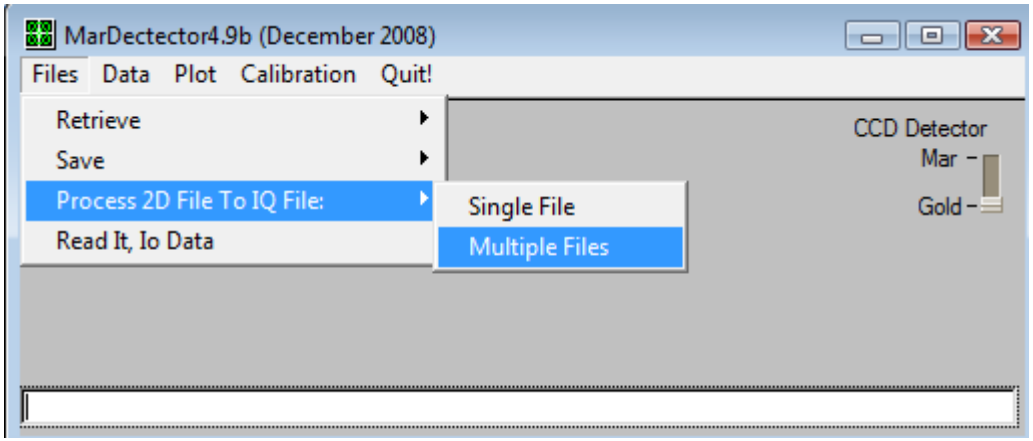
Running 2D→1D conversion program



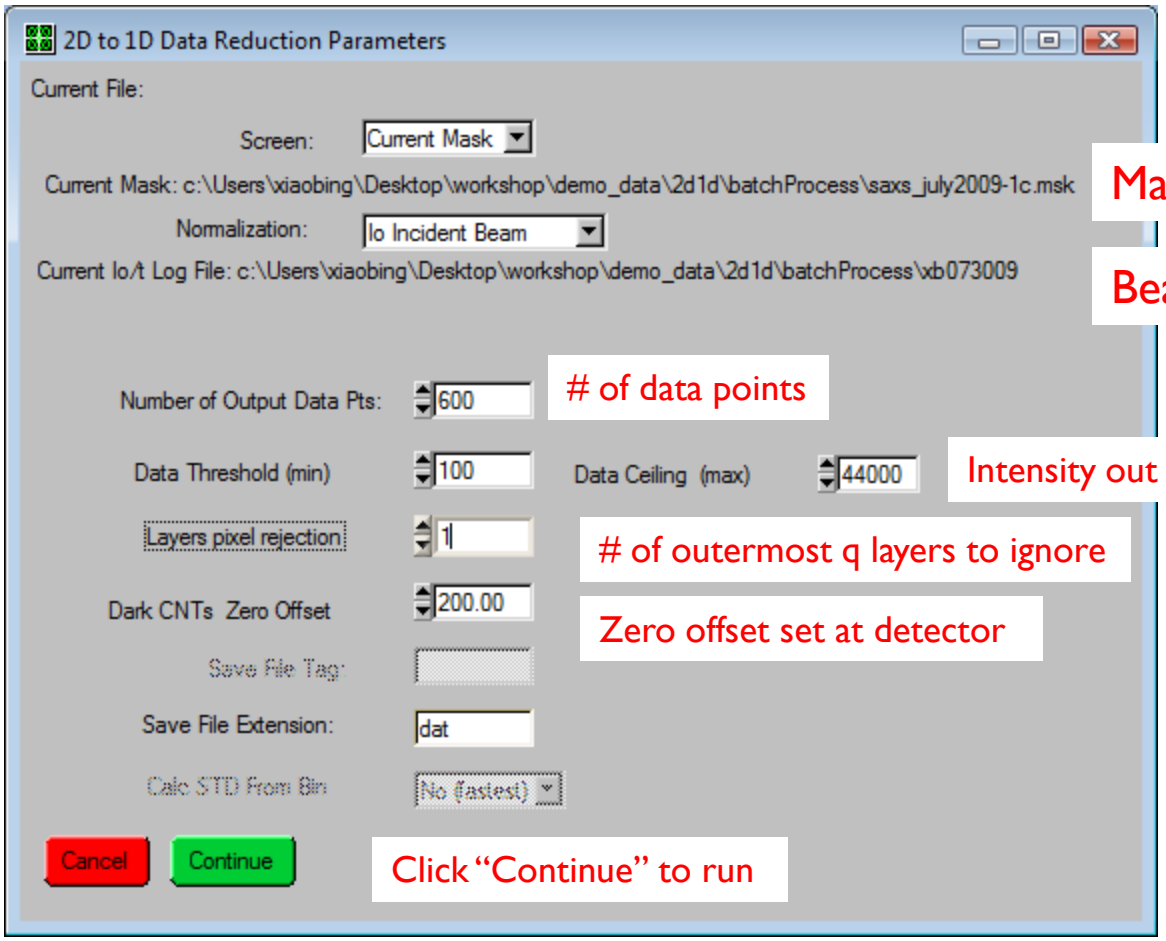
Configuration file

Mask file

Beamline log file: Io & It



2D→1D Parameter Panel



Mask file

Beamline log

# of data points

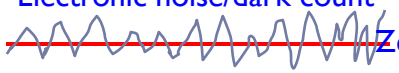
Intensity out of [min, max] will be ignored

# of outermost q layers to ignore

Zero offset set at detector

Click "Continue" to run

Electronic noise/dark count



Zero offset catches
otherwise negative noise

Log files used in image conversion

Configuration log file:

```
CCD Log File version: 4.9
1.000000
21447.285156
0.688800
18.000000
482
572
600
0
100
k:\gold\xiaobing1\program\saxs_july2009-1c.msk
44000
0
0
200.000000
1
```

- set the size of detector pixel as "1.0"
- **Sample-to-detector distance** relative pixel size, calibrated from q-spacing standard
- X-ray wavelength
- X-ray energy in KeV
- **X-ray beam center**
- Number of resulting data points
- Minimum intensity counts threshold for pixel to consider by program
- Mask file
- Maximum intensity counts threshold for pixel to consider by program
- Pixel intensity count offset
- Number of outermost layers to ignore

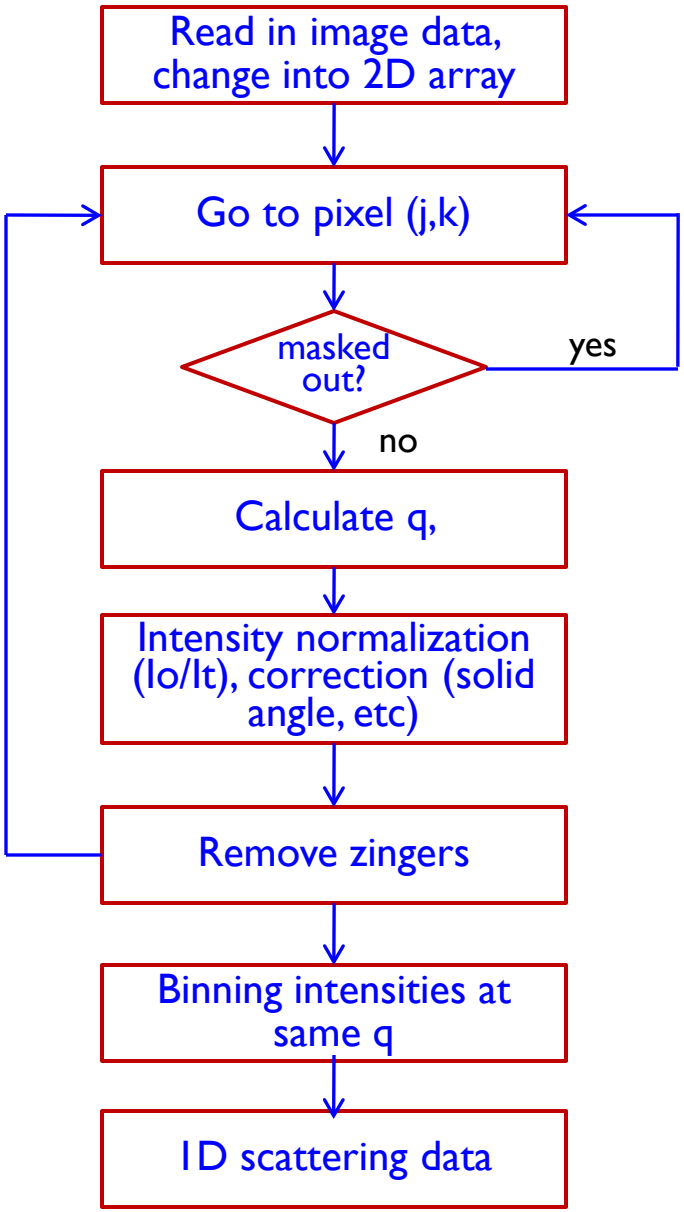
Beamline Io/It log file:

```
#F xb073009
...
#Z swat1_00005_00001.tif 2000044 36471 34051 0.100 18.000 20.435 Thu Jul 30 15:11:59 2009 1248984719.881776
#Z swat1_00005_00002.tif 2000044 36540 34104 0.100 18.000 20.467 Thu Jul 30 15:12:01 2009 1248984721.609769
#Z swat1_00005_00003.tif 2000044 36502 34051 0.100 18.000 20.502 Thu Jul 30 15:12:03 2009 1248984723.401776
#Z swat1_00005_00004.tif 2000043 36574 34127 0.100 18.000 20.540 Thu Jul 30 15:12:05 2009 1248984725.289774
#Z swat1_00005_00005.tif 2000044 36492 34091 0.100 18.000 20.579 Thu Jul 30 15:12:07 2009 1248984727.221829
#Z swat1_00005_00006.tif 2000043 36565 34155 0.100 18.000 20.615 Thu Jul 30 15:12:09 2009 1248984729.129903
#Z swat1_00005_00007.tif 2000044 36433 34044 0.100 18.000 20.644 Thu Jul 30 15:12:10 2009 1248984730.885767
```

File name | Io | It | KeV

2D→1D data conversion

Data conversion flow chart:



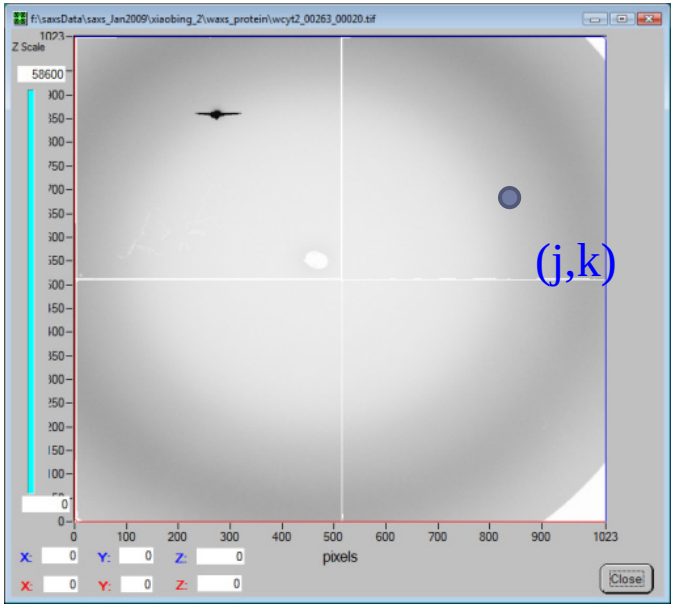


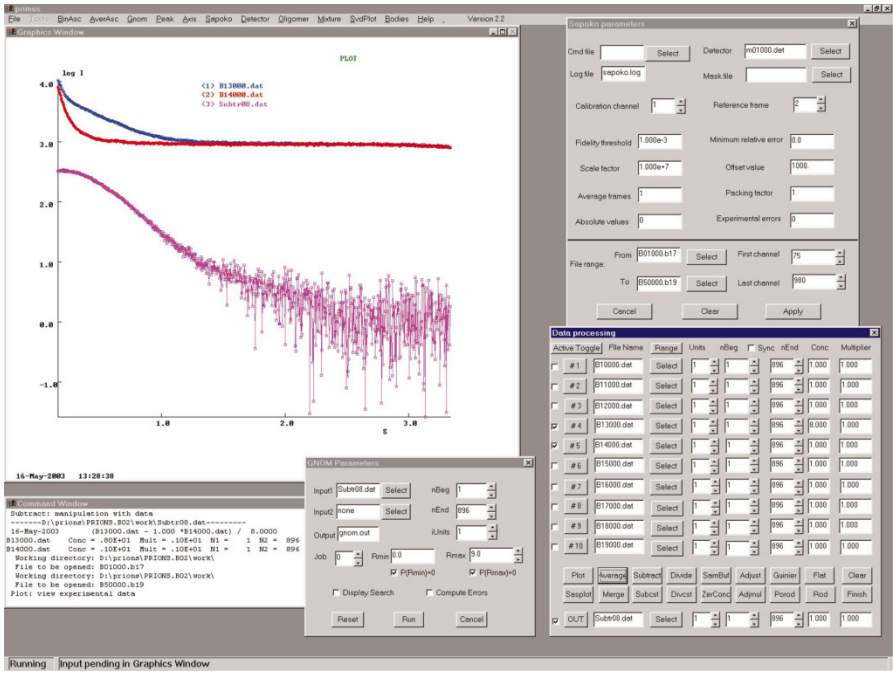
Image data structure:
1D (1kX1k) unsigned integer

X-ray detectors at 12-ID:
gold CCD detector: home-made, 2x2 CCD chips, square, data size 1024X1024
marCCD165 detector: round, 1024X1024

Softwares for data evaluation and manipulation

Primus: developed in Svergun group

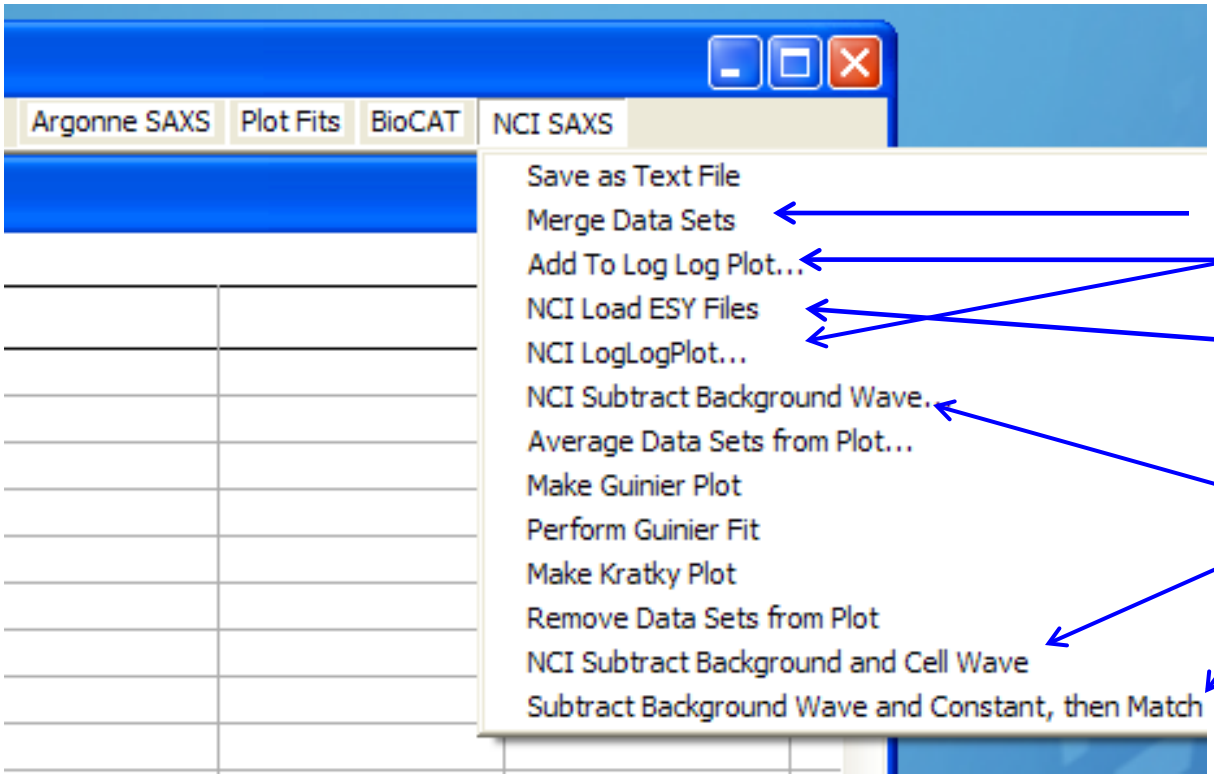
- basic data manipulation functions (averaging, background subtraction, merging of data measured in different angular ranges, extrapolation to zero sample concentration, etc.) and computes invariants from Guinier and Porod plots.
- Link other programs developed in Svergun group, such as GNOM, DAMMIN etc.



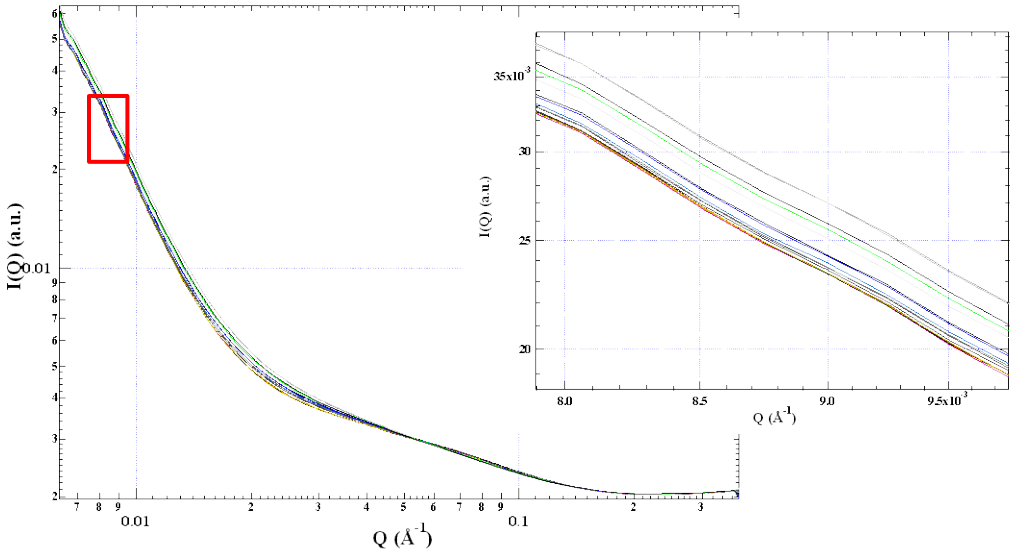
Igor / Origin / matlab:

- Procedures / scripts available. Some of Igor procedures were copied or modified from those written by Argonne groups.
- Basic data manipulations: averaging, background subtraction, merging data, Guinier plot, Kratky plot, data extrapolation, data point sparsing
- Save parameters used in data processing

Data manipulation Procedures in Igor



Check data convergence/reproducibility

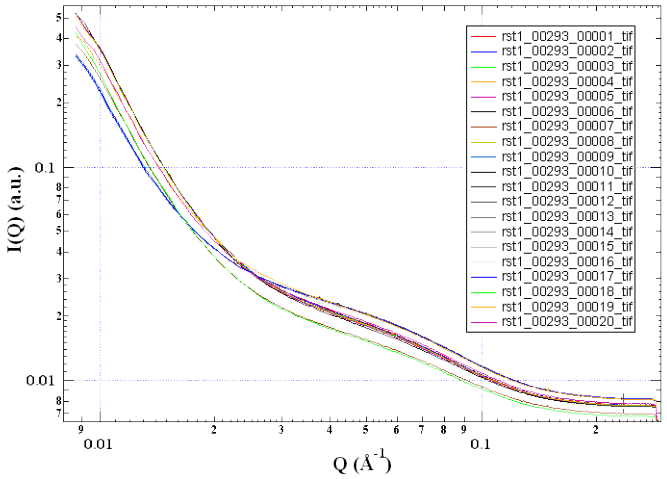


When sample degrades in X-ray, its scattering intensities change along time:

- Decrease if x-ray beaks molecules
- Increase if inducing aggregation

Real time action:

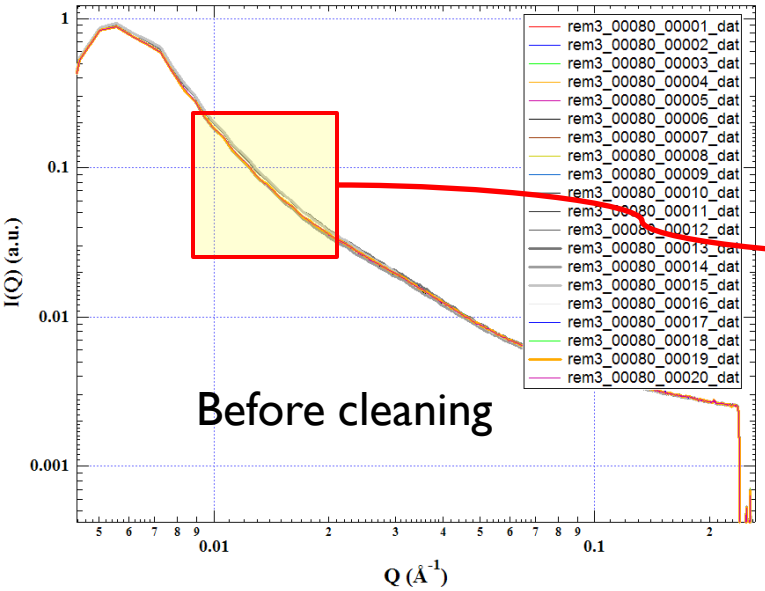
- Attenuate beam intensity/ decrease exposure time
- Increasing flow rate
- Change buffer composition if possible

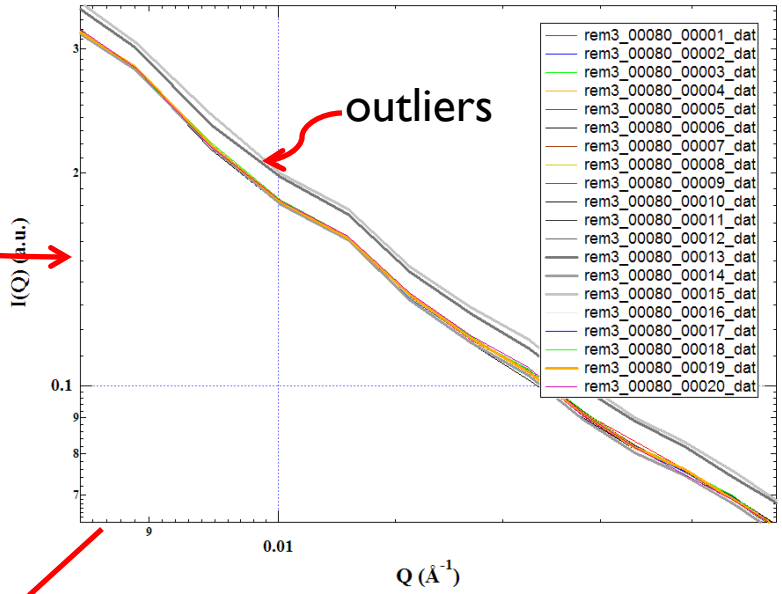


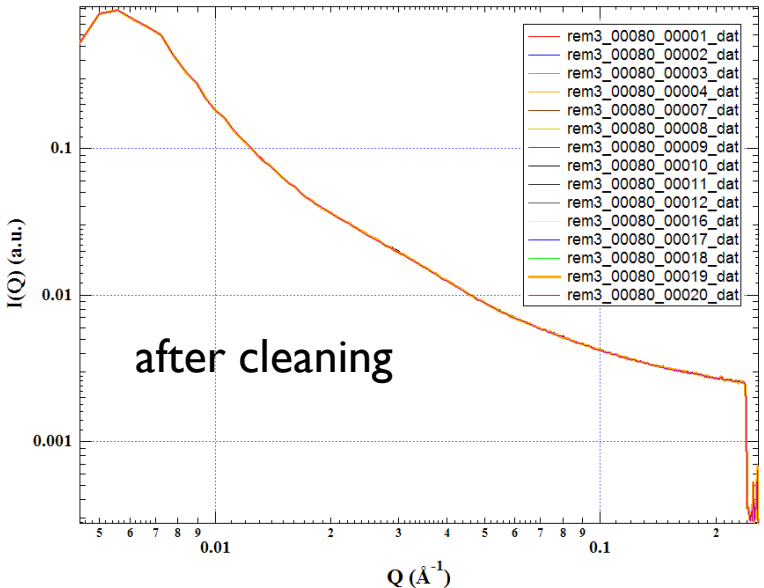
Protein samples are very easy to form bubbles, therefore, scattering data may jump everywhere and become unusable.

To prevent or alleviate this problem, clean the flow cell extensively and use slower pump speeds. Or stop flow if the sample can tolerate X-ray radiation damage.

Data cleaning and averaging







- 10-30 frames of scattering images are often collected for each sample to get good statistics.
- Outliers could exist due to air bubble, small particles, beam instability, and some unknown reasons. Outliers appear more often in sample data than in solvent/buffer data.
- Zoom in, check outerliers and remove them.

b. Background subtraction

- SAXS background subtraction:
  - A. $I_{sample} = I_{solution} - I_{buffer}$

    Scattering intensities normalized against incident beam photon counts (Io)

    - The following factors make the coefficient before I_{buffer} close but not exactly be unit: experimental errors in measuring incident photon counts, different x-ray transmission of solution vs buffer, etc

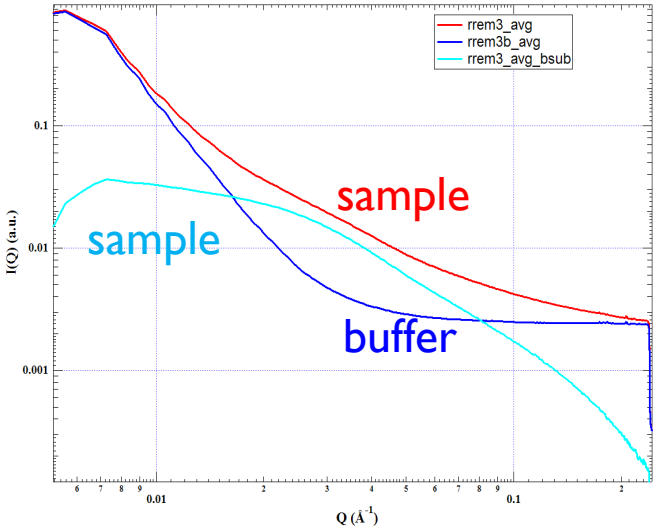
  - B. $I_{sample} = I_{solution} - \alpha * I_{buffer}$
    - protein: $\alpha = 1 - C_{mg/ml} * 0.743/1000$

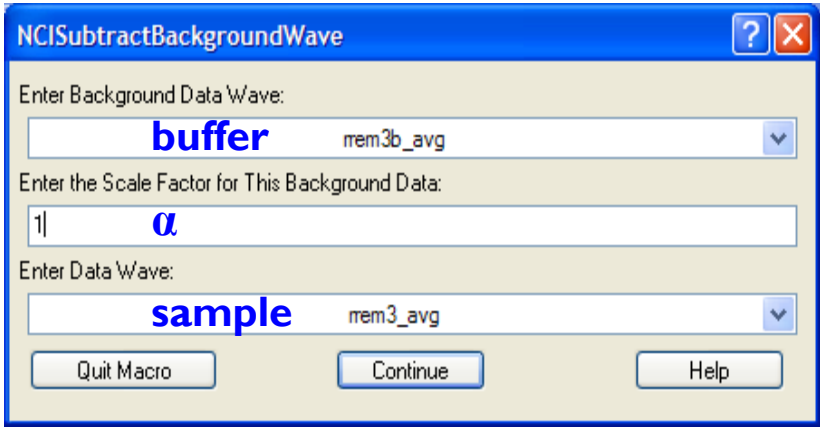
      Nucleic Acids: $\alpha = 1 - C_{mg/ml} * 0.54/1000$

      Widely used in literature

  - C. using WAXS as guide

At synchrotron beamline station, using A or B to quickly get scattering data from your molecule and do real-time data evaluation.

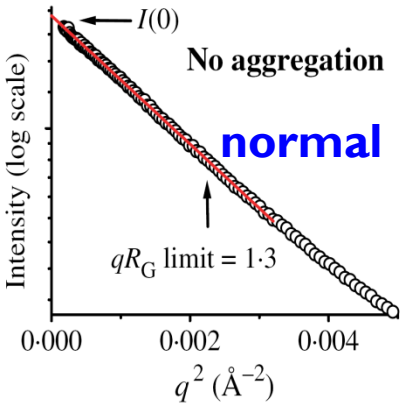




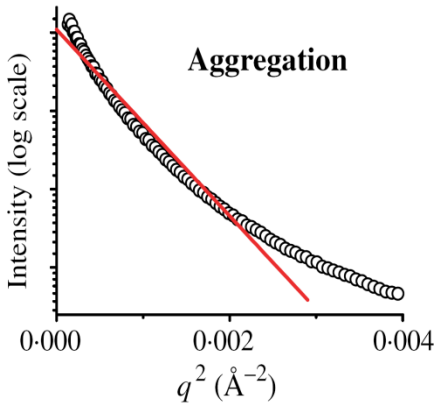
SAXS data evaluation

Guinier equation: $\ln[I(q)] = \ln[I(0)] - R_g^2 q^2 / 3$

q→ 0

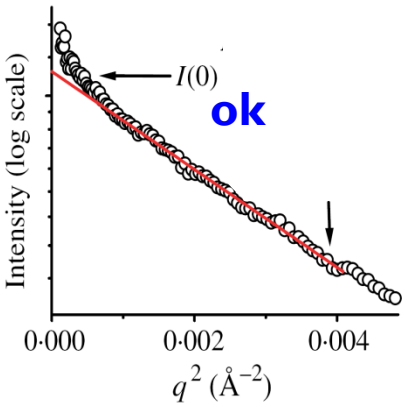


Mono-dispersed sample



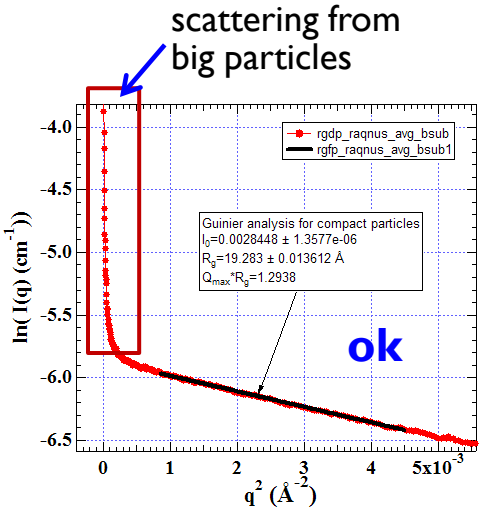
polydispersed (continuous size distribution) sample

Need to re-condition
Data unusable

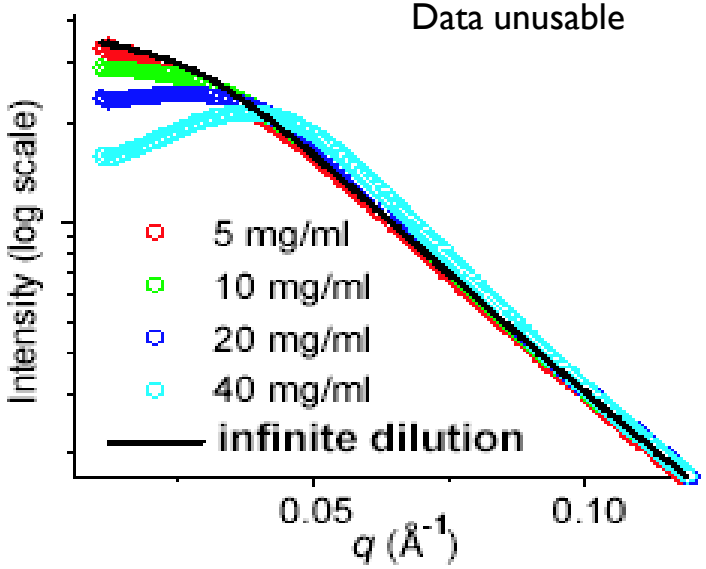


With small amount of aggregation

Data fixable



sample with big particles
Scattering from big particles: huge slope, drop fast
To remove: 0.22μm filter



- Repulsive electrostatic interactions among molecules
- Add salt to shield the interactions
- Structure factor often less obvious for dilute sample and higher salt solution
- Do a series of concentration
- Intensity extrapolation to with 0 structure factor

WAXS background subtraction

Why WAXS?

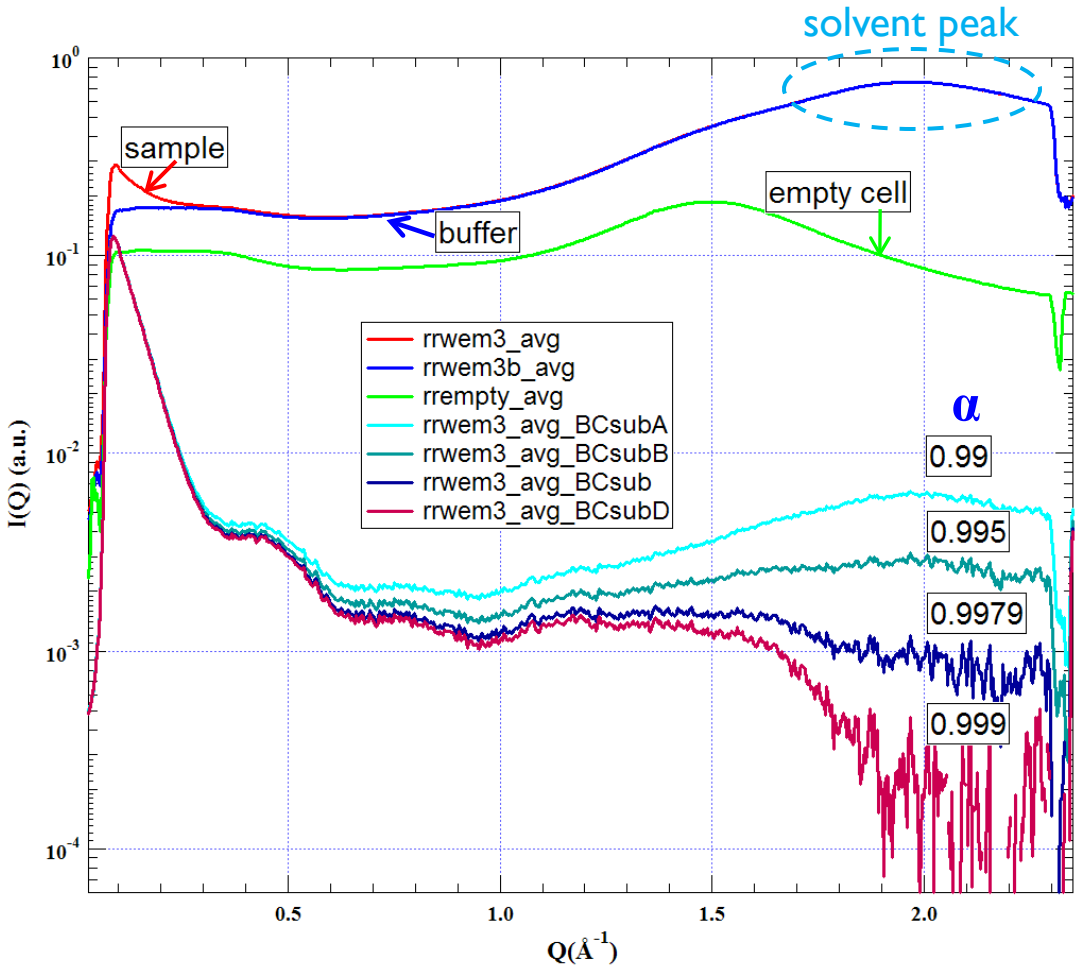
- Wider q range
- WAXS data is good guide for SAXS data subtraction

The total scattering from the sample solution comprises of those from biomolecules, buffer, and x-ray cell.

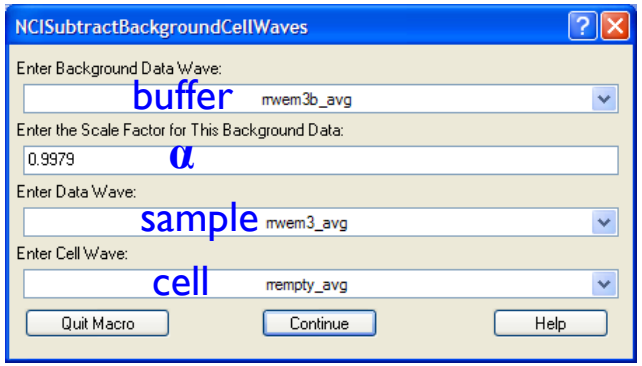
$$I_{molecule} = I_{solution} - \alpha\, I_{buffer} - (1-\alpha)\, I_{empty\_cell}$$

All Intensities have been normalized vs Io

Residual empty cell scattering becomes significant only for dilute samples



Watch the solvent peak (at ~2.0 Å^{-1}) disappear



under-subtraction
(significant amount solvent)

reasonable subtraction

over-subtraction
Derivative type shape

WAXS data guided SAXS background subtraction

How to?

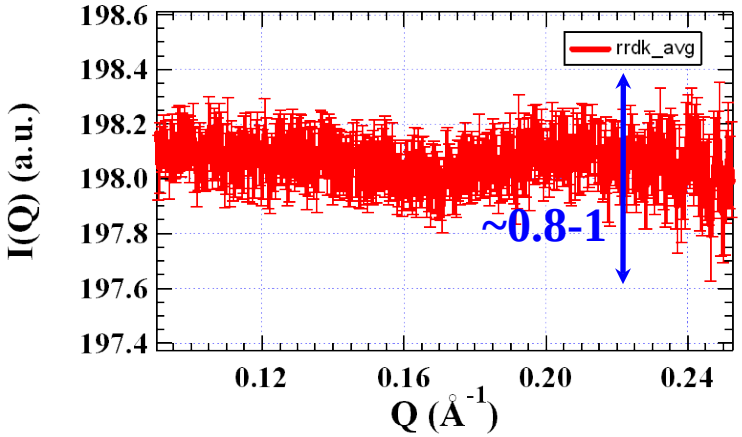
(1). SAXS: $I_{sample} = I_{solution} - \alpha * I_{buffer} - const$

α: account for contribution from buffer

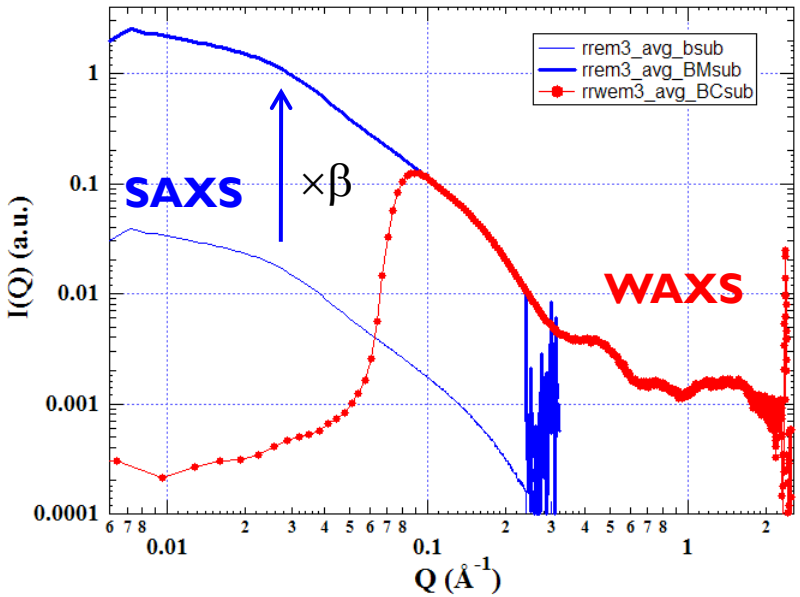
const: q-independent residual electronic noise and drift

(2). Match I^{waxs} and β I^{saxs} at overlapping q ranges:

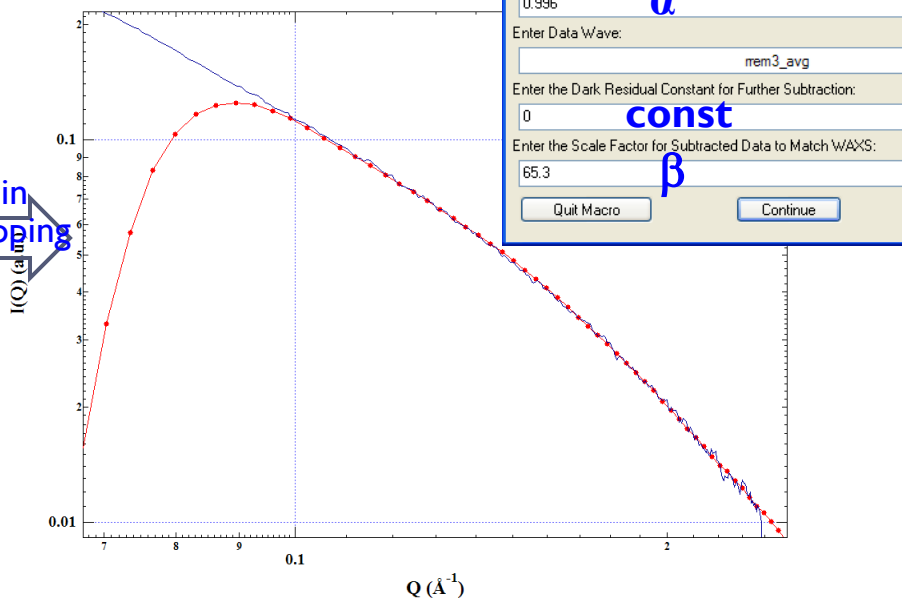
(3). Interactively tune α (**const** if necessary) and β to make SAXS and WAXS have best match.



Average of 10 dark (no x-ray) images.
Residual electronic noise (**const**) should not exceed ~2^0.5x1/Io≈ 2.7e-5



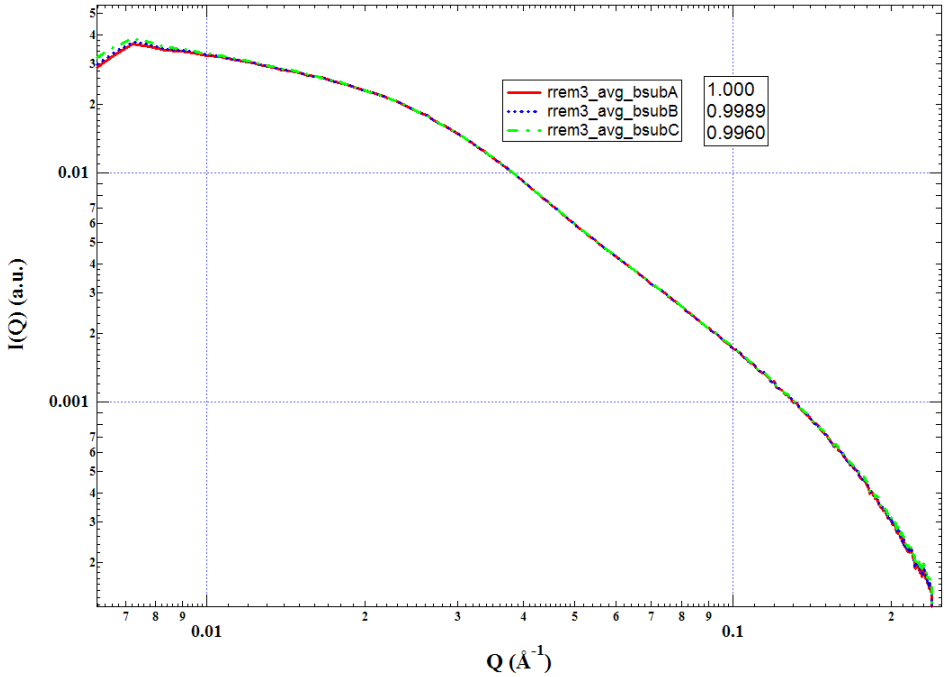
Zoom in overlapping region



SAXS background subtraction

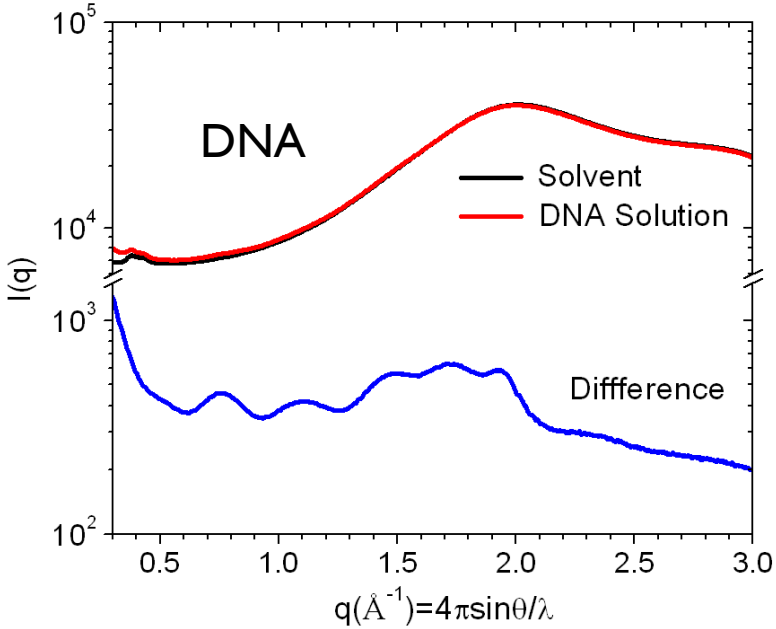
- SAXS background subtraction:
  - A. $I_{sample} = I_{solution} - I_{buffer}$
  - B. $I_{sample} = I_{solution} - \alpha * I_{buffer}$
    - protein: $\alpha = 1 - C_{mg/ml} * 0.743/1000$

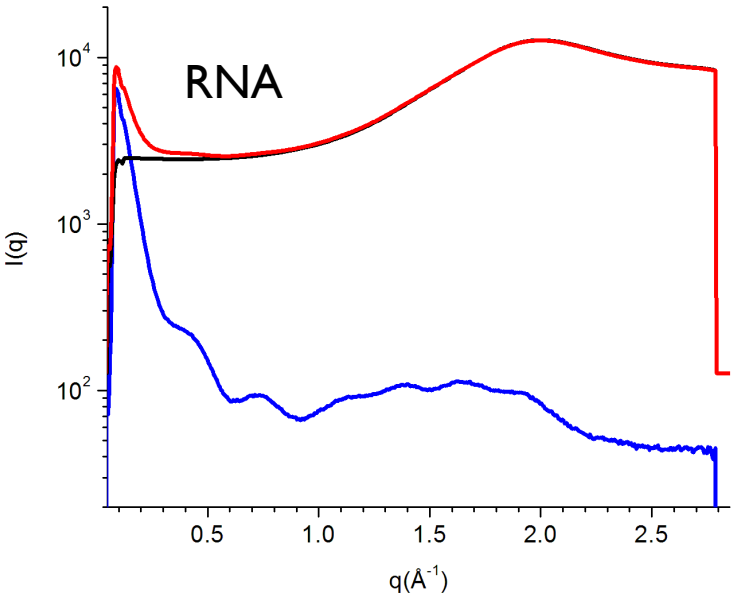
      Nucleic Acids: $\alpha = 1 - C_{mg/ml} * 0.54/1000$
  - C. using WAXS as guide

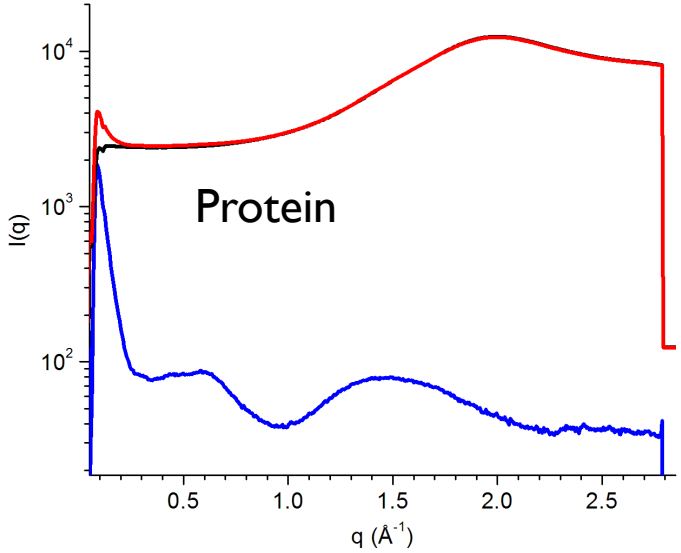


- The difference in Guinier region is often small, biggest difference often appears at the high q end, i.e., the slope at high q end.
- For this sample, SAXS data from three subtractions are very close.

Some good WAXS background subtraction



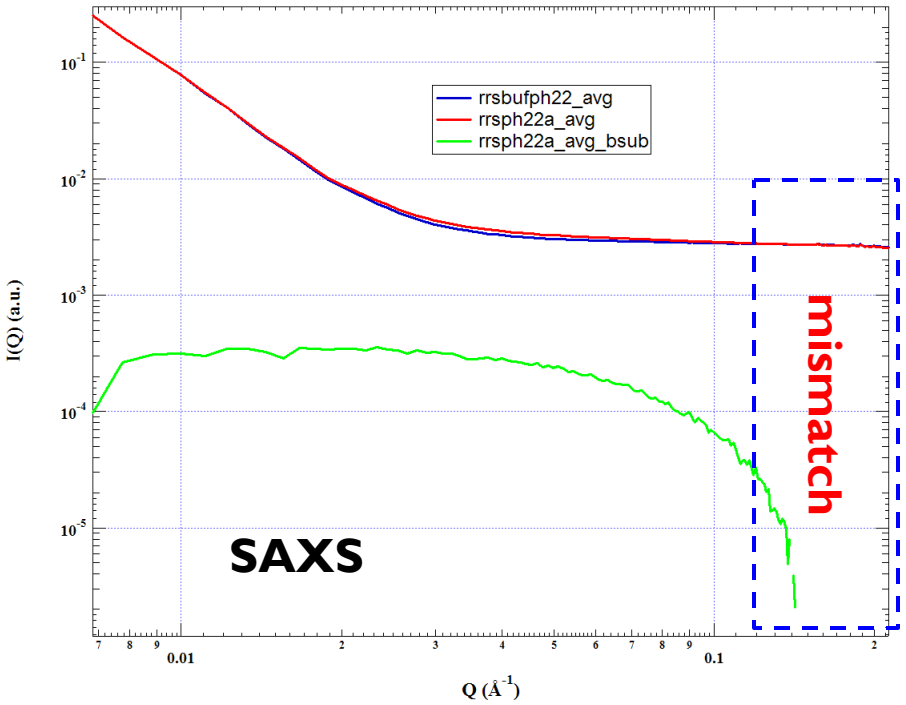


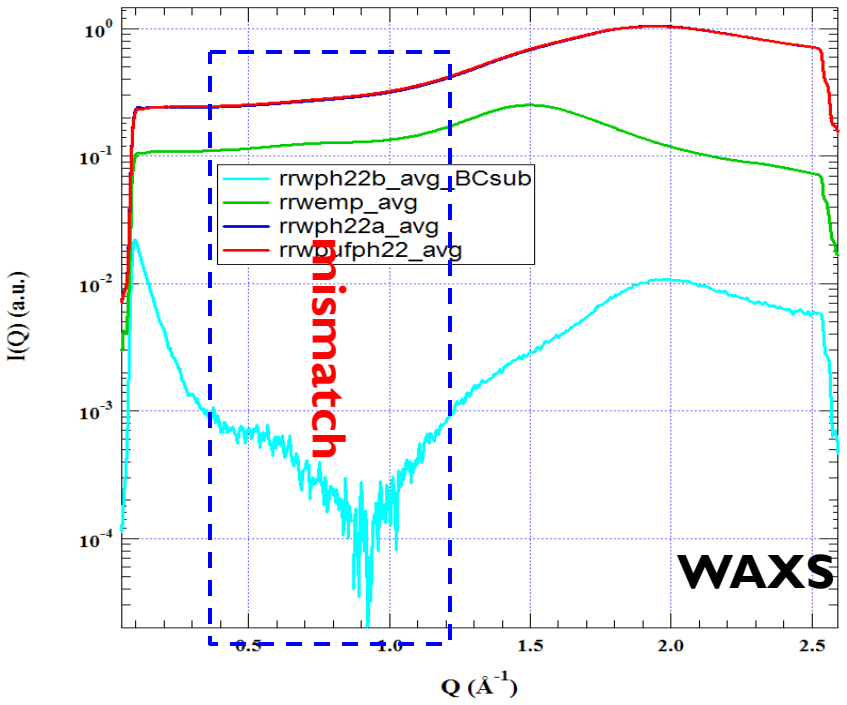


- No negative intensities
- No solvent peak
- Normal high angle feature
- Smooth baseline

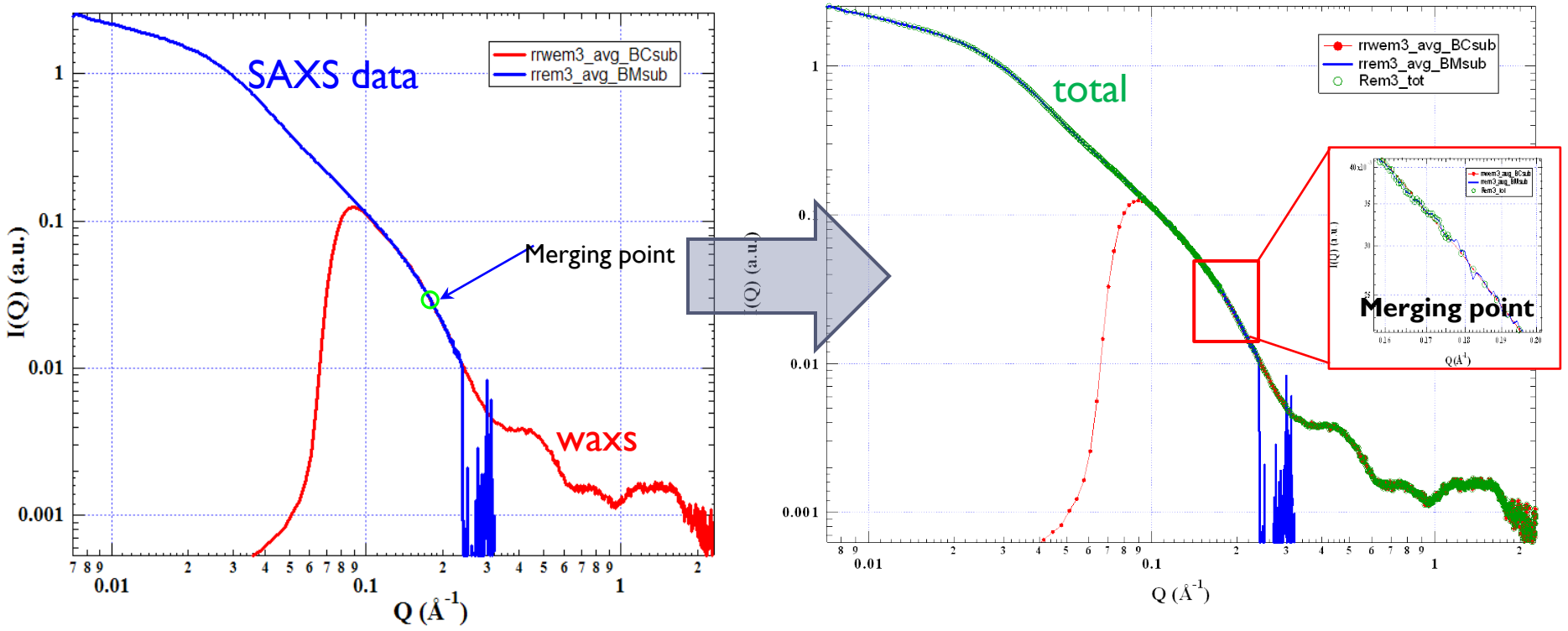
Data quality evaluation: background mis-match

Solution X-ray scattering measurements are very sensitive to the background match. Here is an example of a dilute sample with slightly background mis-match





SAXS and WAXS data merging



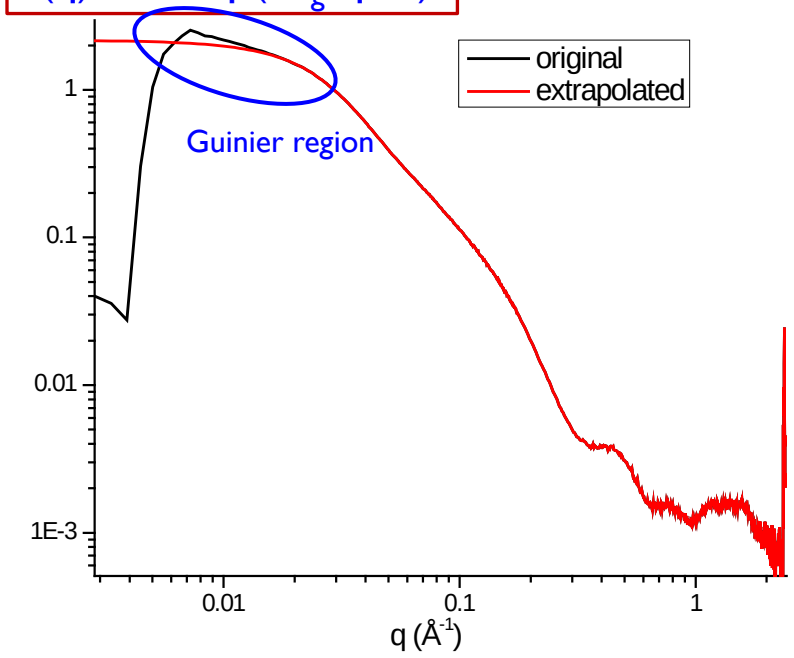
Choose merging point (q_M), then take data lower than q_M from SAXS data set (blue) and higher than q_M from WAXS data set (red) to form the new data set with complete q range (total, green).

Igor procedure available

Additional data manipulations

Data extrapolation q→0

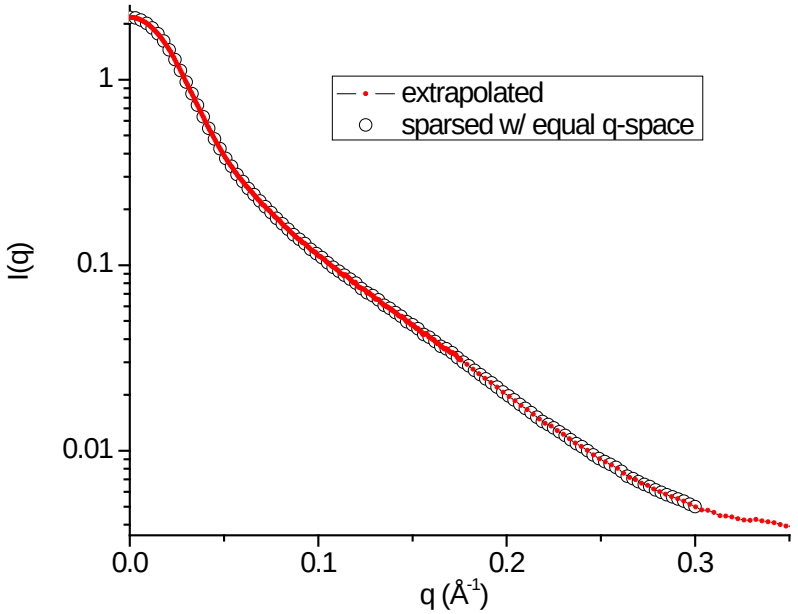
$I(q)=Io*\exp(-R_g^2q^2/3)$



For data with good Guinier plot, it is worthy to extrapolate data to q→0, particularly for large biomolecules.

Origin scripts available

Data points sparsing w/ equal q-space or equal ln(q)-space



The data point densities from SAXS and WAXS are different. Data with equal q-/ln(q)- space to receive same consideration while further data analysis

matlab scripts available