# Cairns Workshop: Combined Small Angle X-ray and Neutron Scattering with Biomolecular NMR

*The XXIVth International Conference on Magnetic Resonance in Biological Systems, Cairns, Australia, 2010*


***Organizers:***

**Terry Mulhern, Yun-Xing Wang**

***Lecturers:***

**Xiaobing Zuo, Alex Grishaev, Andrew Whitten, Charles Schwieters**


***Sebel Cairns Hotel, Cairns Australia***

***9:00 am- 5:00 pm, Saturday***

***August 21, 2010***


Structural Biophysics Laboratory

CCR, NCI, NIH, USA

# Part One

# Fundamentals and Experimental Aspects of Small Angle X-ray Scattering

**Xiaobing Zuo (NCI-Frederick )**

**Alex Grishaev (NIDDK)**

# 1. Fundamentals of solution x-ray scattering

## A. General aspects of x-ray scattering

- History
- Why x-ray scattering
- Study scope
- X-ray  interference pattern
- Solution x-ray scattering experiments

## B. Physics of X-ray Scattering

- Scattering and interference phenomena
- Momentum transfer and phase change
- X-ray form factor and structure factor
- X-ray scattering contrast

## C. Solution x-ray scattering calculations from atomic coordinates

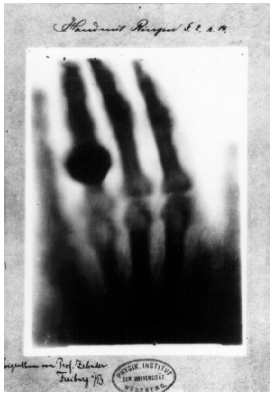
- Solvent contribution and x-ray contrast
- Debye equation
- Fast calculation methods

## A. General aspects of x-ray scattering

- History
- Why x-ray scattering
- Study scope
- X-ray  interference pattern
- Solution x-ray scattering experiments

# History of X-ray and solution scattering

- X-ray discovered in 1895



Wilhelm Conrad Röntgen 1845-1923
Nobel Prize in Physics in 1901

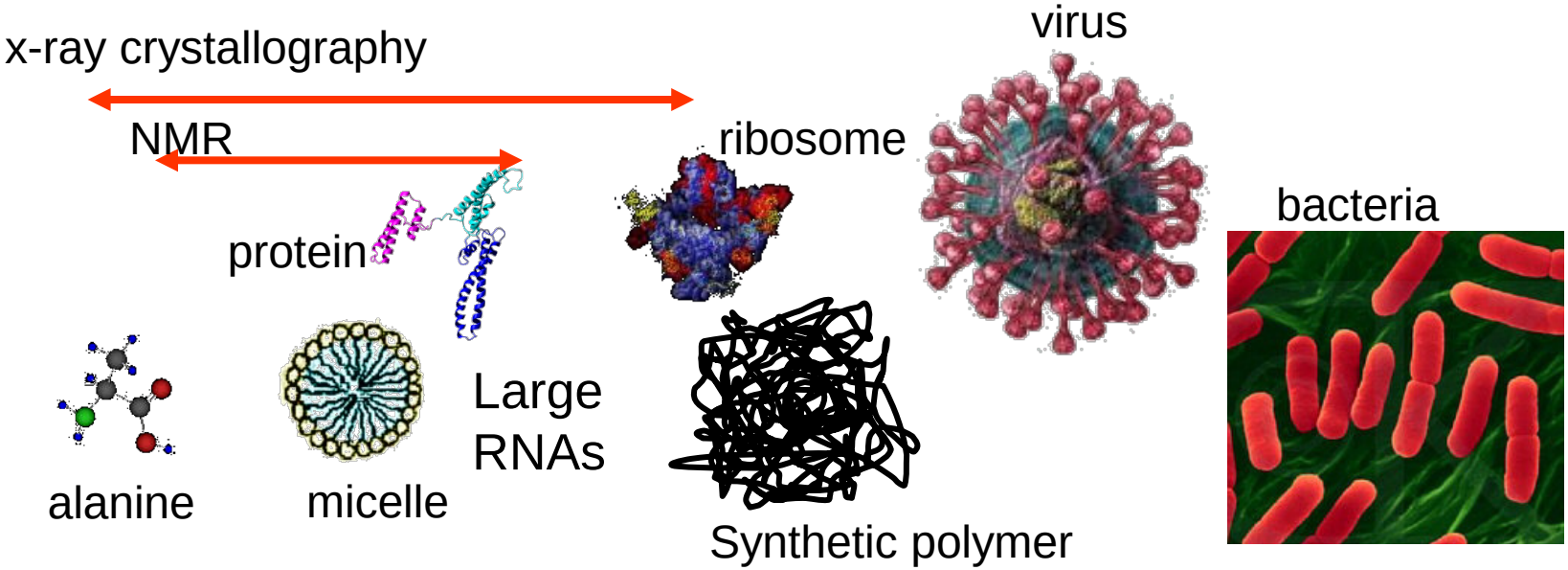
- First solution X-ray scattering performed in late 1930s
  - Measure molecular weight, size (1936)

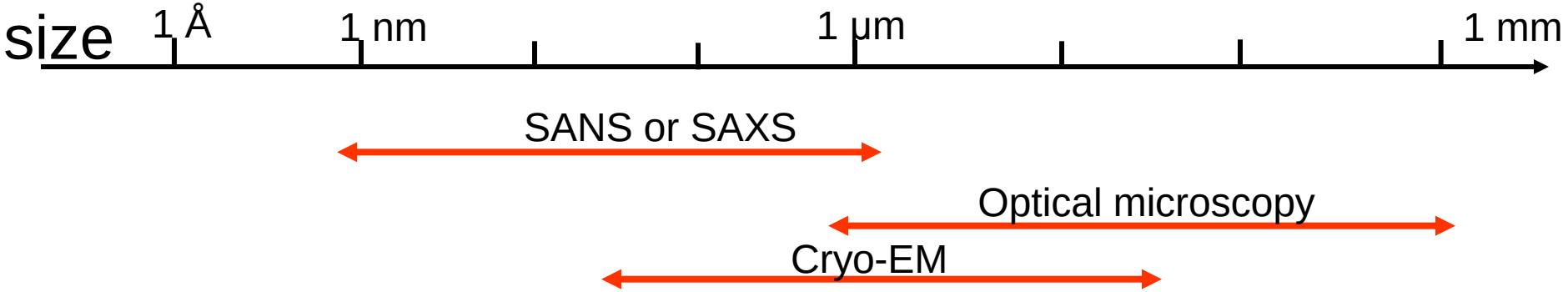
- For a long time, applications had been limited by available X-ray sources, detectors and data analysis methods

# Why solution x-ray scattering

- Becomes an all-purpose structural tool
- Characterizes a variety of samples:
  - From rigid molecules to totally unstructured molecules
  - Flexible parts also contribute signals (vs crystallography, NMR)
  - Larger molecules scatter stronger (vs solution NMR)
- Studies biomolecules under most physiologically relevant conditions
- requires relative smaller amount of sample (vs NMR and crystallography)
- Fast data acquisition and analysis
- Covers a wide size range of biomolecules/assembly

# Scales of various methods

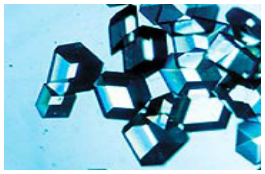
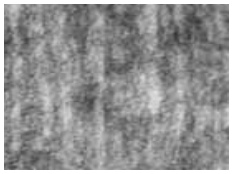
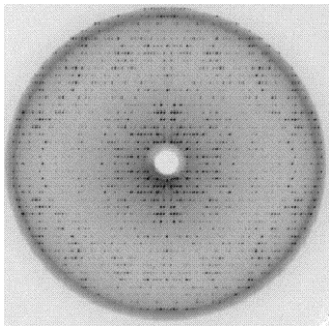
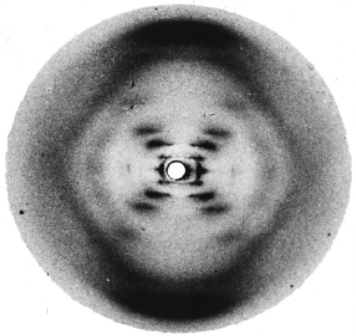
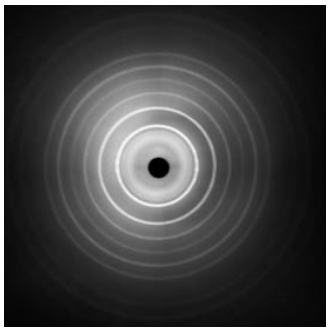
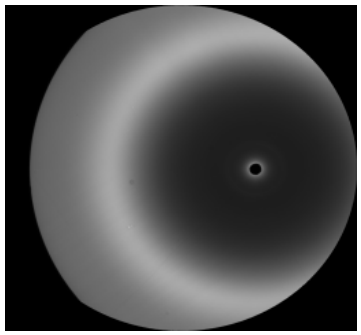
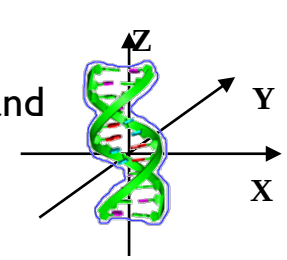
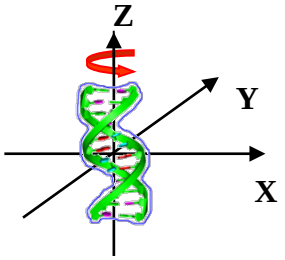
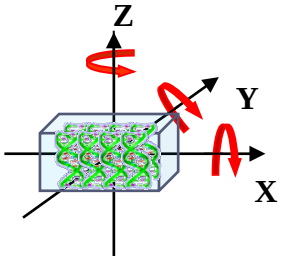
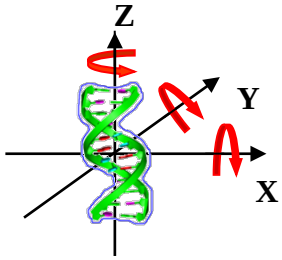



The scope of small angle X-ray scattering in terms of spatial dimension covers ~nm to ~μm ranges, perfectly suitable for biomolecular structrual study.

# 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.

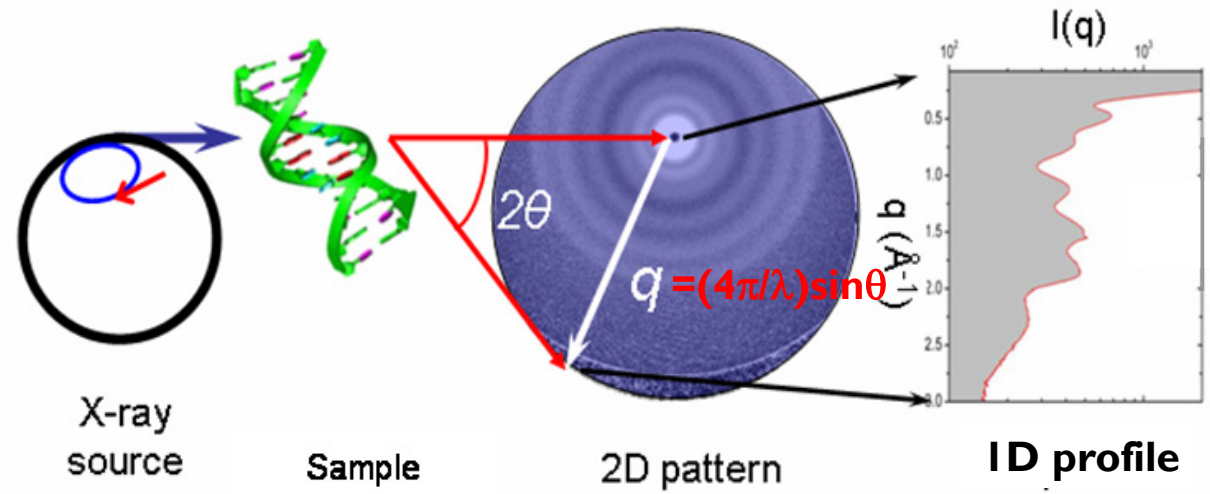
| | Single Crystal | Fiber/ Membrane | Powder/ Micro-crystals | Solution |
|---|---|---|---|---|
| Sample states |  |  |  |  |
| Interference pattern |  |  |  |  |
| matrix symmetry and molecular freedom |  |  |  |  |



Matrix symmetry: high ← low

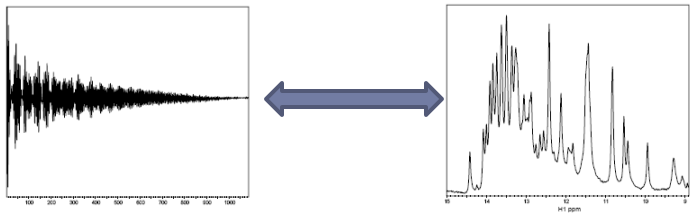
Molecular freedom: low → high

# Solution X-ray Scattering Experiment




Scattering profile
I(q) vs q

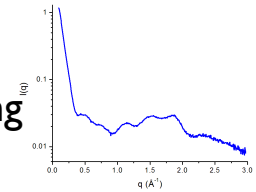
NMR



Time domain

frequency domain

scattering



Reciprocal space

Real space

**Tasks today:**

How scattering works

How to get structural information

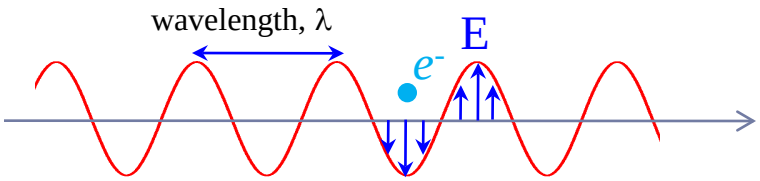
How to combine scattering with NMR

## B. Physics of X-ray Scattering

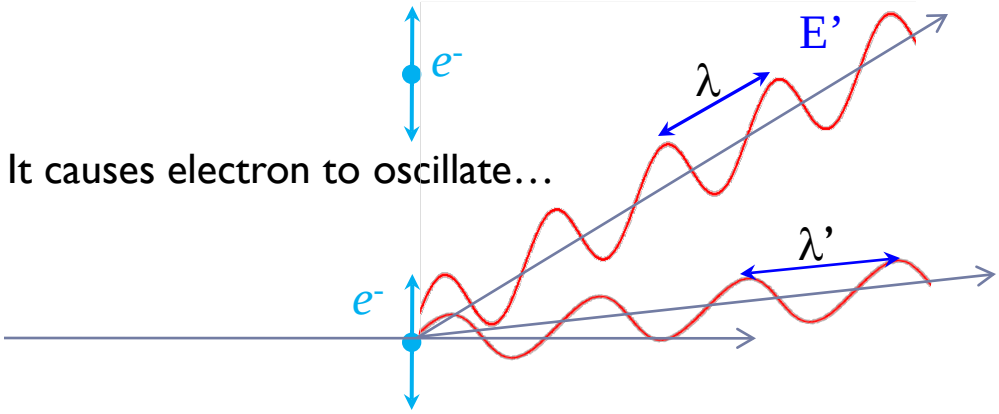
- **Scattering and interference phenomena**
- **Momentum transfer and phase change**
- **X-ray form factor and structure factor**
- **X-ray scattering contrast**

# X-ray scattering by a free electron

X-ray is traveling wave…




X-ray electric field hits electron…




It causes electron to oscillate…

The oscillating electron radiates secondary X-ray!
This is scattering!

X-ray Wave function:

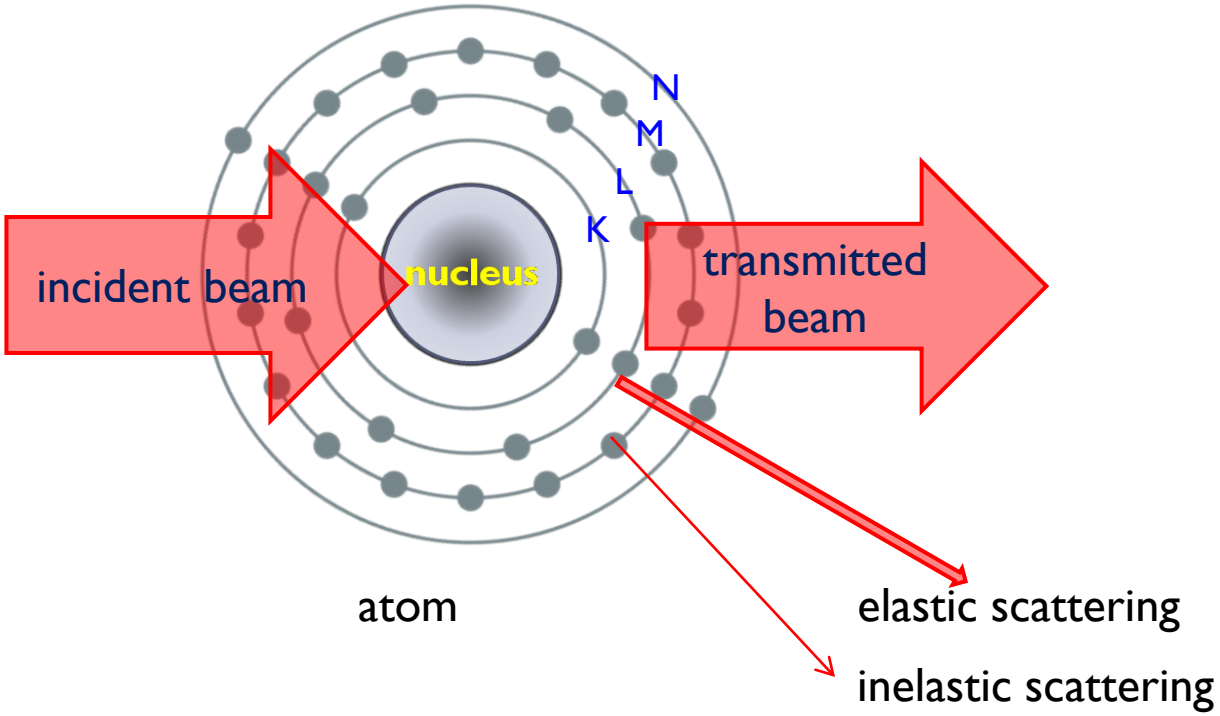
$$E(t) = E_0 e^{i(2\pi\nu t + \phi_0)}$$

frequency: $\nu = c/\lambda$
Initial phase: $\phi_0$
Phase change: $\Delta\phi = 2\pi d/\lambda$ (radian)
after travelling distance $d$.

Elastic scattering: $\lambda \rightarrow \lambda$ not change
Inelastic (Compton) scattering: $\lambda \rightarrow \lambda'$ 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




- Atoms scatter x-ray by electron cloud, elastically and inelastically.
- Inelastic scattering only add background to scattering data because they don't interfere with elastically scattering x-ray due to different wavelength!
- Inelastic scattering is very weak and often ignorable for atomic scattering.

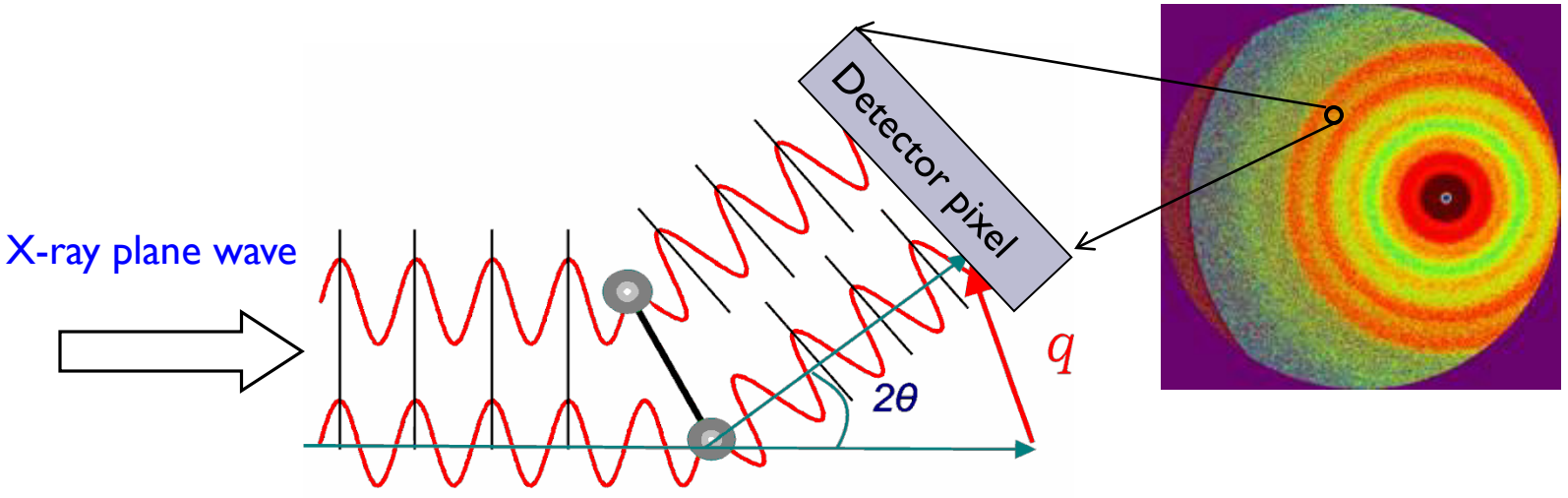
# X-ray measures distance/structure by interference

- Interference in pond:

  Interference pattern of water waves codes distance between sources

- Elastic scattering Interference:

  Elastic X-ray scattering interferes.
  Before hitting 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

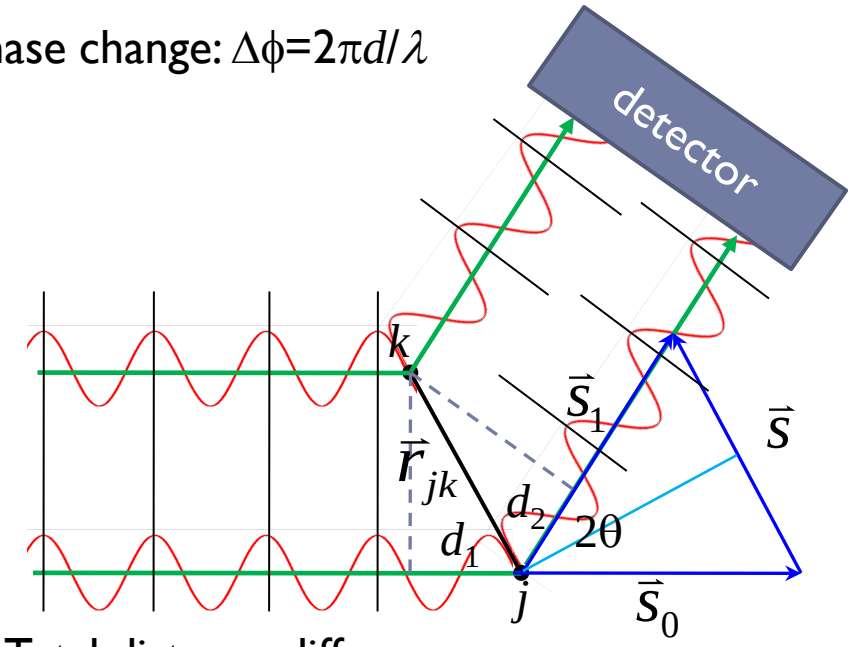
# Scattering and Interference at various levels

| scatterer | description | information | scattering length/factors |
|---|---|---|---|
| electron | Smallest scattering entity | Electron pair distance, electron density theoretic experiment | $f_e$=2.8179x10-13cm$^{-1}$ |
| atom | Smallest entity for structure | Atom-pair distances in molecules, molecular structure | Form factor: $f(q)$ $f(\mathbf{q}) = \int_V f_e \rho(\mathbf{r}) \exp(i\mathbf{q}\cdot\mathbf{r}) d\mathbf{r}$ |
| molecule | | Inter-molecular/particle interaction | Structure factor: $S(q)$ $S(q) = 1 + \rho \int_0^\infty 4\pi^2 (g(r) - 1) \frac{\sin qr}{qr} dr$ |

# Momentum transfer and Phase change

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

Phase change: $\Delta\phi = 2\pi d/\lambda$




unit incident wavevector $\mathbf{s}_0$

unit scattering wavevector $\mathbf{s}_1$

wave vector change: $\mathbf{s}$

$$\mathbf{S} = \mathbf{S}_1 - \mathbf{S}_0$$

$$|\mathbf{S}| = 2\sin\theta$$

Total distance difference:

$$d = d_1 + d_2 = -\mathbf{s}_0 \cdot \mathbf{r}_{jk} + \mathbf{s}_1 \cdot \mathbf{r}_{jk}$$

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

$$= (\mathbf{s}_1 - \mathbf{s}_0) \cdot \mathbf{r}_{jk} = \mathbf{s} \cdot \mathbf{r}_{jk}$$

Definition: momentum transfer

$$\mathbf{q} = (2\pi/\lambda)\mathbf{S}$$

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

Phase change: (radian)

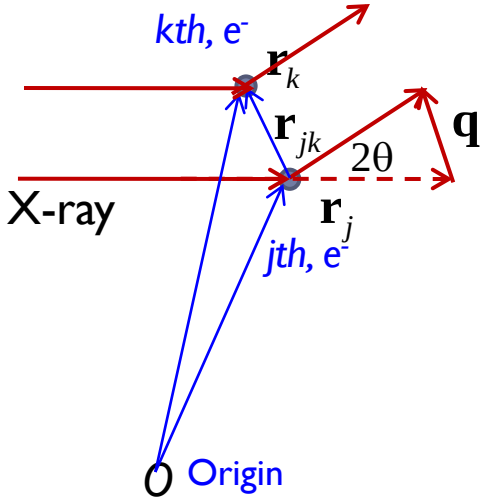
$$\Delta\phi = 2\pi * \frac{d}{\lambda} = \mathbf{q} \cdot \mathbf{r}_{jk}$$

# Scattering by an electron pair

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

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

additional phase change




Total amplitude at $\mathbf{q}$ of scattered X-ray by electrons $j$ & $k$:

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

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

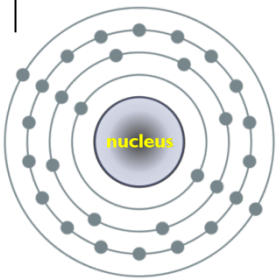
$$I(\mathbf{q}) = \left|A_{tot}(\mathbf{q})\right|^2 = A_{tot}(\mathbf{q})A_{tot}^*(\mathbf{q})$$
$$= f_e^2 \left|\exp(i\mathbf{q}\cdot\mathbf{r}_j)\right|^2 \times \left|1 + \cos(\mathbf{q}\cdot\mathbf{r}_{jk}) + i\sin(\mathbf{q}\cdot\mathbf{r}_{jk})\right|^2$$
$$= f_e^2 \times 1 \times \left[\underbrace{1+1}_{\text{from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

- 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

# X-ray scattering form factor/scattering length

In principle, scattering of an object with N e- can be calculated as:

$$I(\mathbf{q}) = \left| \sum_{j=1}^{N} f_e \exp(i\mathbf{q}\cdot\mathbf{r}_j) \right|^2 \xrightarrow{\substack{M \text{ relative independent,} \\ \text{stable units (atoms)}}} I(\mathbf{q}) = \left| \sum_{k=1}^{M} f_k(\mathbf{q}) \exp(i\mathbf{q}\cdot\mathbf{r}_k) \right|^2$$




For a unit / entity with an ensemble of multiple electrons:

Define form factor $f(\mathbf{q})$:

$$f(\mathbf{q}) = \sum_{j=1}^{n} f_e \exp(i\mathbf{q}\cdot\mathbf{r}_j) \equiv \int_V f_e \rho(\mathbf{r}) \exp(i\mathbf{q}\cdot\mathbf{r}) dV$$

$$I(\mathbf{q}) = f^2(\mathbf{q})$$

Scattering form factor/length

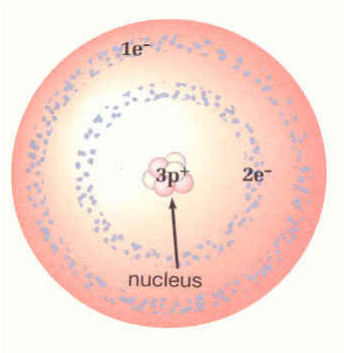
$$\left| \frac{f(\mathbf{q})}{f_e} \right|^2 = \left| \frac{I(\mathbf{q})}{I_e} \right|$$

- X-ray scattering form factor (or scattering length) describes the scattering capability of an object along **q** comparing to a free electron.
- For a electron, $\rho(\mathbf{r})=\delta(\mathbf{r}-\mathbf{r'})$. 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 (such as atoms) with radialy uniform electron density $\rho(r)$:

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

# Atomic form factor



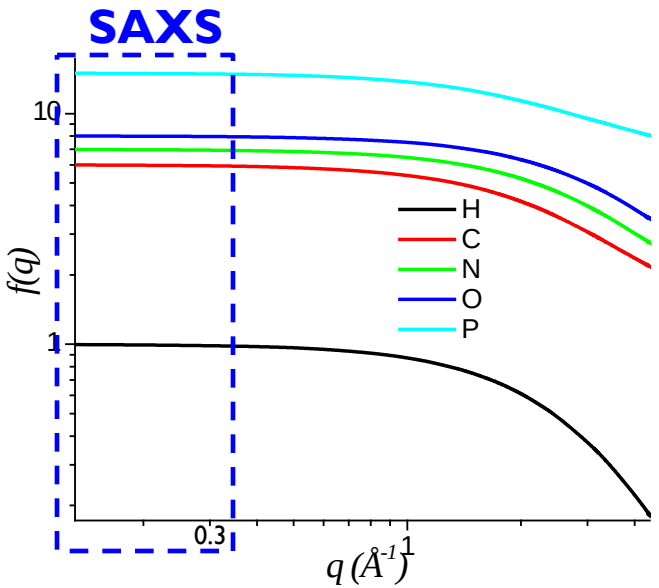

Atomic electron cloud

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

For an atom with electron density $\rho(r)$:

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

ρ(r) were obtained from quantum chemistry 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**

# 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 uniform electron density ($\rho(r)=1$)and radius $R$:

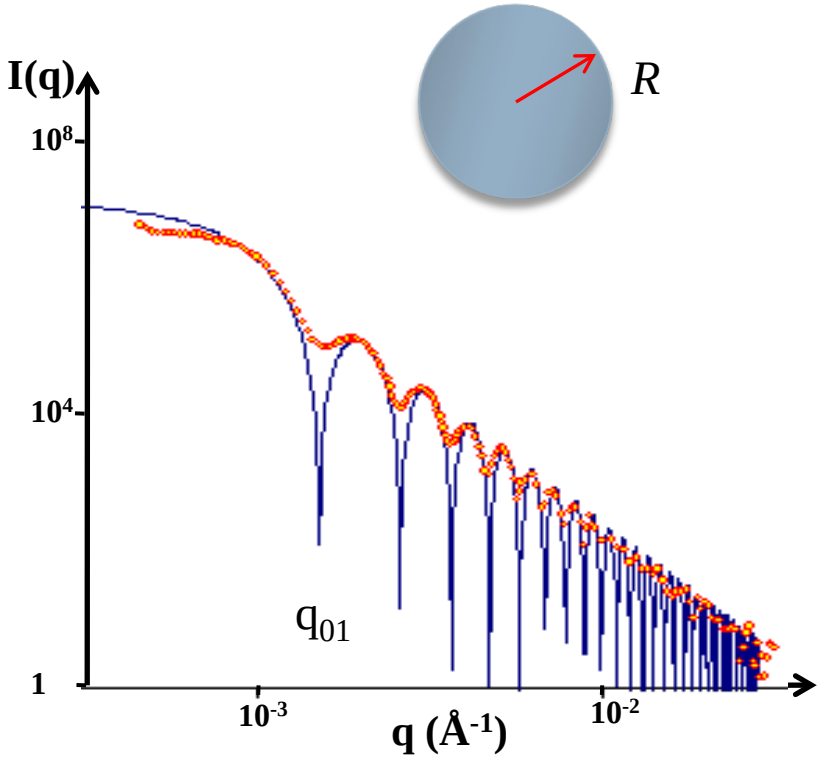
$$f(q) = 4\pi \int \rho(r) r^2 \frac{\sin(qr)}{qr} dr$$

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

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

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.

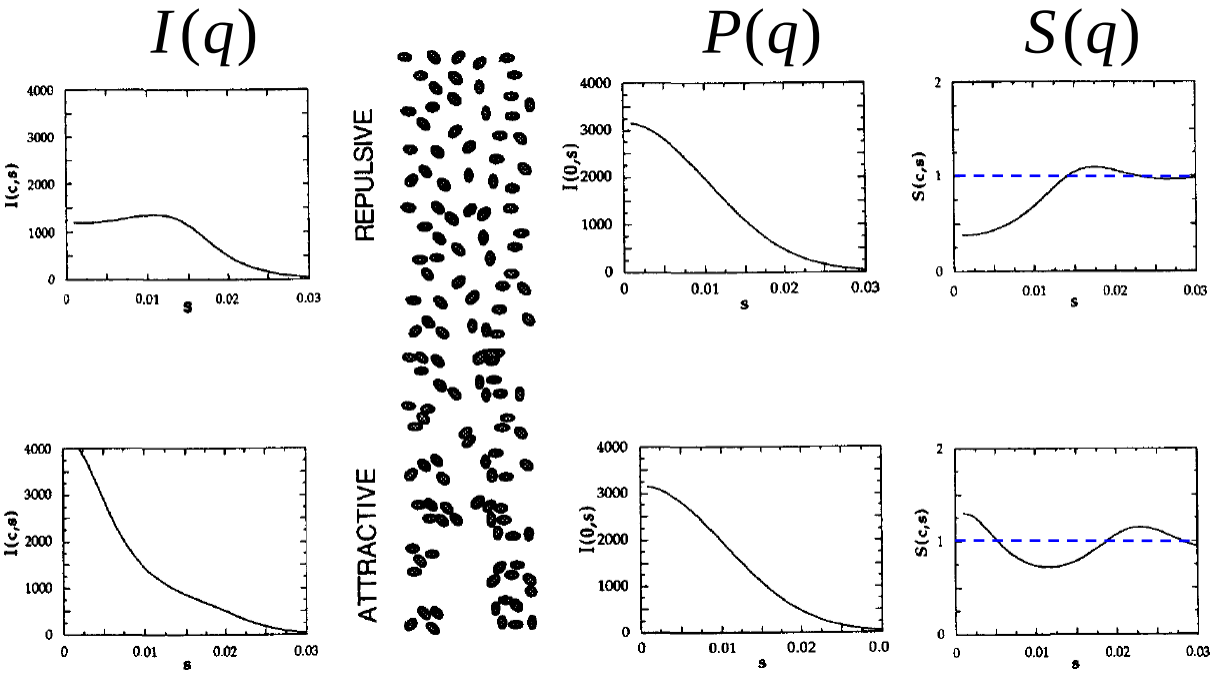
Figure modified from: Lambard, Lessieur and Zemb, J. Phys. I France, 1992, (2) 1191

# structural factor

In solution X-ray scattering, structural factor describes the interactions among macromolecules or particles.

$$I(q) = P(q) * S(q)$$

I(q): apparent scattering profile
P(q): scattering profile in absence of inter-particle interaction
S(q): structural factor




$$S(q) = 1 + \rho \int_0^\infty 4\pi^2 (g(r) - 1) \frac{\sin qr}{qr} dr$$

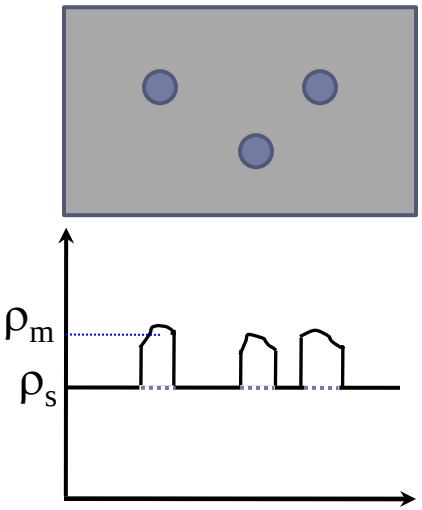
Auto-correlation function of the particle distribution

- Structure factor is a undesired effect.
- One should try to reduce or eliminate the inter-particle interactions.
- There are two widely used ways to deal with structural factor: measure samples with a serial of concentrations and add salt to repulsive cases.


Ducruix, A., Guilloteau, J. P., Riès-Kautt, M., & Tardieu, A. (1996) J. Crystal Growth 168, 28-39.
Svergun, DI, & Koch, MH, (2003) Rep. Prog. Phys. 66, 1735–1782

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

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




Solution sample scattering:

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

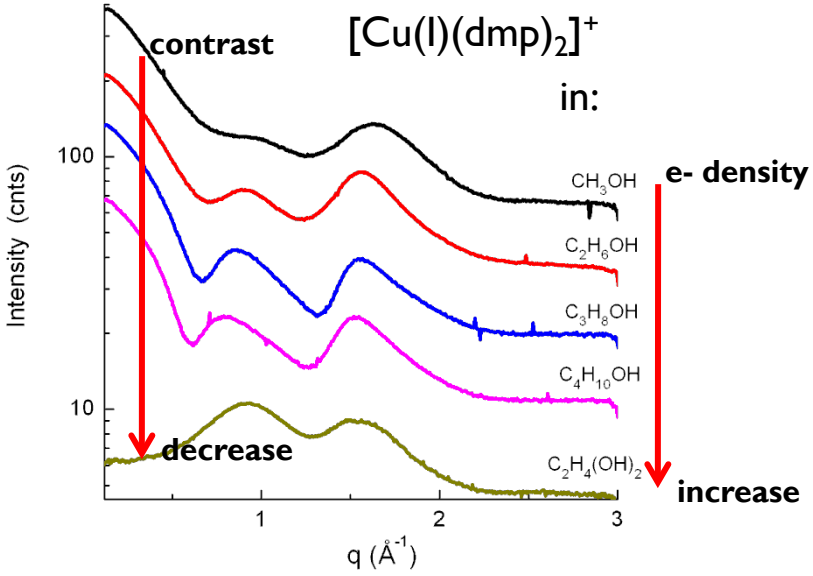
What X-ray scattering measures:

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

excess electron density/scattering length against solvent/buffer

Solution X-ray scattering contrast matching:

- Increasing solvent electron density, X-ray scattering contrast match occurs at average electron density (large spatial scale) [I(q~0) → 0], but electron density difference still exist locally / at high spatial resolution scale.
- X-ray contrasting possibly renders additional structural information. But there is very little room for playing x-ray contrast for aqueous biomolecular samples.




Tiede, et al, unpublished data

## C. Solution x-ray scattering calculations from atomic coordinates

- **Solvent contribution**
- **Debye equation**
- **Fast calculation methods**

# X-ray scattering

In vacuo, for a molecular with fixed orientation:

$$I(\mathbf{q}) = \left| \sum_{j=1}^{N} A_j(\mathbf{q}) \right|^2 = \left| \sum_{j=1}^{N} f_j(q) \exp(i\mathbf{q}\cdot\mathbf{r}_j) \right|^2$$

In solution,

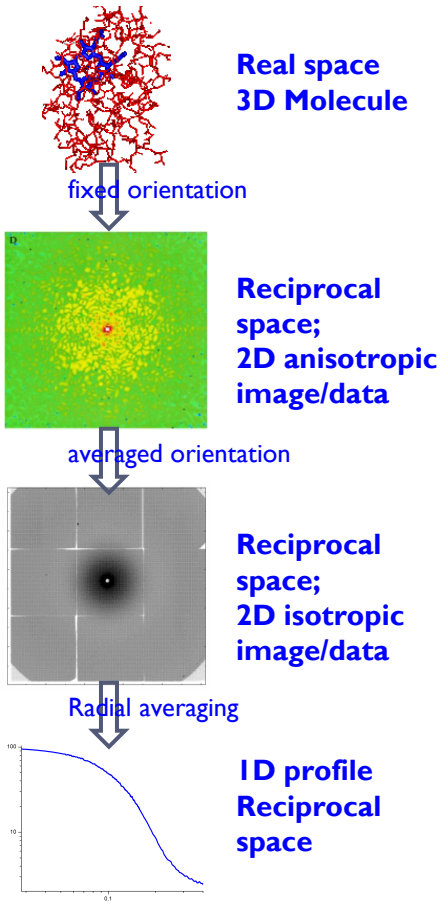
- Molecular orientation
  - In solution, x-ray sees all orientations

$$I(q) = \langle I(\mathbf{q}) \rangle_\Omega \qquad \langle \exp(i\mathbf{q}\cdot\mathbf{r}) \rangle_\Omega = \frac{\sin(qr)}{qr}$$

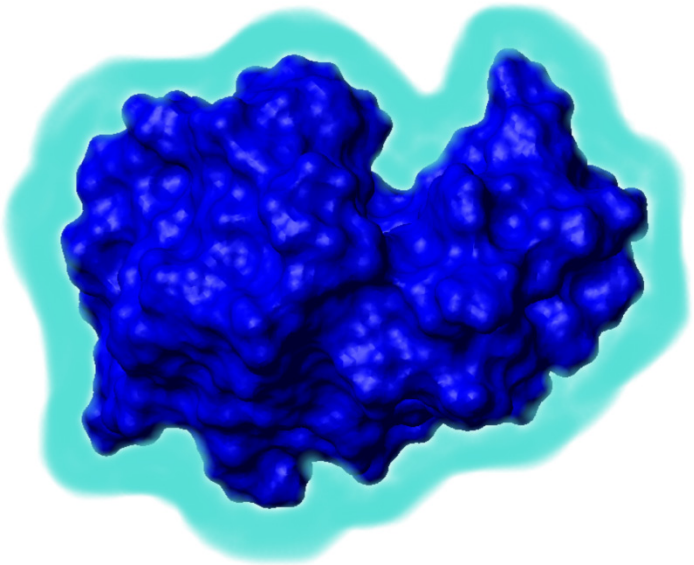
average over all orientation

Loss of the direction the momentum transfer/ angular term

- Solvent contribution



# Solution x-ray scattering calculation from atomic coordinates



Scattering occurs on the contrast between the macromolecule and the displaced solvent. A surface layer of solvent results in additional scattering.

$$I(q) = \left\langle \left| A_m(\mathbf{q}) - A_s(\mathbf{q}) + \delta\rho A_l(\mathbf{q}) \right|^2 \right\rangle_\Omega$$

Scattering intensity - average over all orientations Ω.

1. Scattering amplitude *in vacuo*

$$A_m(\mathbf{q}) = \sum_{j=1}^{N} f_j(\mathrm{q}) \exp\left(i\mathbf{q}\mathbf{r}_\mathbf{j}\right)$$

2. Scattering amplitude from the excluded volume using *dummy solvent* approximation

$$A_s(\mathbf{q}) = \sum_{j=1}^{N} g_j(\mathrm{q}) \exp\left(i\mathbf{q}\mathbf{r}_\mathbf{j}\right)$$

Dummy atom model with Gaussian function type form factor

$$g_j(q) = G(q)\rho_s V_j \exp\left(-\frac{q^2 V_j^{2/3}}{4\pi}\right)$$

Expansion factor

3. Additional scattering from solvent layer

$\delta\rho\, A_l(\mathbf{q})$ where δρ is the electron density difference between the layer solvent and the bulk solvent and the second term is the form factor of the surface layer envelope.


Fraser, R., Macrae T. P., Suzuki, E. , *J. Appl. Cryst. (1978). 11, 693-694*

Svergun, D., Barberato, C., & Koch, M. H. J. (1995) *J. Appl. Cryst.* **28,** 768-773.

# Debye formula

$$I(q) = \left\langle \left| \sum_j (f_j(q) - g_j(q)) \exp(i\mathbf{q}\mathbf{r}_j) + \delta\rho A_l(\mathbf{q}) \right|^2 \right\rangle_\Omega$$


Petrus J.W. Debye (1884-1966)
Nobel Prize in Chemistry, 1936

Atomic apparent form factor / contrast :

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

atomic form factor in vacuum

form factor of excluded solvent

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

$$I(q) = \langle I(\mathbf{q}) \rangle_\Omega = \left\langle \left| \sum_{j=1}^{N} A_j \exp(i\mathbf{q}\cdot\mathbf{r}_j) \right|^2 \right\rangle_\Omega$$

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

Debye Formula

Individual contribution

interference

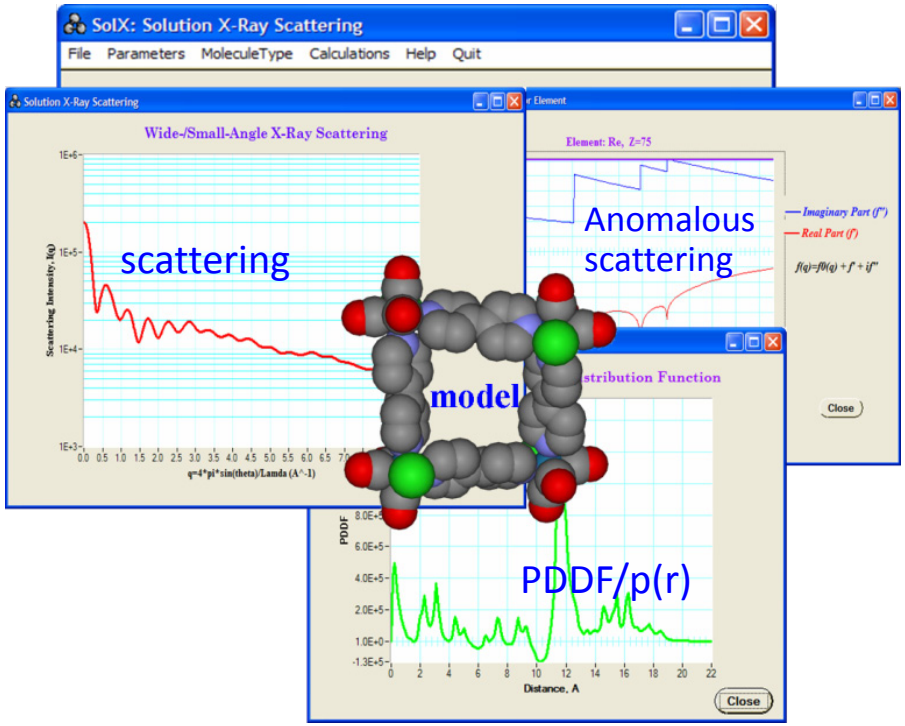
Atom pair distance / Structural information

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

# SolX: Solution X-ray scattering simulation software

SolX:

- Implements Debye equation, Complexity $\propto N^2$
- Indirectly deal with solvation layer contribution
- Coordinate-based calculations of solution x-ray scattering curve, PDDF, anomalous scattering data
- Designed for biomolecules and supramolecules with a flexible molecular dictionary
- Recognizes common biomolecules including proteins, nucleic acids and complexes.
- Easy to extend to user-defined molecules such as ligands.




$$I(q) = \sum_{j=1}^{N}\sum_{k=1}^{N} A_j A_k \frac{\sin(qr_{jk})}{qr_{jk}}$$

$$g_j(q) = G(q)\rho_s V_j \exp\left(-\frac{q^2 V_j^{2/3}}{4\pi}\right)$$

Expansion factor

expanding recognizable molecular scope by extending atom conversion map (SolxAtomMap.txt):

```
// Valine                          (VAL)
ATOM   OXT    OH      VAL     PRT
ATOM   NT     NH2     VAL     PRT
ATOM   N      NH      VAL     PRT
ATOM   CA     CH      VAL     PRT
ATOM   C      C       VAL     PRT
…

// Undetermined                    (UNK)
ATOM   O      H2O     HOH     PRT
ATOM   Zn     Zn      HOH     PRT

// HEM
ATOM   FE     Fe2     HEM     PRT
ATOM   NA     N       HEM     PRT
ATOM   NB     N       HEM     PRT
ATOM   NC     N       HEM     PRT
ATOM   ND     N       HEM     PRT
…
```

Atoms in regular biomolecules

User defined heteroatoms

User defined ligands/unusual residues


Available on request:
X. Zuo: xiaobing.zuo@gmail.com
D. Tiede: tiede@anl.gov

# Fast calculation methods

**One time calculation:**
SolX (Debye equation): HIV protease (22kDa, ~1550 non-H atoms), 300pts, 2.0GHz single cpu, 68 seconds

**Many-time calculation:**
Scattering profile calculation is computationally very expansive. The computation time depends on the number of atoms (N) in the molecule, number of q-value points ($N_p$), orientation averaging order, etc.

$$I(q) = \langle I(\mathbf{q}) \rangle_\Omega = \left\langle \left| \sum_{j=1}^{N} A_j \exp(i\mathbf{q} \cdot \mathbf{r}_j) \right|^2 \right\rangle_\Omega$$

Complicity

$\propto N^2 * N_p$

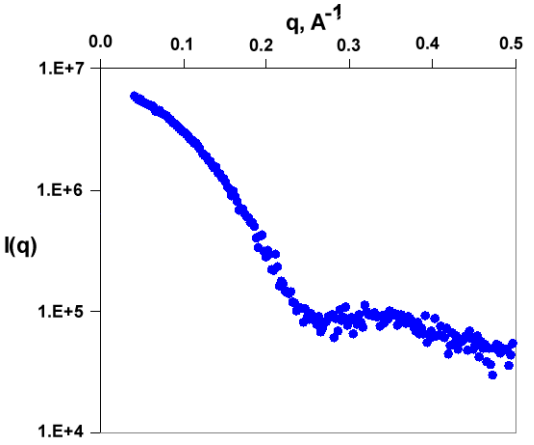
$$I(q) = \langle I(\mathbf{q}) \rangle_\Omega = \left| \left\langle \sum_{j=1}^{N} A_j \exp(i\mathbf{q} \cdot \mathbf{r}_j) \right\rangle_\Omega \right|^2$$

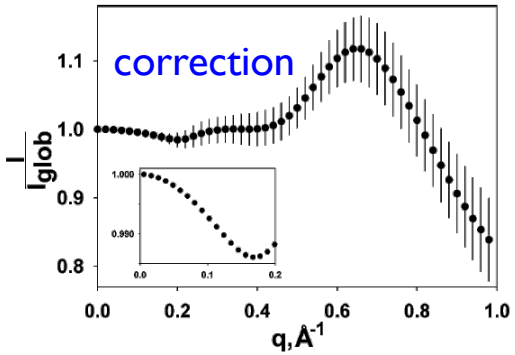
$\propto N * N_p$

# Fast SAXS evaluation strategies in MD refinement

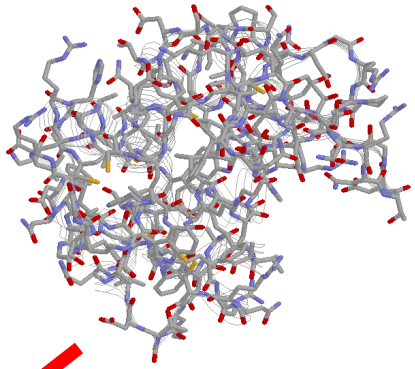
Cost function: $$\chi^2 = \frac{1}{N_q - 1} \sum_{k=1}^{N_q} \left[ \frac{c_k I_{calc}(q_k) - I_{expt}(q_k)}{\sigma(q_k)} \right]^2$$

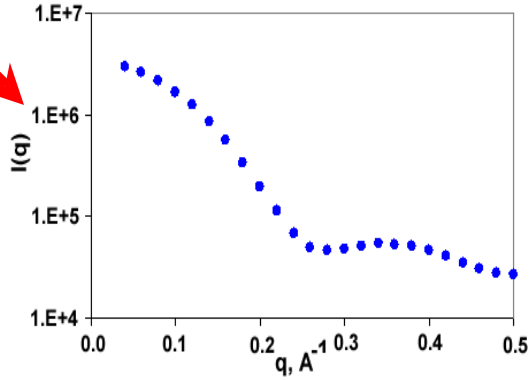
The problem : number of operations ~ $N^2 * N_q$

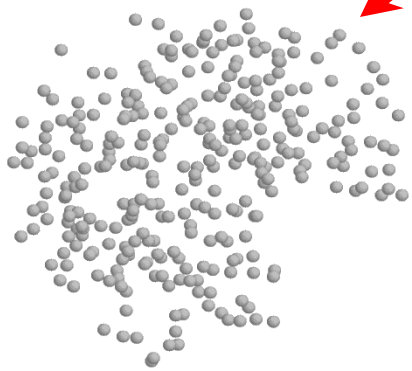



Grishaev, A., et al, JACS, (2005), 127, 16621





glob
(2-3 linked heavy atoms + Hs)

# Quasi-uniform vector grids for approximate angular averaging

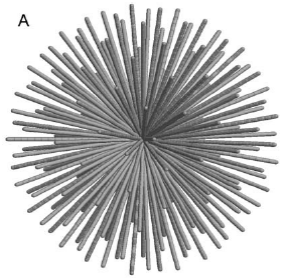
$$I(q) = \langle I(\mathbf{q}) \rangle_\Omega = \left| \left\langle \sum_{j=1}^{N} A_j \exp(i\mathbf{q}\cdot\mathbf{r}_j) \right\rangle_\Omega \right|^2$$

$O(N^2 * N_p) \rightarrow O(N * N_p)$

Fibonacci number-based algorithm

F(1)..... F(n)
F(i) = F(i-1)+F(i-2)
0, 1, 1, 2, 3, 5, 8, 13, 21, 34 ...
Ngrid = F(n) + 1

q(j) = arccos(1-2(j-1)/F(n))
f(j) = 2*p*mod(j*F(n-1)/F(n))

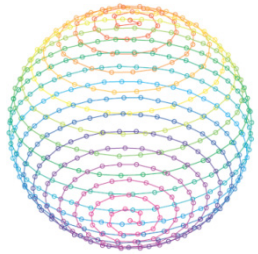


F(11) =145
...
F(18) = 4181

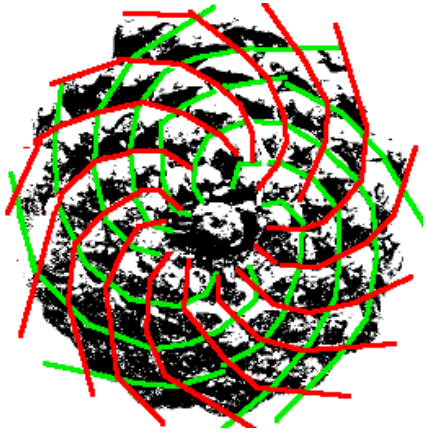
spiral algorithm

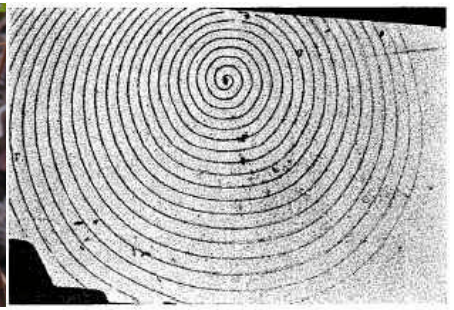
1.....N
h(j) = -1+2(j-1)/(n-1)

q(j) = arccos(h(j))
f(0) = f(N) = 0
f(j) =f(j-1)+3.6/sqrt(n*(1-$h^2$))

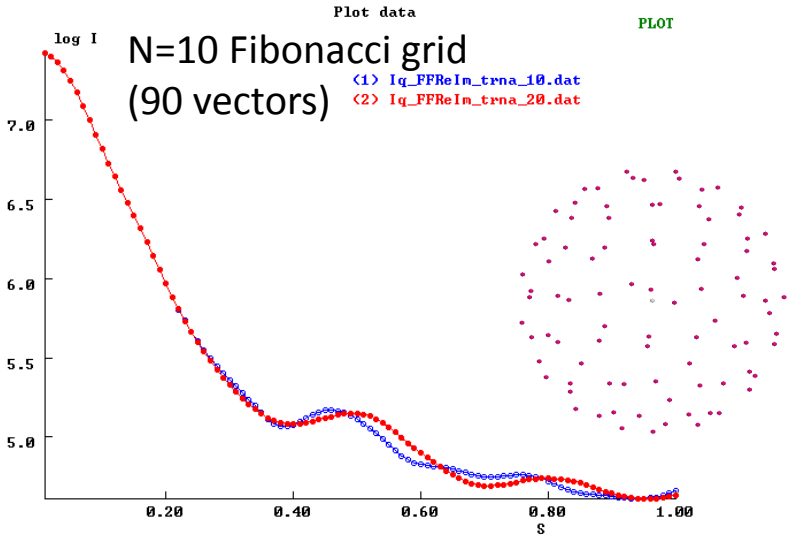


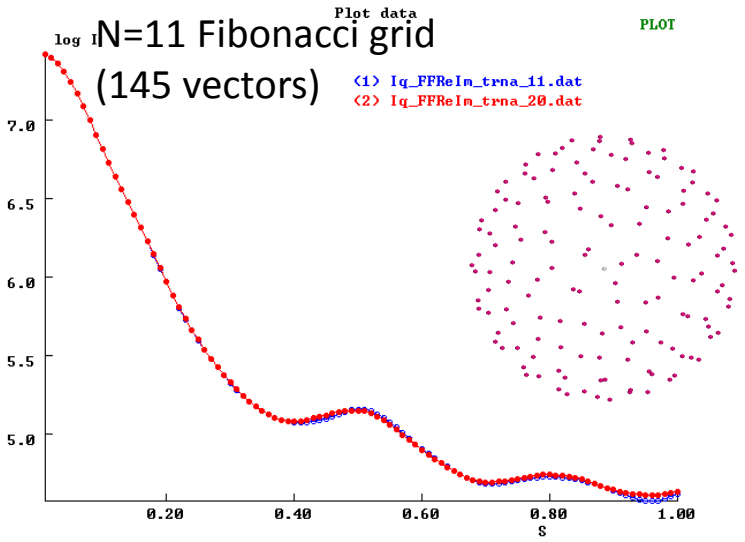
For a large grid (> ~$10^3$ vectors )
angular averaging is nearly exact

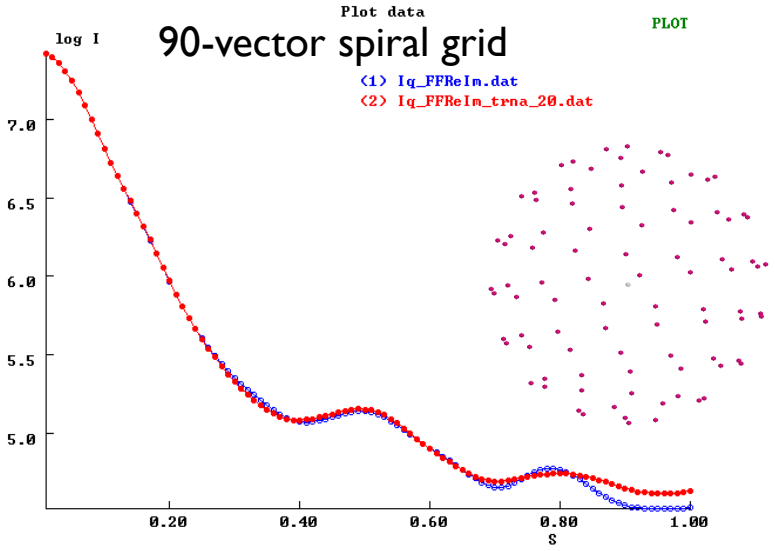




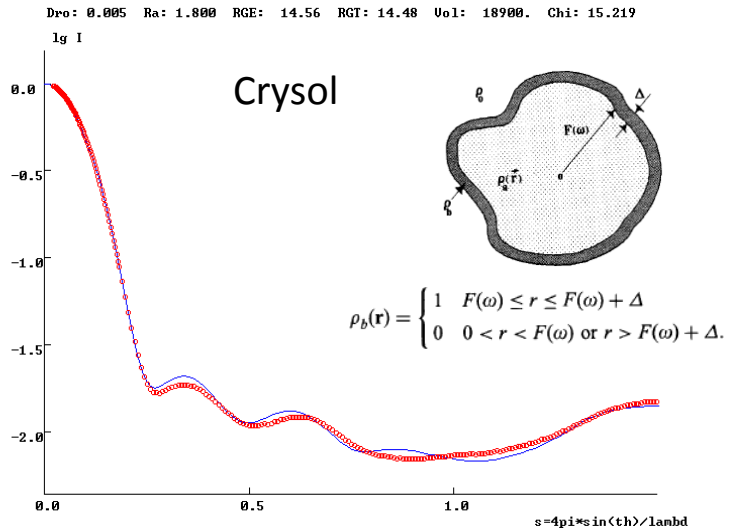
# Accuracy of data representation with small grid sizes



# CRYSOL




$$I(q) = \left| \left\langle \sum_j (f_j(q) - g_j(q)) \exp(i\mathbf{qr}_j) + \delta\rho A_l(\mathbf{q}) \right\rangle_\Omega \right|^2$$

Multipole expension:

$$\exp(i\mathbf{qr}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l j_l(qr) Y^*_{lm}(\omega) Y_{lm}(\Omega)$$

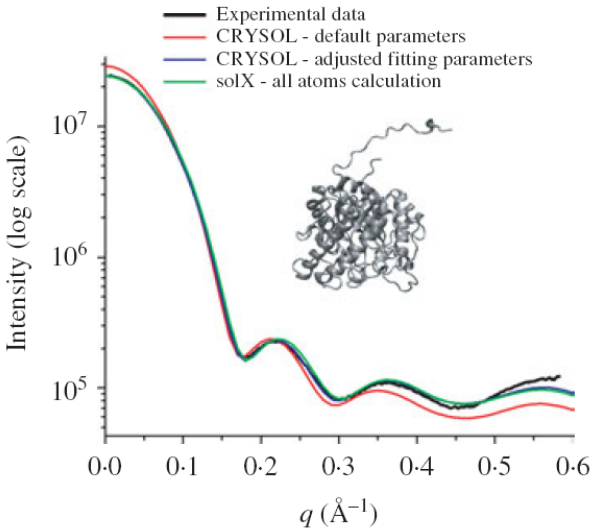
Approximation: finite $l$ harmonics

Complexity $\propto l^2 N$

Crysol predicts the I(q) data from the atomic coordinates or optimize the fit between predicted and experimental data by optimizing the following adjustable parameters:
*overall scale, average atomic displaced solvent multiplier, total excluded volume, and contrast of the bound solvent layer.*

Prediction accuracy at high angles depends on harmonics order and Fibonacci grid order. Highest level of calculations works best: 50 (1-50) for the harmonics order and 18 (10-18) for Fibonacci grid order.

Predictions for elongate or highly asymmetric molecules are less accurate.




Svergun, D. et al. (1995) J. Appl. Cryst. 28, 768-773.

Putnam, D., et al. (2007) Quart. Rev. Biophys. 40, 191-285.

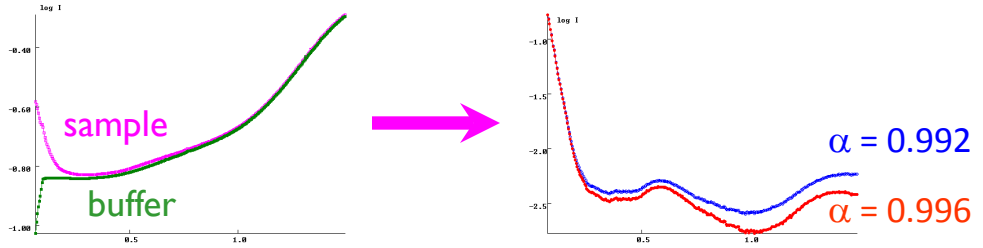
# AXES (Analysis of X-ray Scattering Explicit Solvent):

## Fitting SAXS data to atomic models using explicit solvent representation and variable sample/buffer rescaling

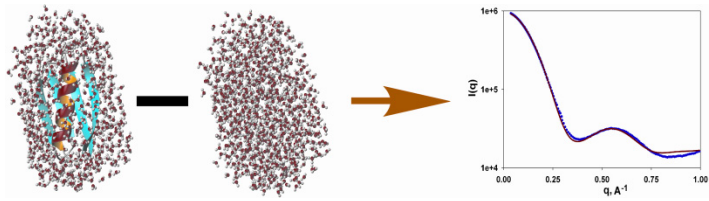
server: http://spin.niddk.nih.gov/bax/nmrserver

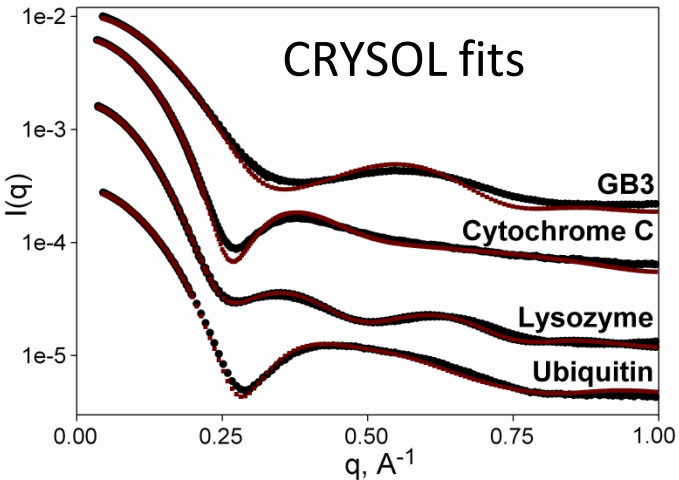
- Accurate description of arbitrarily complex macromolecular surfaces
- Macromolecular ensemble fits with the same ease as those for single models
- Explicit and accurate (TIP5P) representation of displaced and surface solvent
- Optimization of the sample/buffer rescaling constant and offset to account for variability of the experimental data (as opposed to Crysol's variation of the excluded solvent density)

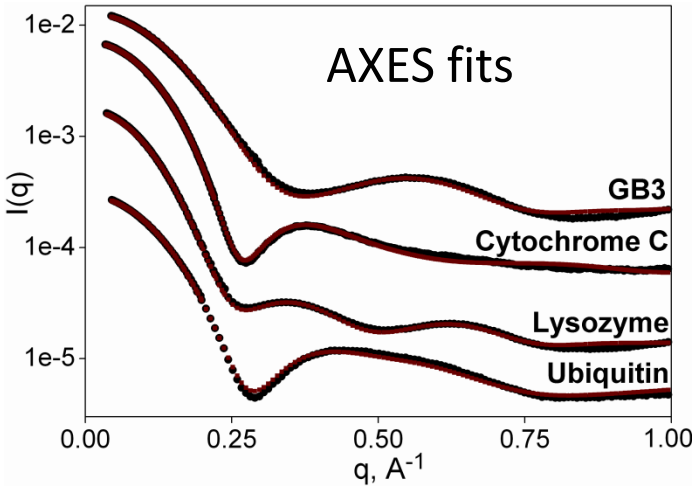
$$I_{expt}(q) = I_{sample}(q) - \alpha I_{buffer}(q) + const$$




$$I_{calc}(q) = I_{mol-mol}(q) + I_{displ-displ}(q) + \rho_{surf}{}^{2} I_{surf-surf}(q)$$
$$- 2I_{mol-displ}(q) + 2\rho_{sur} I_{mol-surf}(q) - 2\rho_{surf} I_{displ-surf}(q)$$






Grishaev et al. (2010) JACS, *submitted.*

# 2. Experimental Aspects of Scattering

**Instruments**

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

**Data acquisition**

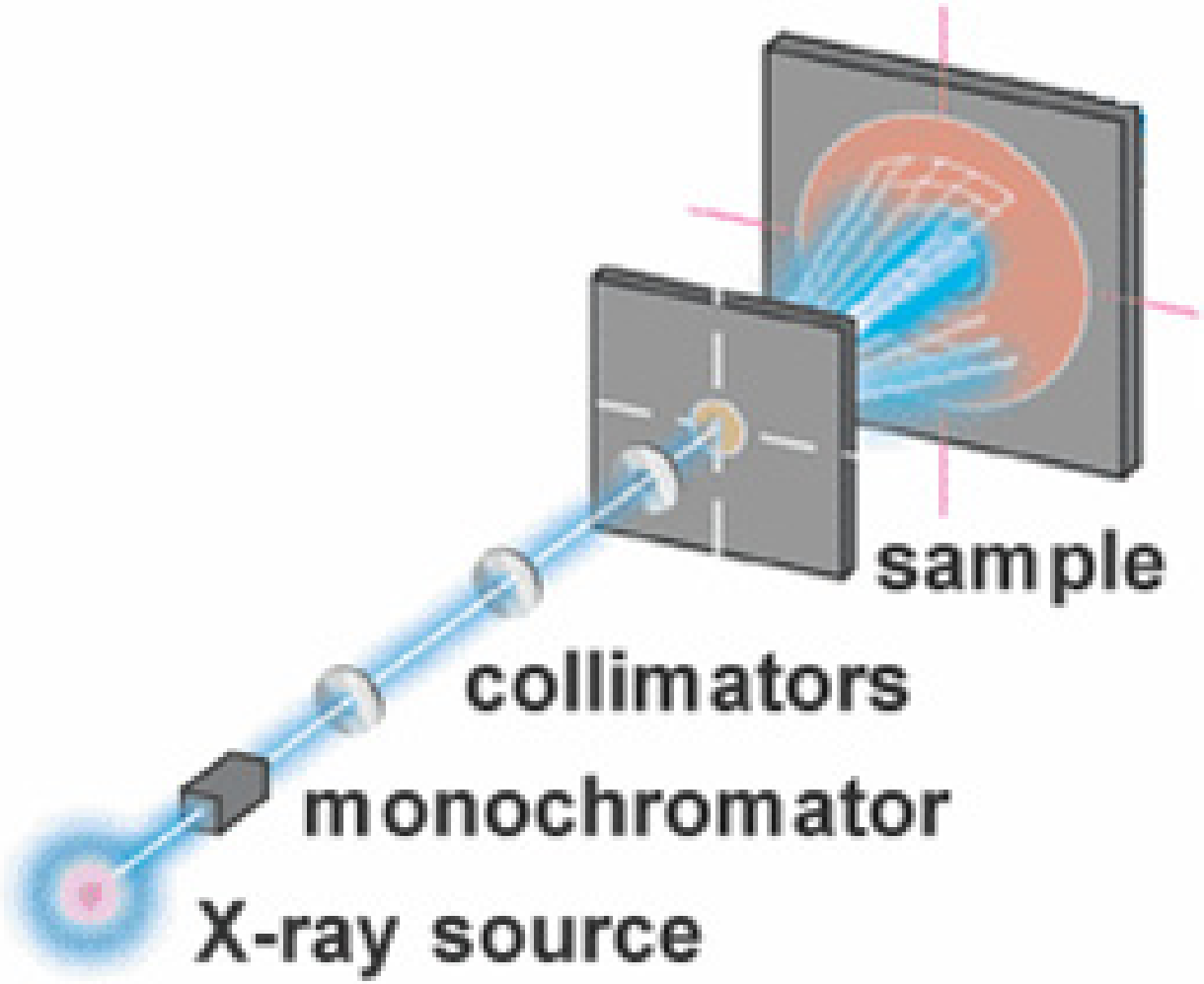
- Synchrotron based SAXS experiments
- Bench-top X-ray source based SAXS experiments

**Scattering sample preparations**

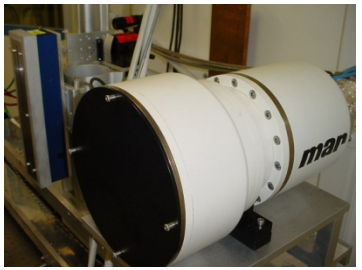
# Instruments

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

# Basic design of the SAXS instrument




Multi-wire gas-filled detector



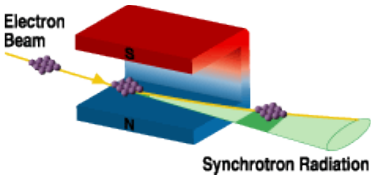
CCD area detector

Lab-based sources:

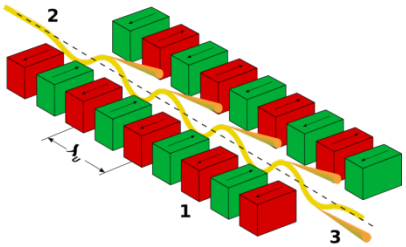
Rotating anode

Sealed tube

Synchrotron sources:

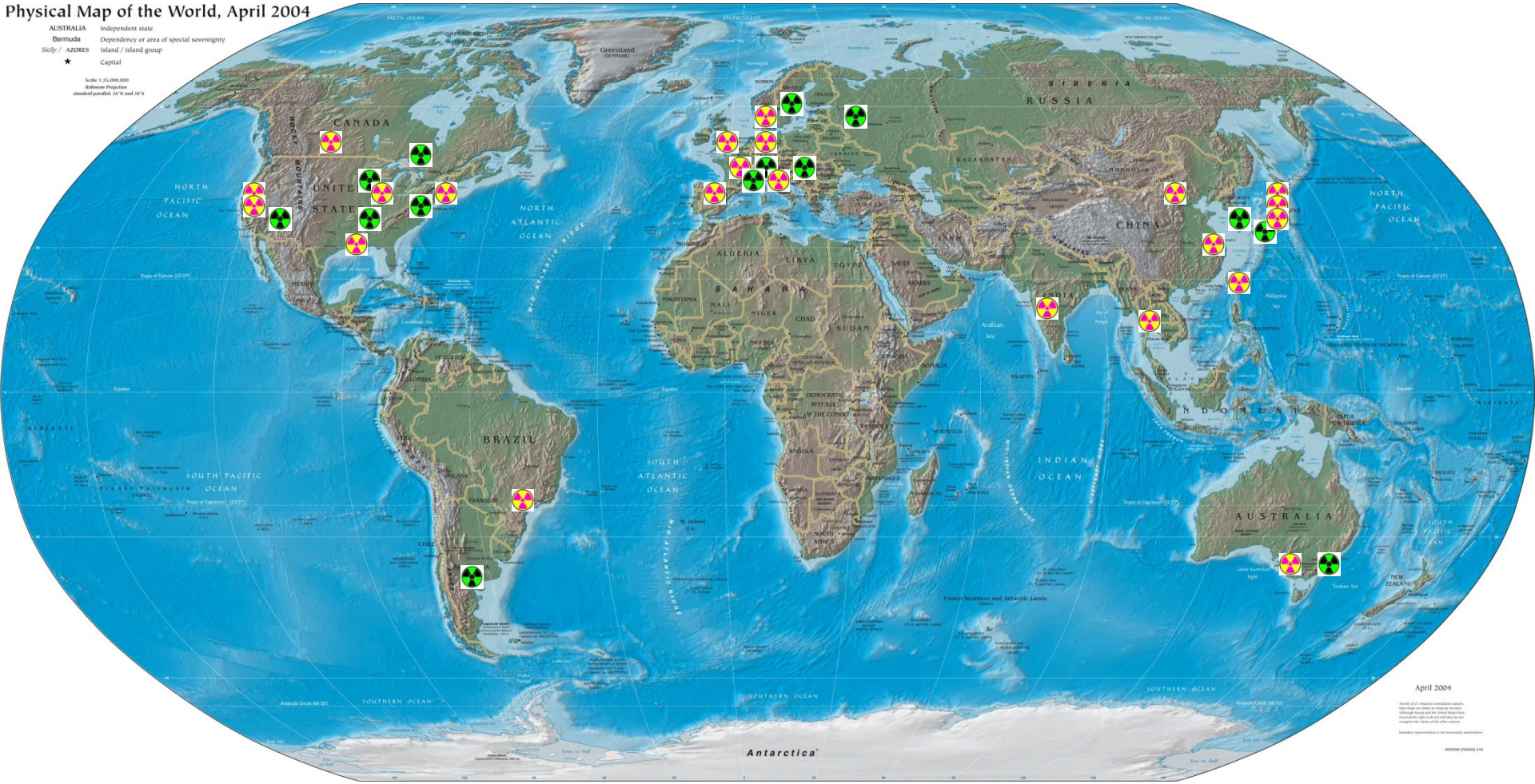



Bending magnet




Undulator/wiggler

# SAXS & SANS beamlines worldwide




SAXS beamlines

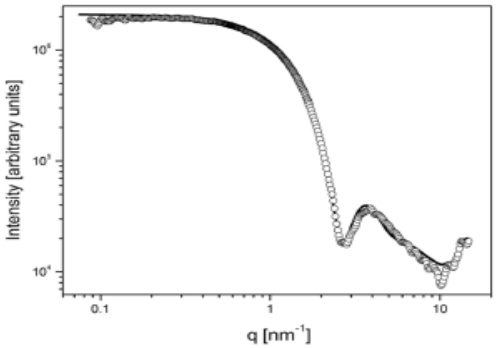
SANS beamlines

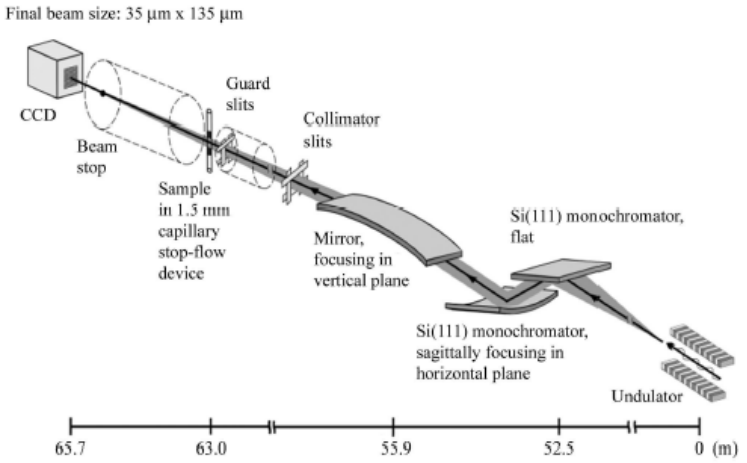
# Examples of synchrotron SAXS beamlines: APS (Argonne, IL, USA)

**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

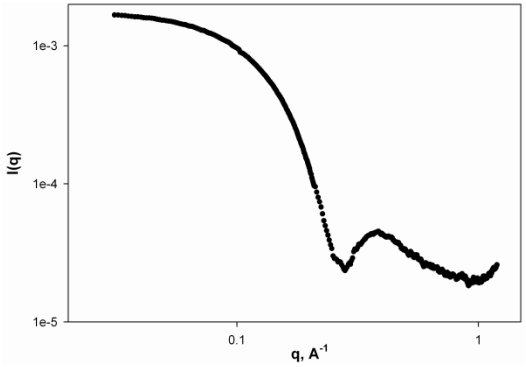


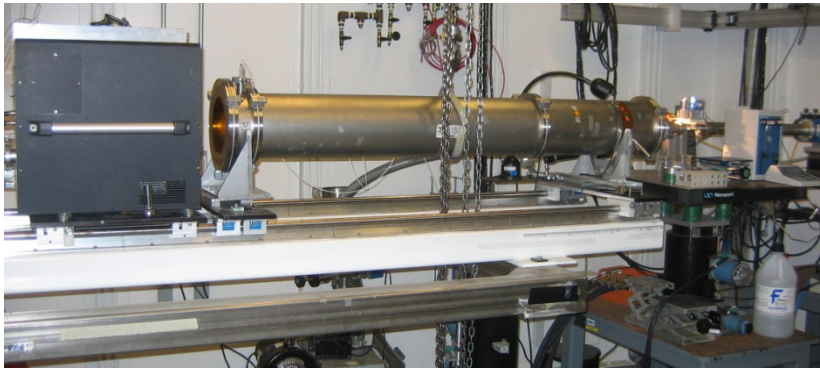




sample data above: 2.5 mg/mL cytochrome c, 1s data collection at 2m and 0.3 m

Fiscetti, F et al. (2004) 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)

# Data Acquisition

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

# Synchrotron data collection overview

Measurements:

empty capillary,
capillary with buffer,
capillary with sample,
the corresponding incident and transmitted intensities.

Measure in 2 configurations (SAXS and WAXS whenever possible), then merge.

Use SAXS to evaluate inter-particle interference, aggregation & radiation damage
and WAXS to evaluate sample/buffer matching.

Watch for radiation damage, attenuate beam and decrease exposure when suspected.

Adjust exposure times to avoid detector saturation – limited by parasitic scattering near the beam stop for SAXS and by the water peak intensity for WAXS.

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

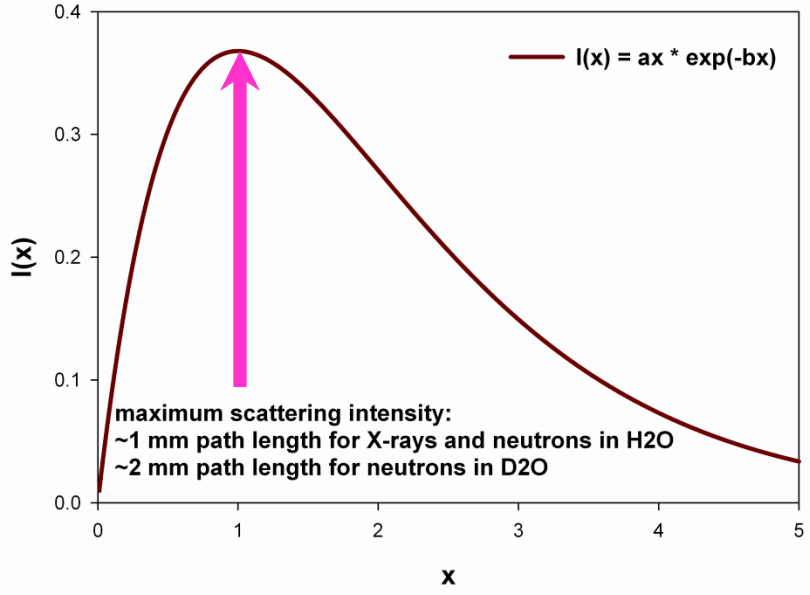
Sample volumes are 15-30 μL for a static cell and 50-150 μL for a flow-cell.

Measure buffer before and after each sample measurement to detect protein deposition in the cell.

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.

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) are necessary.

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

If line source is used, desmearing with be necessary.

Minimum concentration is ~ 1 mg/mL protein.

# 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. 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.

To decrease bubble formation during data collection, dialyzate buffer should be degassed.

The samples should be passed through a 0.22 μm filter before dialysis.

Sample concentration should be measured after the dialysis (UV-Vis).

# Composition of the buffer

**Salts** are useful to suppress long-range electrostatic interactions between solutes (structure factor).

They also increase background and decrease solute/solvent contrast but these effects are 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 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.

**Detergents** are best avoided unless absolutely necessary (membrane proteins) due to their often strong signal.

# 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.

**Solute concentration**: 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 via I(0)/c or $R_{gyr}$ analysis. In those cases, zero-concentration extrapolation should be performed.

For proteins in 150 mM salt buffer, structure factor is often quite small up to 5-8 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.

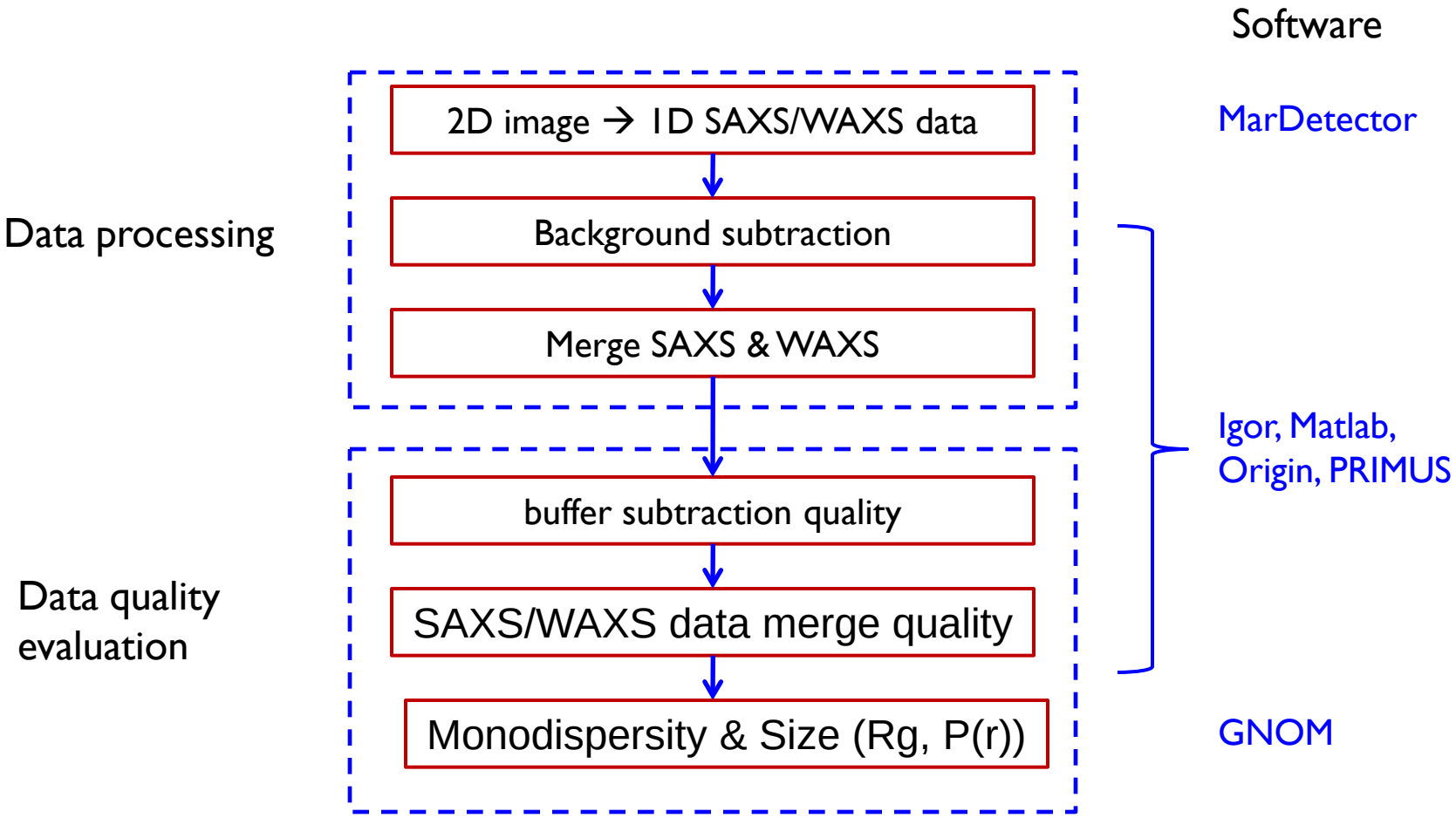
# 3. Data processing

2D → 1D data conversion

Background subtraction

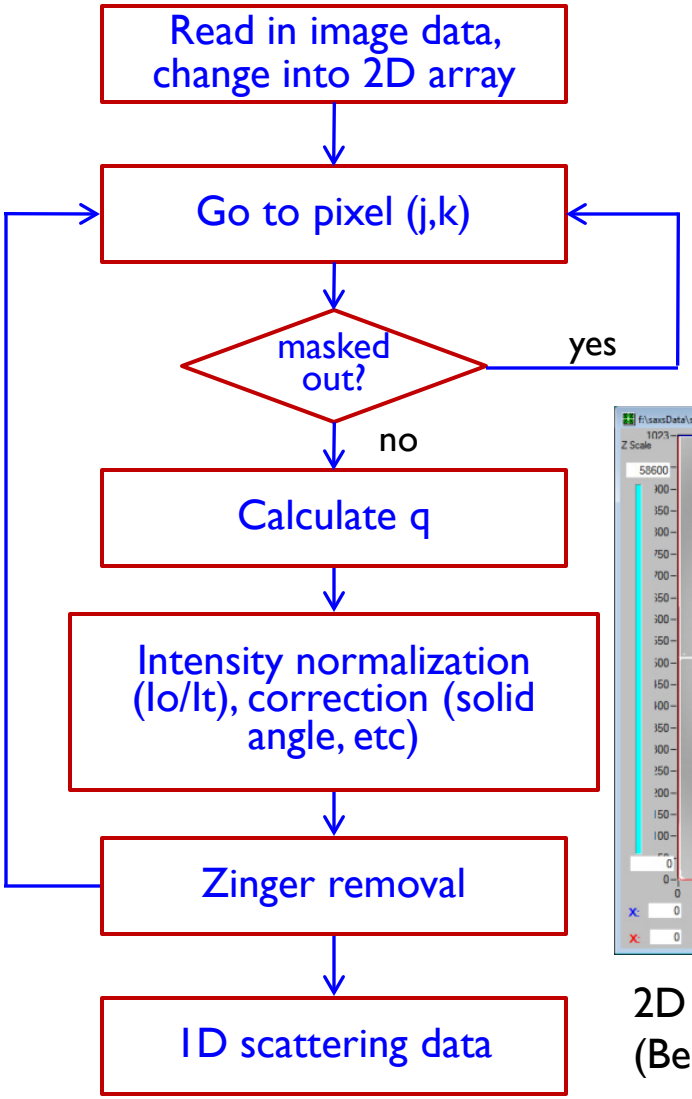
Data quality evaluation

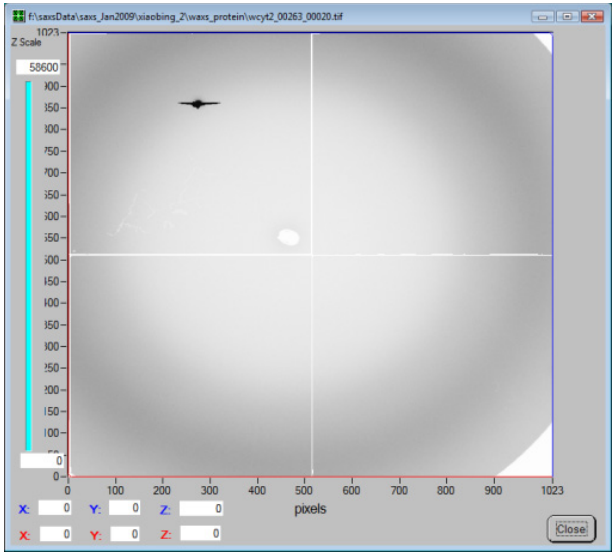
# Data processing and analysis flow chart



# 2D image → 1D data conversion

- beam center positioning
- q calibration/mapping
- Image masking
- Sensitivity, dark current,
- & solid angle corrections
- Averaging



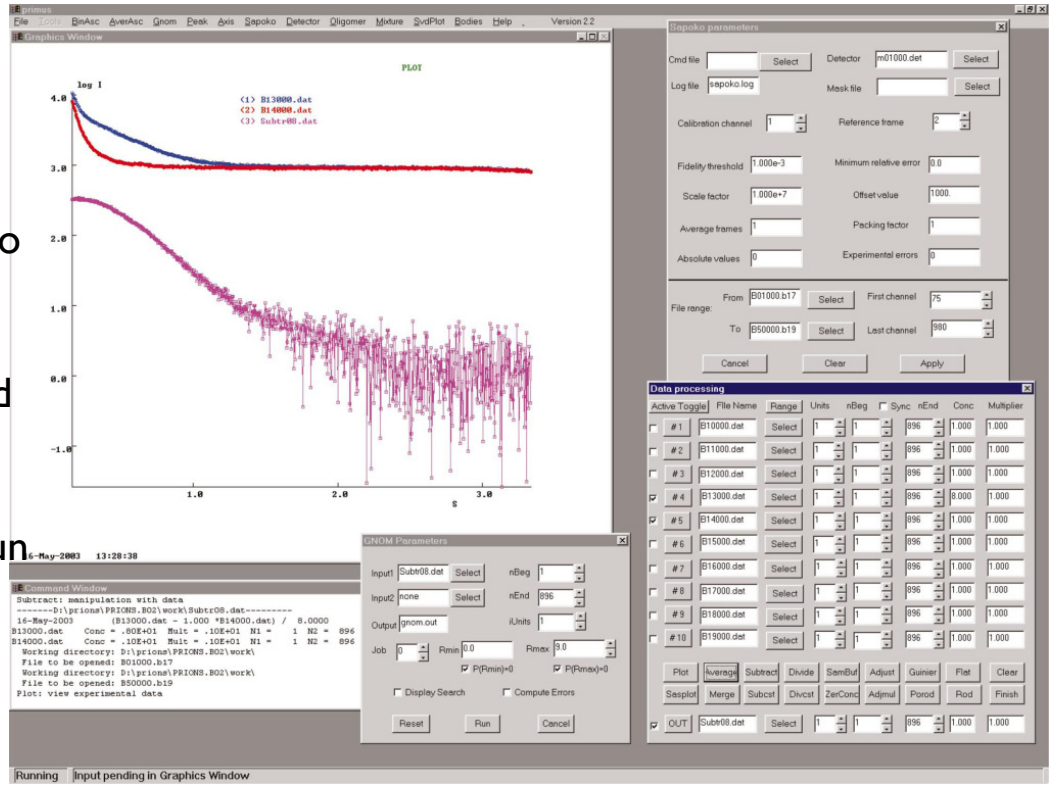




2D solution scattering image
(Beam line 12-ID, Argonne)

# Software for data manipulation

**Primus:** developed in Svergun group

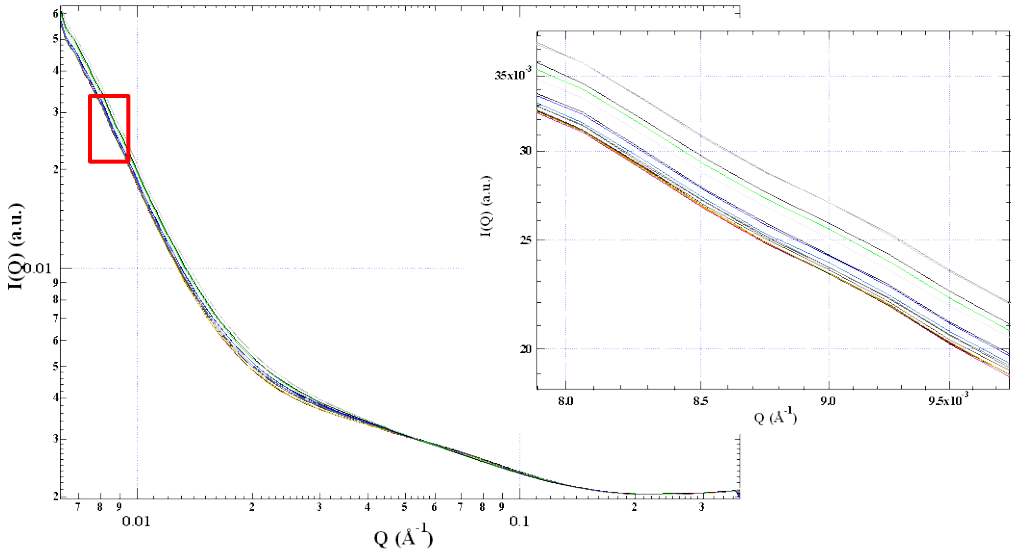
- Basic data manipulation functions (averaging, background subtraction, data merging at different angular ranges, extrapolation to zero sample concentration, etc.)
- Computes invariants from Guinier and Porod plots.
- Links to other programs developed in Svergun group: GNOM, Dammin etc.

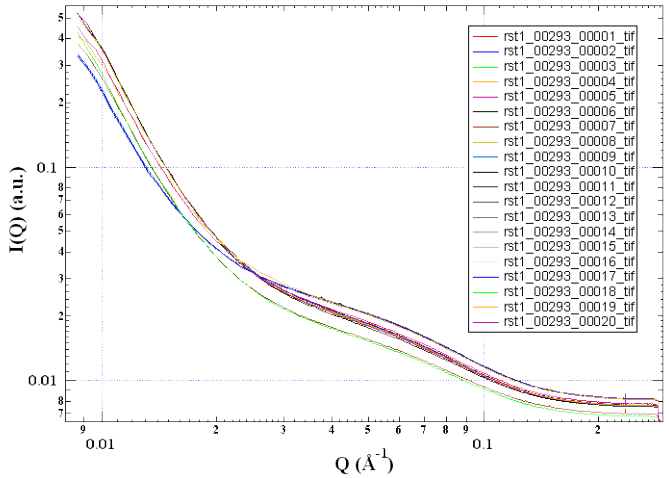



**Igor / Origin / Matlab:**

- Basic data manipulations: averaging, background subtraction, data merging,
- Guinier plot, Kratky plot, data extrapolation, data point sparsing
- Saves parameters used in data processing

# Data divergence check



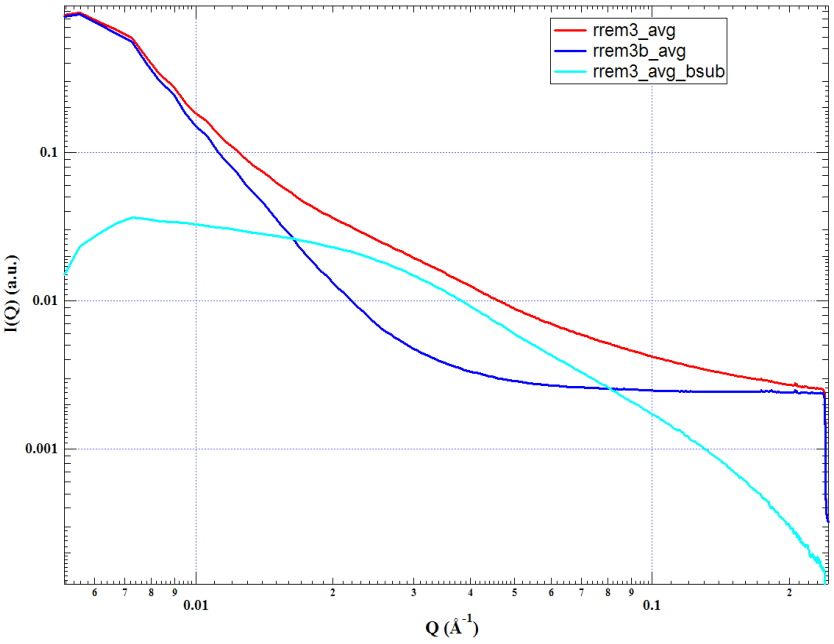




- 10-30 scattering image frames are often collected to get good statistics.
- Outliers could occur due to air bubbles, small particles, beam instability, or radiation damage. They tend to appear more often in sample rather than buffer data.
- Zoom in, find outliers and remove them.
- To overcome bubbles during data collection:
  - clean the flow cell extensively
  - use slower pump speeds.
  - use stopped flow with attenuation if the sample can tolerate radiation damage.
- With radiation damage, scattering intensities change in time.
- To decrease radiation damage during data collection:
  - Attenuate beam/decrease exposure
  - Change buffer composition
  - Decrease sample concentration

# SAXS background (buffer) subtraction

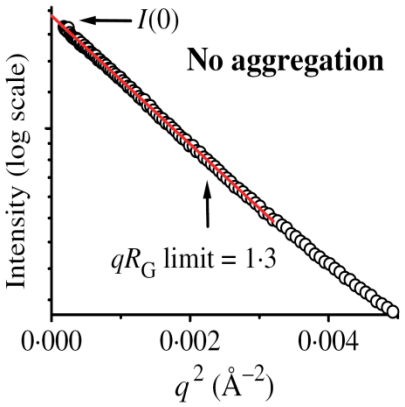
- SAXS background subtraction:
  - A. $I_{sample} = I_{solution} - I_{buffer}$
  - B. $I_{sample} = I_{solution} - \alpha * I_{buffer}$
    - protein: $\alpha = 1 - C_{mg/ml} * 7.43\ 10^{-4}$

      Nucleic Acids: $\alpha = 1 - C_{mg/ml} * 5.4\ 10^{-4}$
  - C. using WAXS as guide

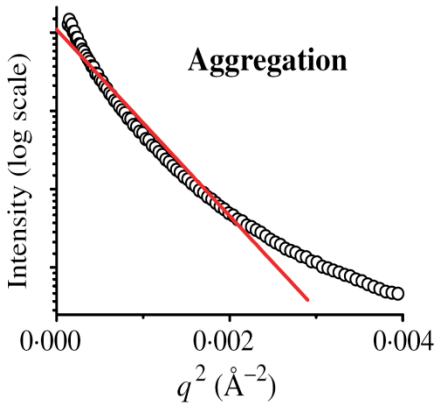



At the beamline, methods A & B should be used
to quickly get scattering data for real-time evaluation.

# Guinier plot for SAXS data evaluation

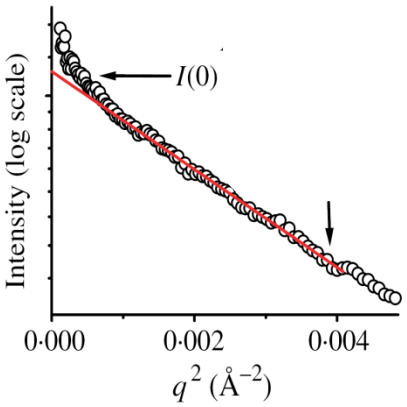



Monodisperse sample



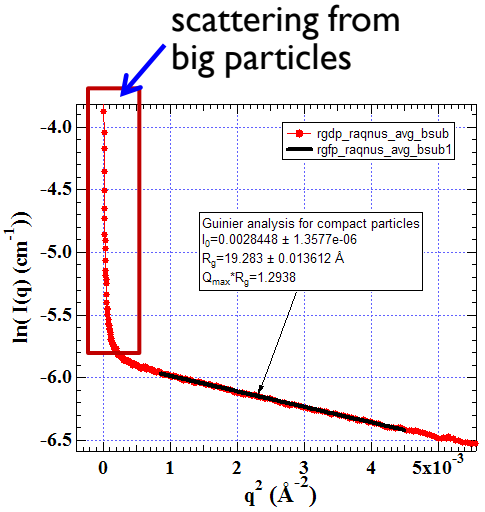

Polydisperse (continuous size distribution) sample

Data unusable unless intrinsically unfolded




Small amount of aggregation

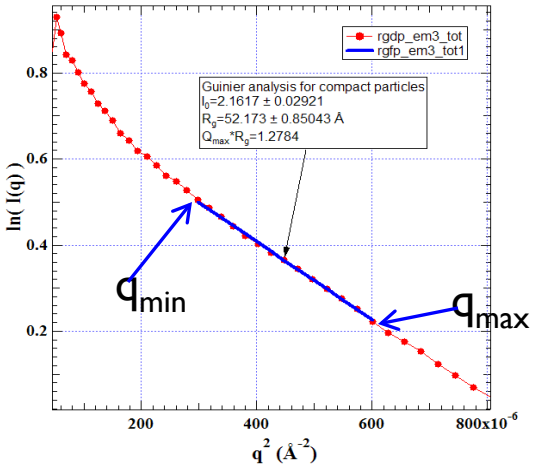
Data fixable



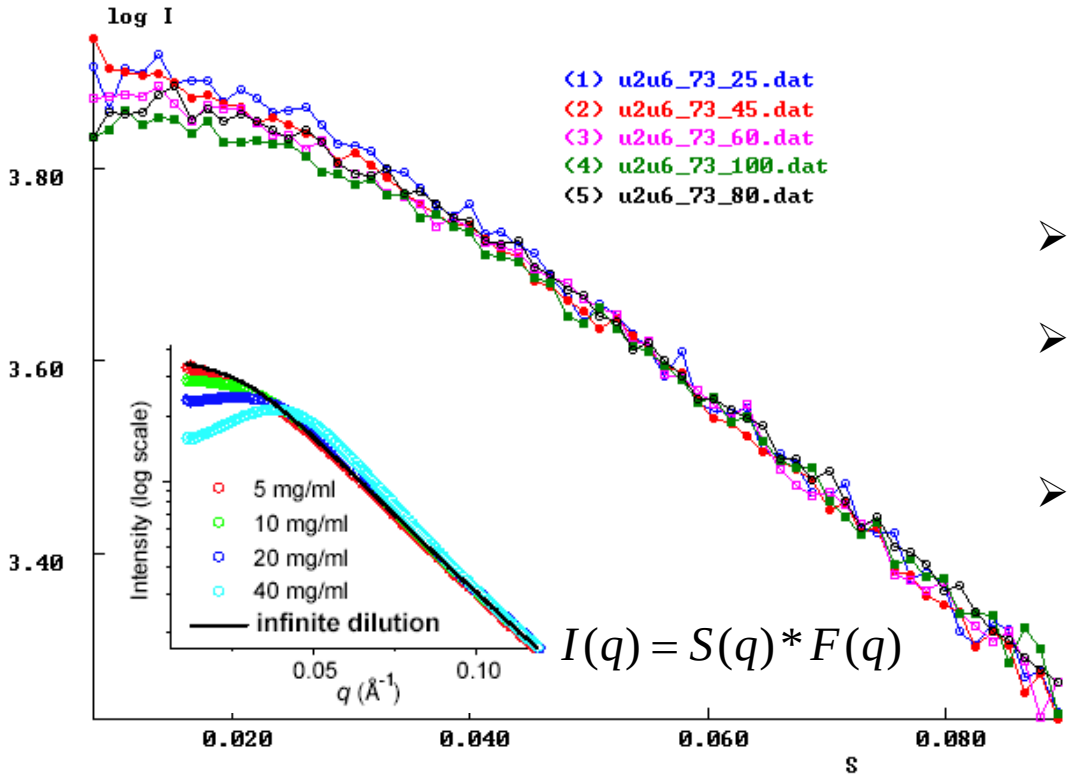

Sample with large aggregates: huge slope at lowest q, drops fast
To remove: 0.22μm filter

To get reliable Guinier plot / $R_g$ analysis:

- $q_{min} \leq \pi/D_{max}$
- $q_{max}*R_g<1.3$ for globular; <0.8 for elongated
- Multiple (≥5) data points in the Guinier region



# Structure factor (inter-particle correlations)



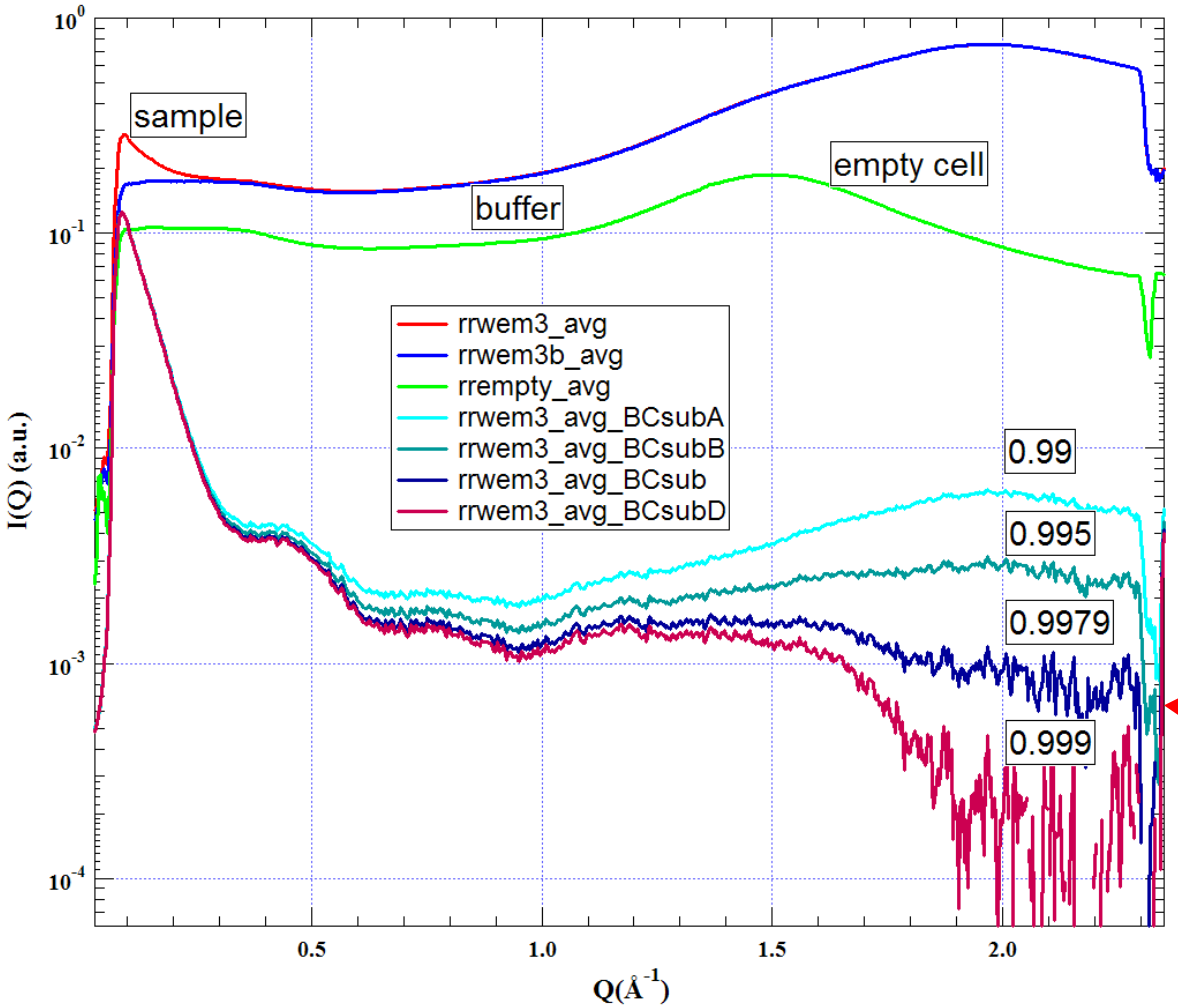

- Disappears at low concentrations
- Concentration series need to be recorded
- Concentration dependence of Rg, Io/c needs to be evaluated

# WAXS sample/buffer subtraction

Solution sample scattering originates from biomolecules, buffer, and cell.

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

Residual empty cell scattering becomes significant only for dilute samples



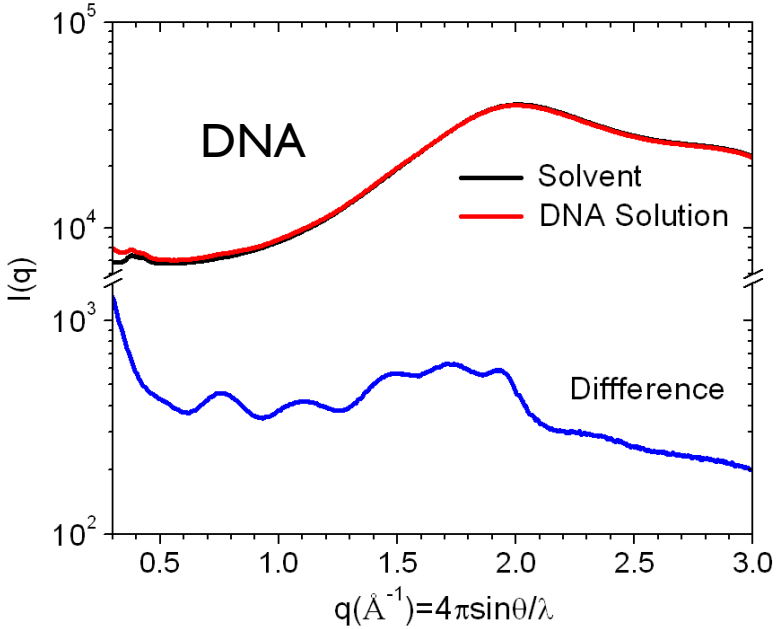

Watch the solvent peak (at ~2.0 Å$^{-1}$) disappear

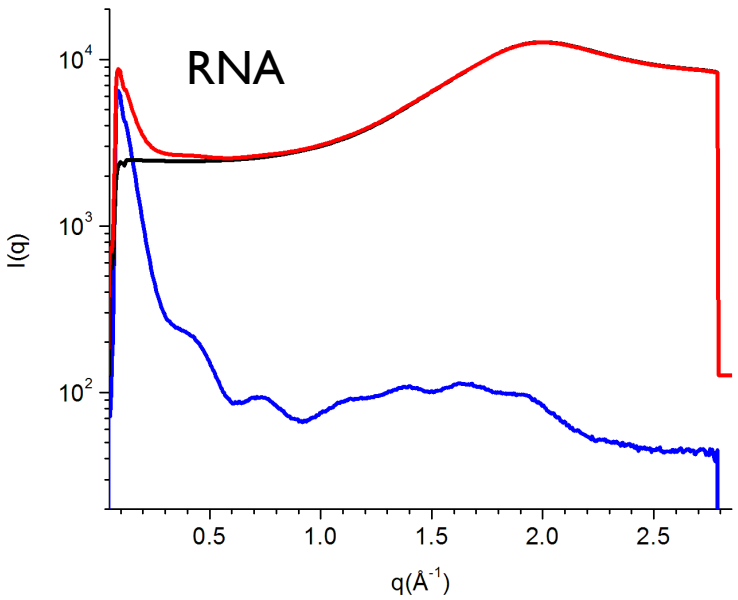
under-subtraction
(significant amount of solvent)

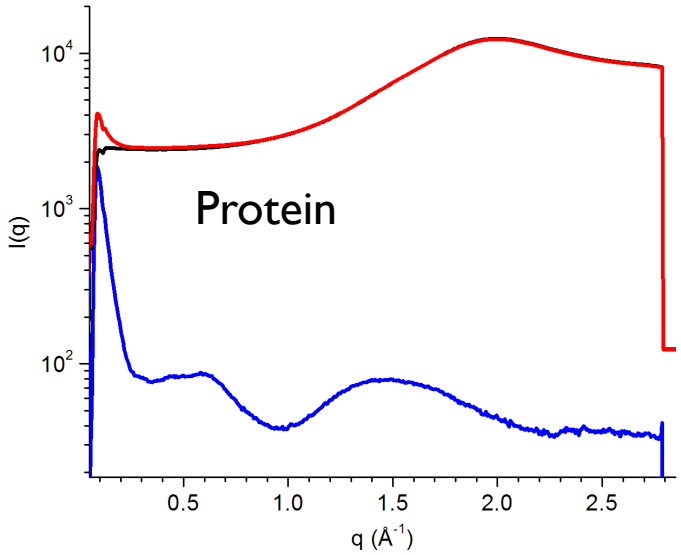
reasonable subtraction

over-subtraction

# Examples of good WAXS background subtraction




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

# WAXS data-guided SAXS background subtraction

**Why use WAXS?**

- results of A/B SAXS subtraction do not match WAXS
- WAXS background subtraction is often more accurate

**How?**

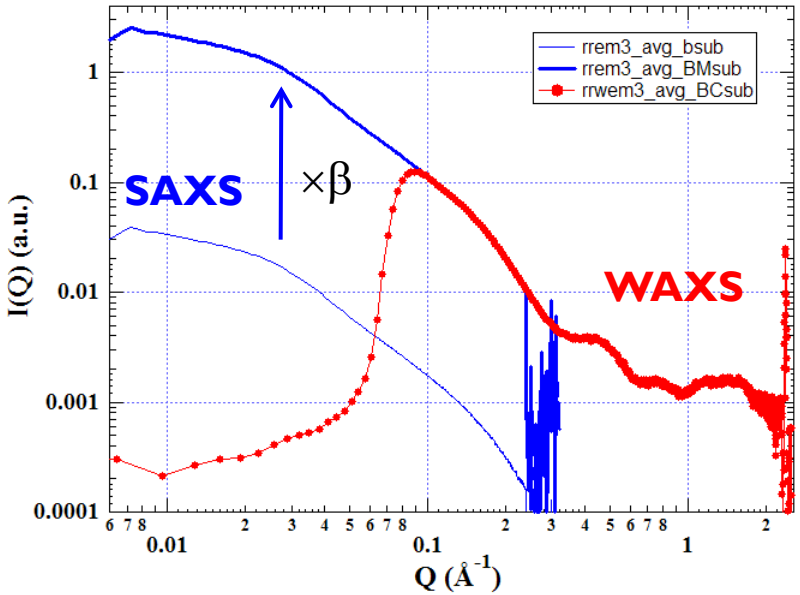
(1). SAXS: $I_{sample} = I_{solution} - \alpha * I_{buffer}$-const

$\alpha$: account for contribution from buffer

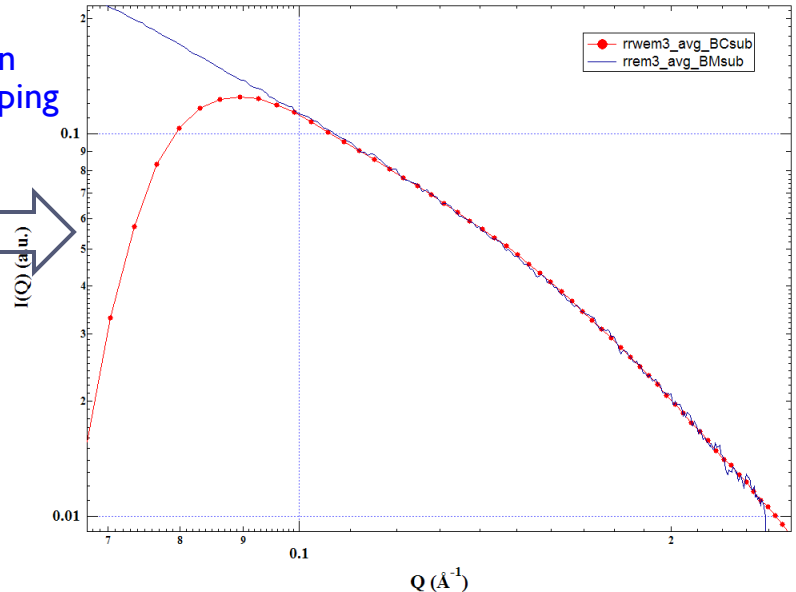
const: residual electronic noise

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

(3). Tune $\alpha$ (**const** if necessary) and $\beta$ to make SAXS and WAXS to have best match.



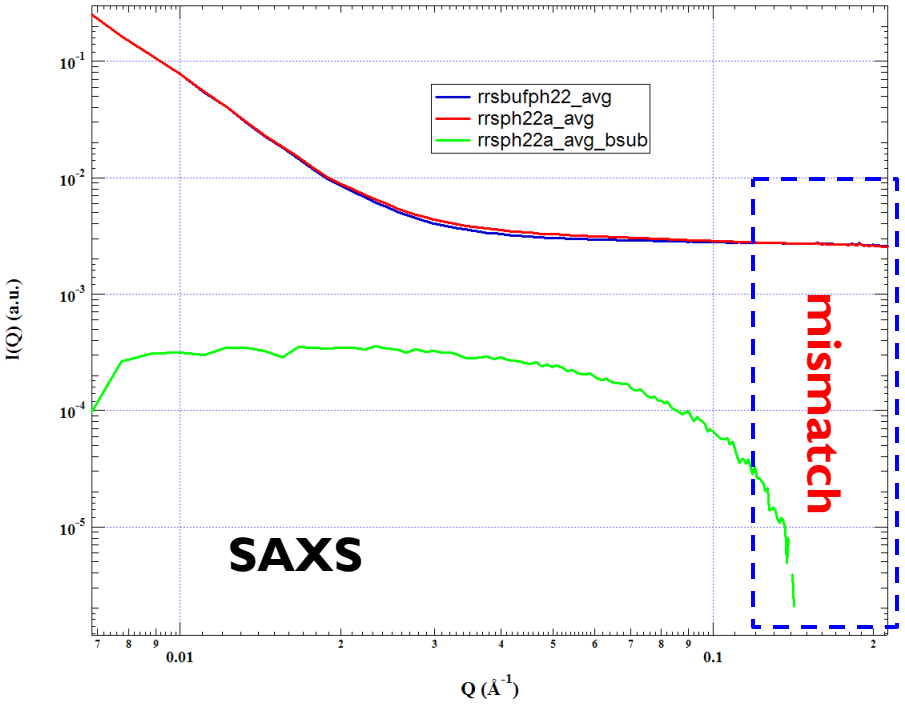

Zoom in overlapping region

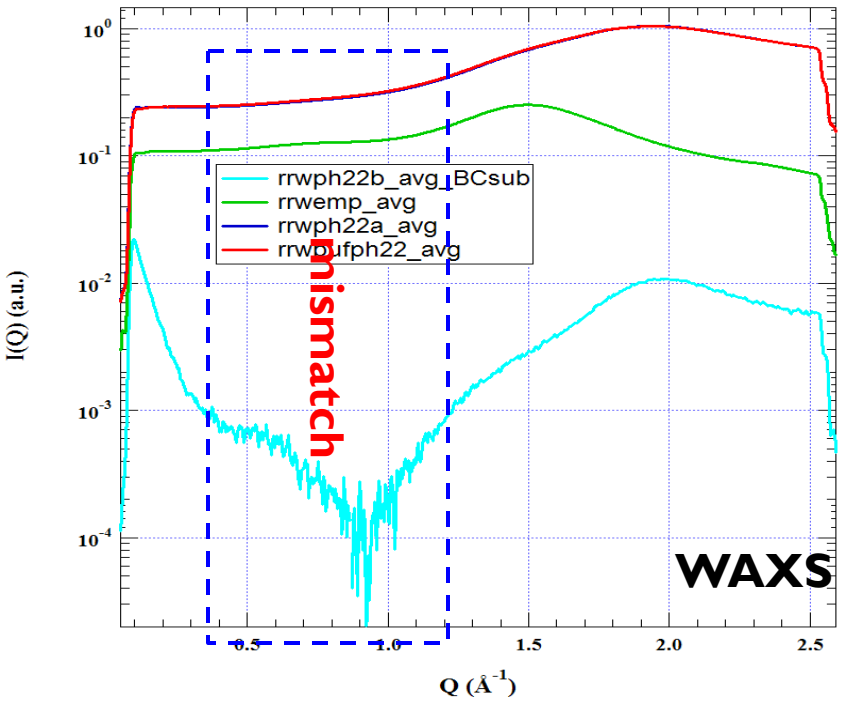


# Data quality evaluation: background mis-match

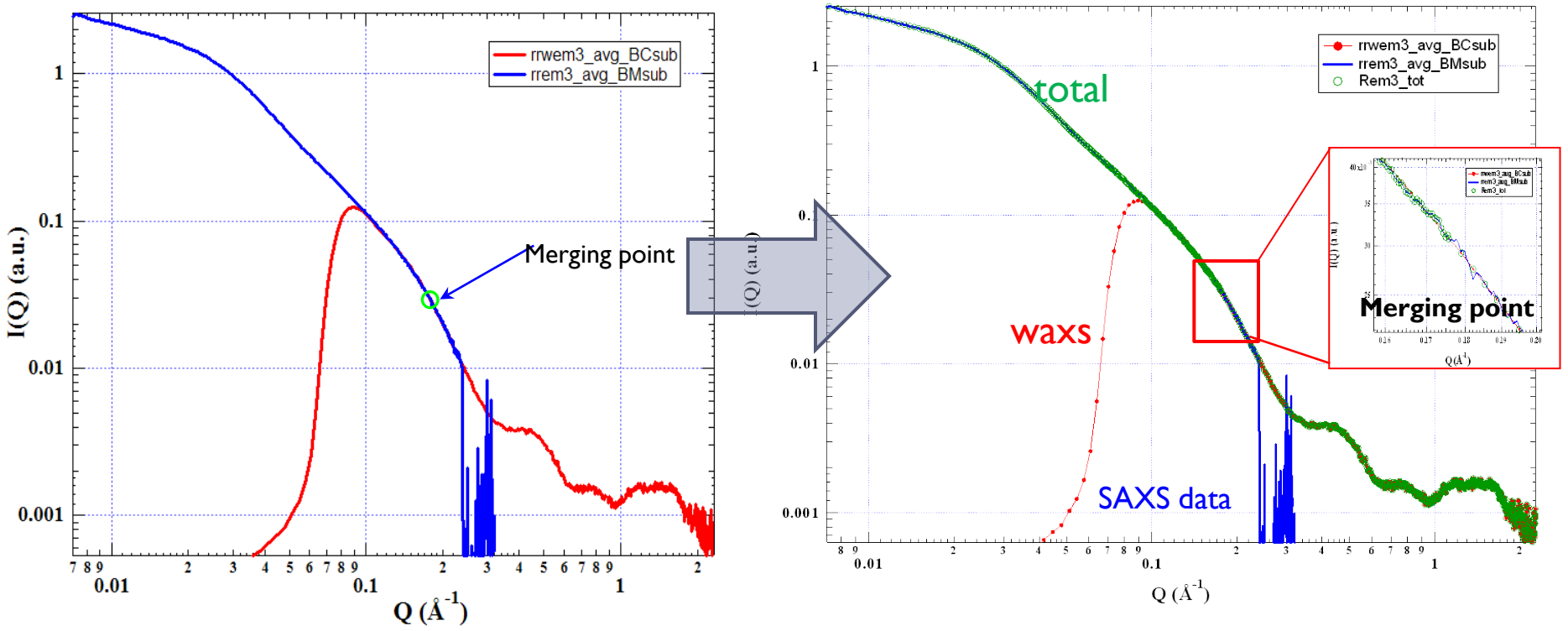
Solution X-ray scattering measurements are very sensitive to the background match.

Here is an example of a dilute sample with slight background mis-match



# SAXS / WAXS data merging



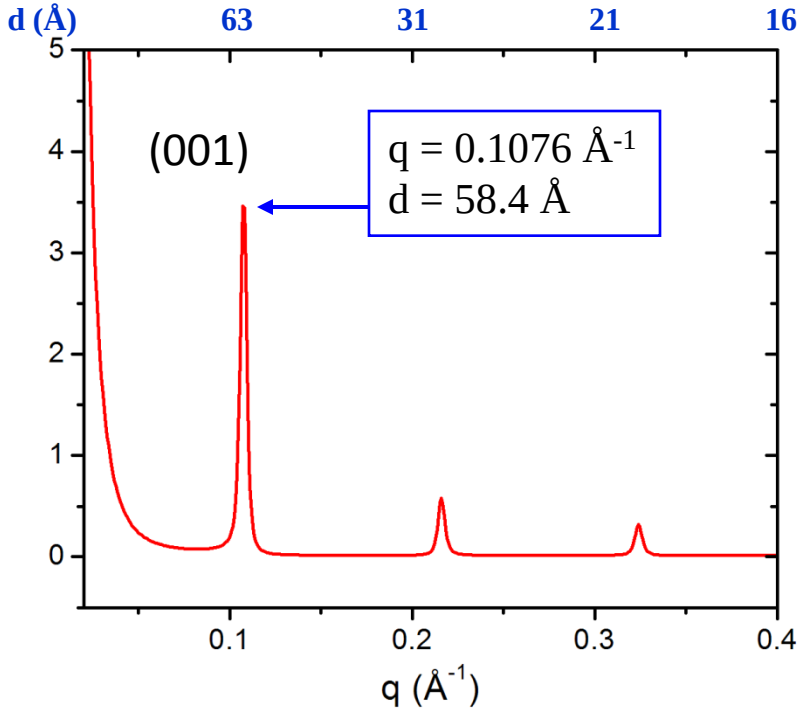

Choose $q_{min}$ and $q_{max}$ merge points that are reliably measured in both SAXS and WAXS data to form the combined data set with complete q range.

# 1. X-ray scattering profile and embedded structural information

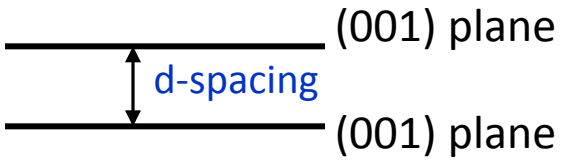
- d-spatial resolution
- hierarchical structural information
- SAXS vs. WAXS
- Guinier plot and Radius of gyration
- Molecular weight deterimination
- Porod's law, Porod invariant and Porod volume estimation
- Kratky plot
- Pair distance distribution function (PDDF) and PDDF calculation

# d-spacing/Characteristic Length/Spatial Resolution

d-spacing / characteristic length: $d = 2\pi / q$



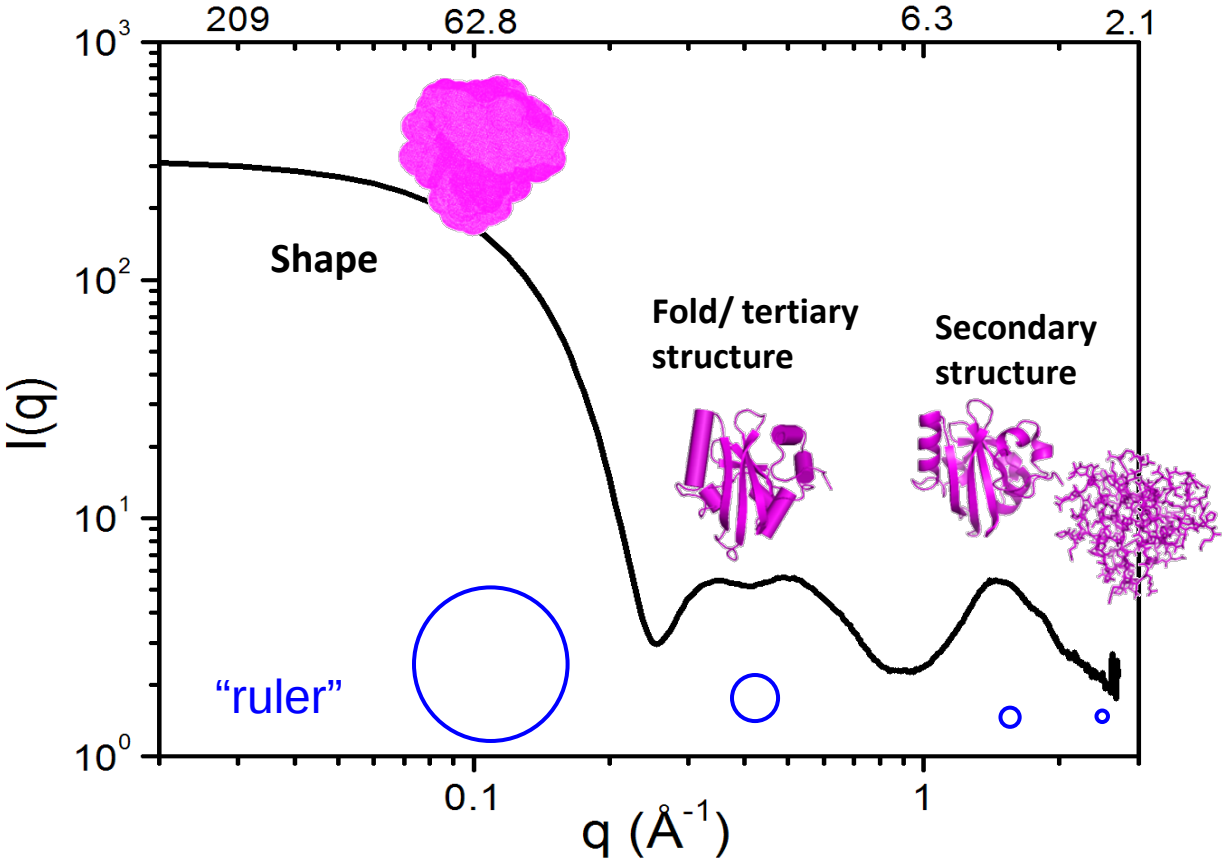

powder scattering / diffraction of silver behenate




- In powder diffraction pattern, peak positions represent certain characteristic distance (d-spacing, $d = 2\pi/q$) in the sample.
- A certain q in reciprocal space corresponds to a certain spatial resolution that could be possibly achieved in real space.

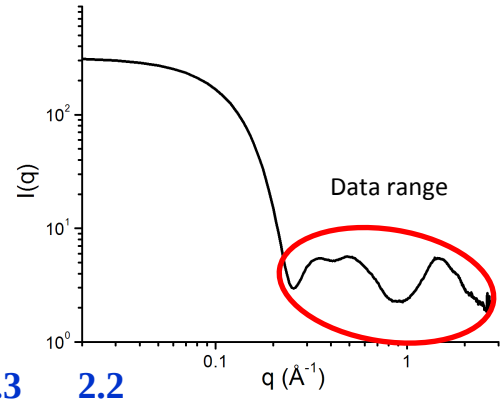
# Hierarchical structural information

characteristic length / spatial resolution: $d = 2\pi/q$

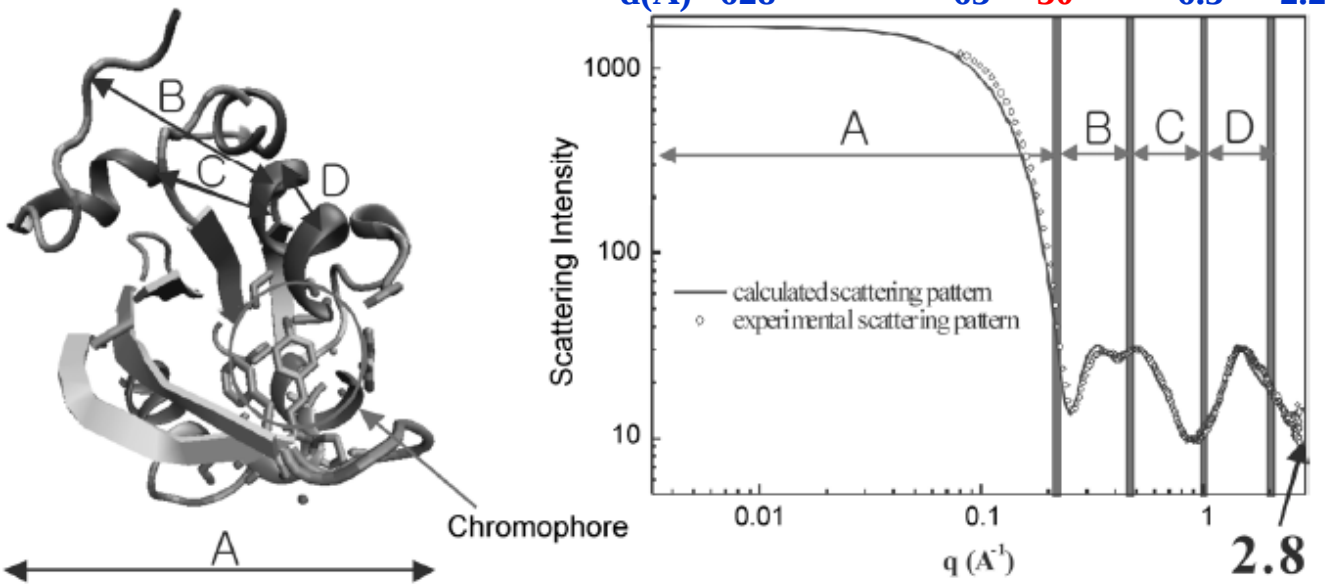



- In various q-regions, one views the molecule at different scale/resolution, see different levels of resolution on the structural details.
- SAXS(0- 0.2-0.4 Å$^{-1}$): shape, size, inter-particle interactions, etc
- WAXS (> 0.2 Å$^{-1}$): internal structural details

# WAXS feature assignment/structural mapping




$$d = 2\pi/q$$



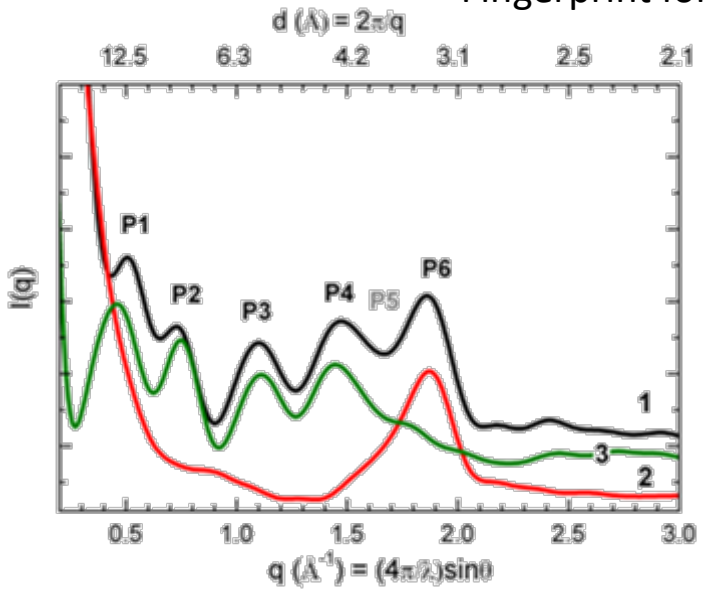

**Photoactive Yellow Protein (PYP)**

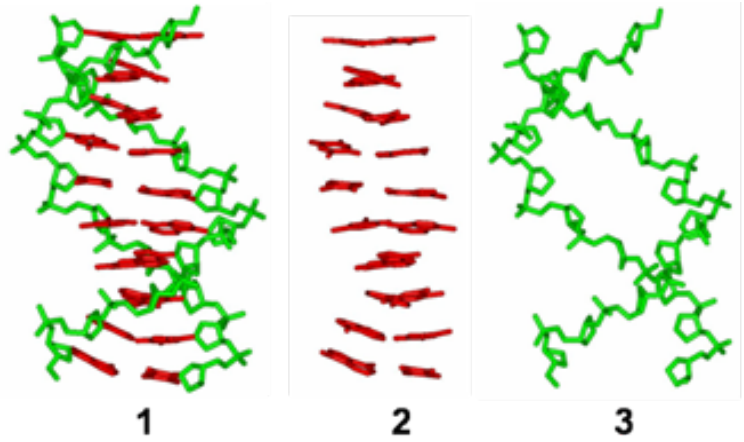
- SAXS provides information of size, shape, interparticle interaction, etc;
- WAXS fingerprints higher resolution structural characters.
- If coordinates available, further assignment/structural mapping is possible.
- For highly regular molecules, it is possible to extract structural information from waxs

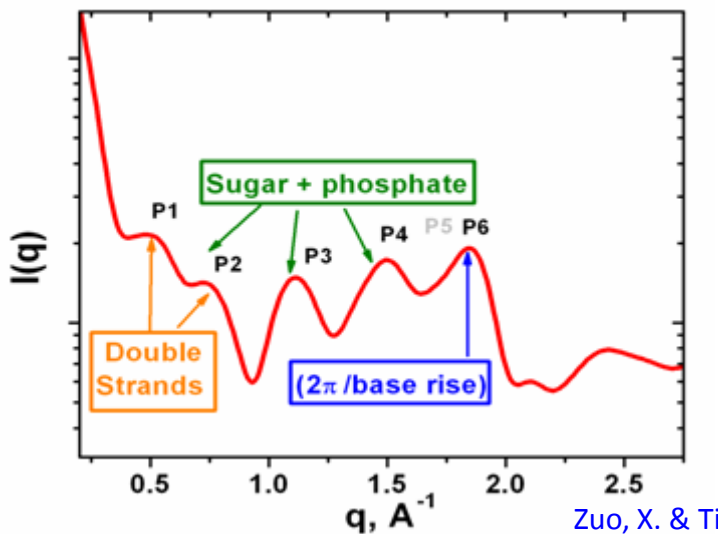
Kim et al; Bull. Korean Chem. Soc. 2004, 25:1677

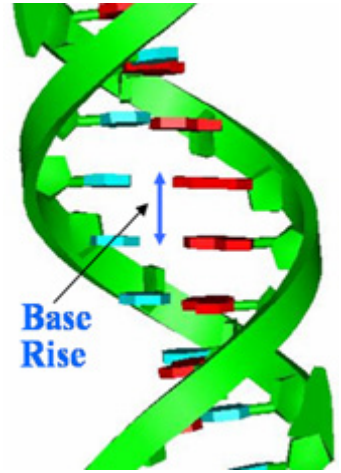
# WAXS “fingerprints” for DNA conformation and parameters

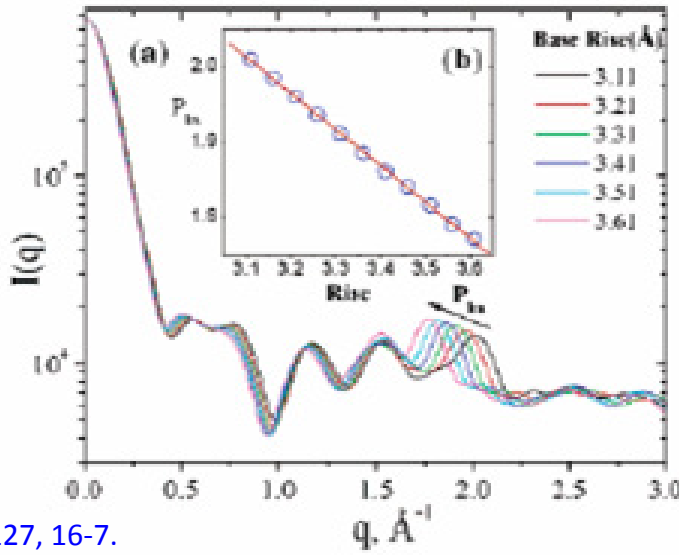
Fingerprint for canonical B-form Duplex DNA:




Zuo, X. & Tiede, D. M. (2005). J Am Chem Soc 127, 16-7.

# Guinier equation

Scattering intensity can be expanded in powers of $q^2$:

$$I(q) = I(0)\left[1 - \frac{R_g^2 q^2}{3} + kq^4 + \cdots\right]$$

When $q \to 0$,

$$I(q) \cong I(0)\exp\left(\frac{-R_g^2 q^2}{3}\right)$$

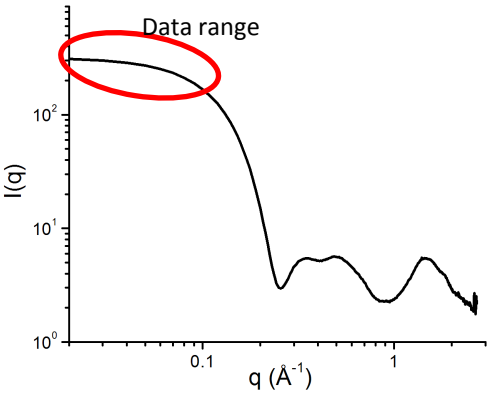
$qR_g$<1.3 for globular;
$qR_g$<0.8 for elongate

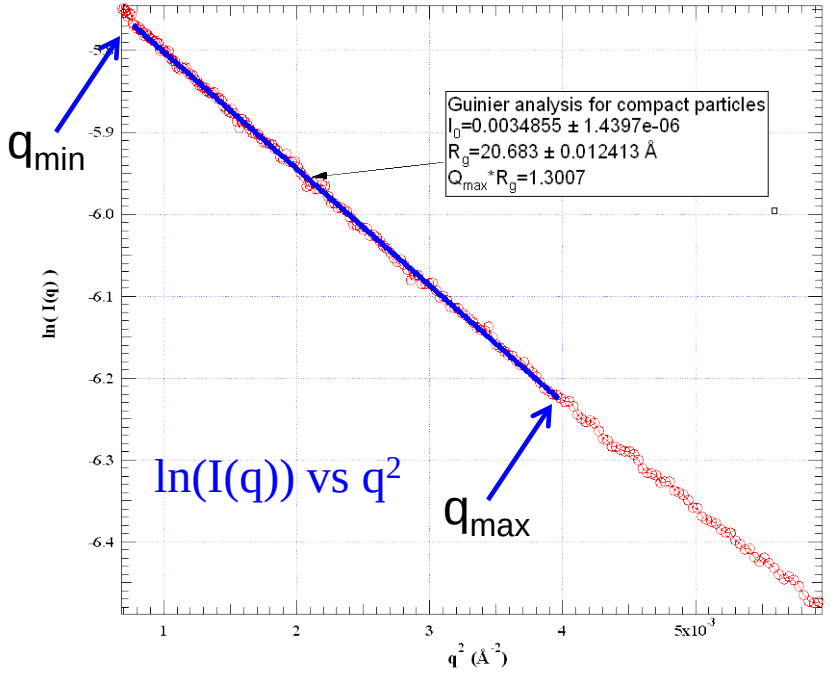
***$R_g$***: radius of gyration
***I(0)***: forward scattering

$$I(q) = \frac{2I(0)}{q^4 R_G^4}\left(q^2 R_G^2 - 1 + e^{-q^2 R_G^2}\right)$$

$qR_g$<1.4 for elongate

André Guinier (1911-2000)



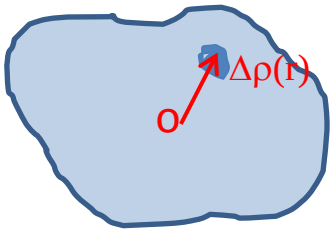

To get reliable Guinier plot / $R_g$ analysis:

- $q_{min} \leq \pi/D_{max}$
- $q_{max}$*$R_g$<1.3 for globular; <0.8 for enlongate
- Multiple (≥5) data points in linear fashion

# Radius of gyration: a size parameter

Radius of gyration ($R_g$) is the square root of the averaged squared distance of each scatterer point from the center weighted by excess electron.

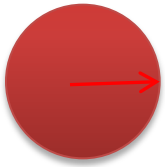
$$R_g = \left( \frac{\int \Delta\rho(r) r^2 dV}{\int \Delta\rho(r) dV} \right)^{\frac{1}{2}}$$




➢For sample geometric objects:

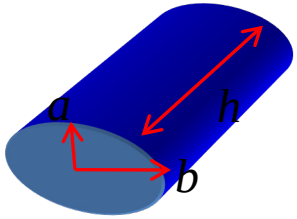
Sphere
(radius R)

$$R_g^2 = (3/5)R^2$$




Elliptic cylinder
(semi-axes a,b; height h)

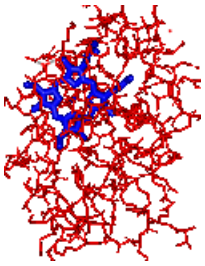
$$R_g^2 = \frac{a^2 + b^2}{4} + \frac{h^2}{12} = R_c^2 + \frac{h^2}{12}$$




➢For a molecule with known model:

$$R_g^2 \approx \frac{\sum_j \Delta n_j r_j^2}{\sum_j \Delta n_j}$$

excess electrons



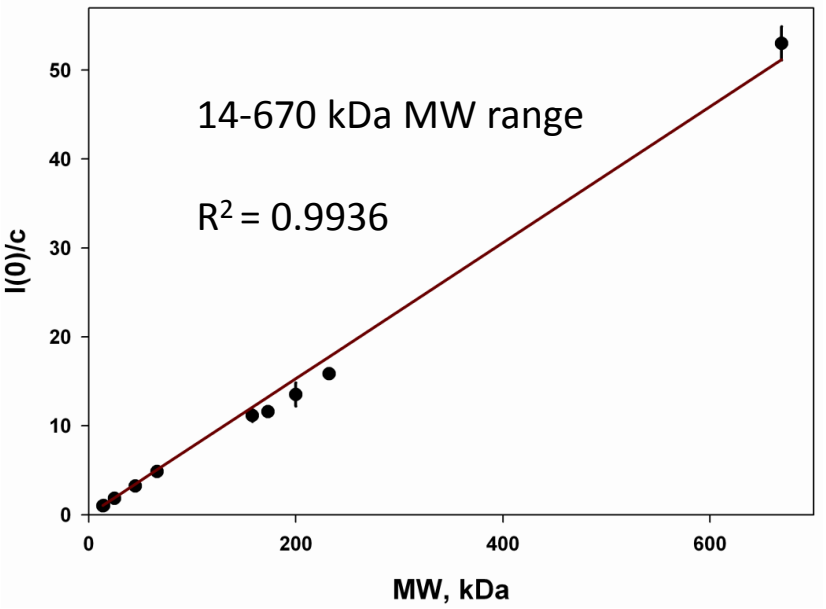
- Rg of a molecule in well-folded conformation is likely smaller than that in extended conformation
- Rg of a molecule in moneric state is likely smaller than that in oligomeric state

# Forward scattering I(0) measures molecular weight

$$I(0) \propto C_{(mg/ml)} MW$$

$$MW_p = I(0)_p / C_p \frac{MW_{st}}{I(0)_{st} / C_{st}}$$

MW: molecular mass/weight

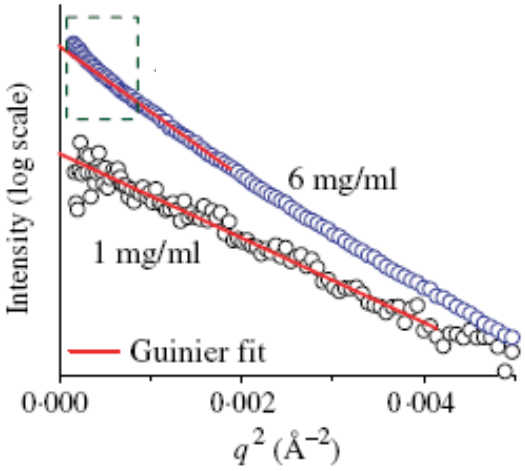



Data taken from: Mylonas, E., Svergun, D. (2007) J. Appl. Cryst. 40, s245-s249.

- Oligomeric state of biomolecules in solution is often unpredictable and difficult to determine by other means.
- Structural factor should be eliminated before using I(0) for molecular weight determination.
- The errors in MW with secondary standard or water calibration **should not exceed 10%.**
- Standard should has **same nature** of the molecules to determine, and with close MW. Using multiple standards is suggested.

# SAXS determines oligomeric state

Monomer vs dimer:




| Sample | $R_G$ | $I(0)/c$ |
|---|---|---|
| 6 mg/ml | 30 Å | 94 |
| 1 mg/ml | 22 Å | 40 |

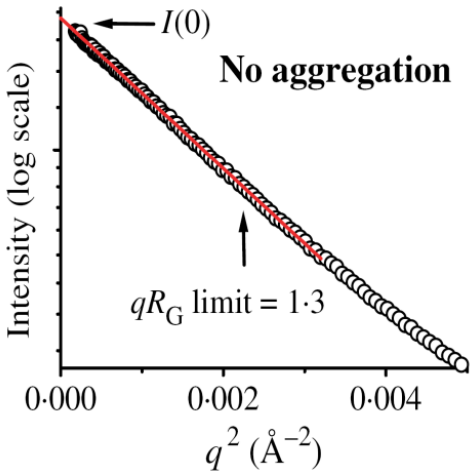
Geometric size of an aggregate grows slower than its mass.
I(0) More sensitive than $R_g$!

Figures from: Putnam, D., et al. (2007) Quart. Rev. Biophys. 40, 191-285.

# Guinier Plot: sample condition

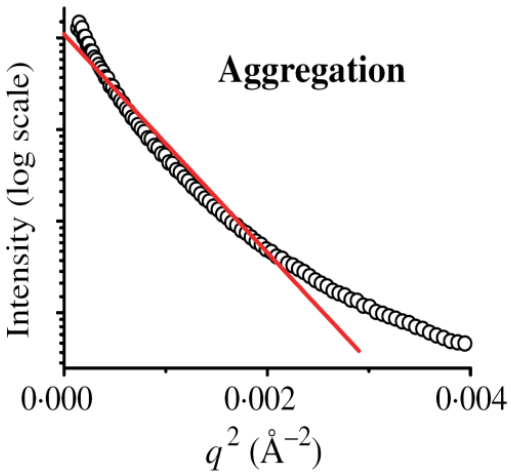
Mono-dispersed

**Normal / linear**



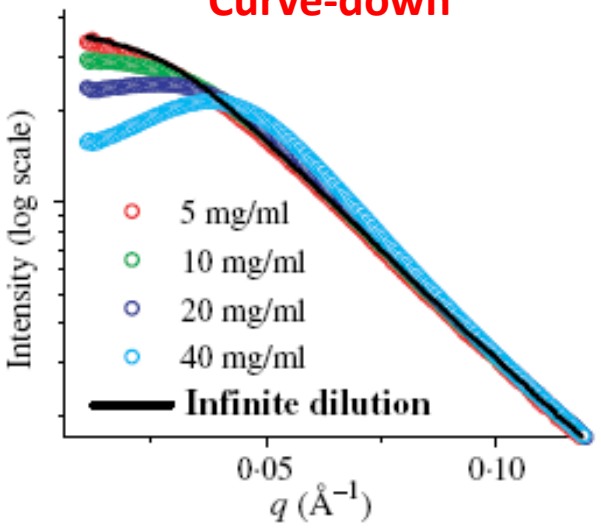

Poly-dispersed aggregates

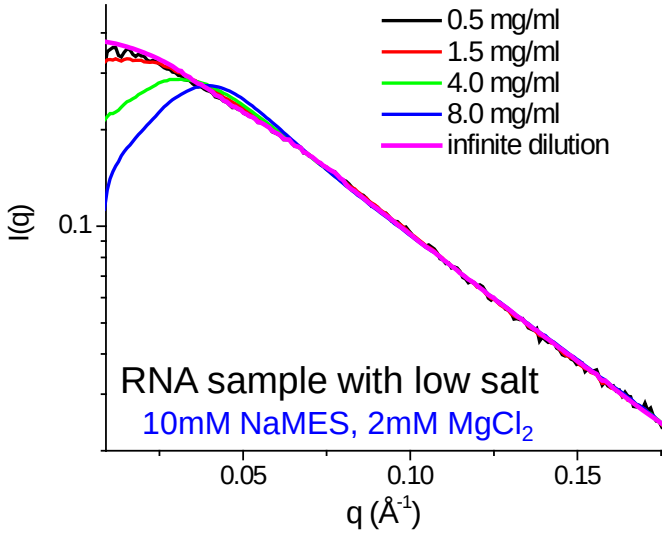
**Curve-up**




Repulsion / Structure factor

**Curve-down**

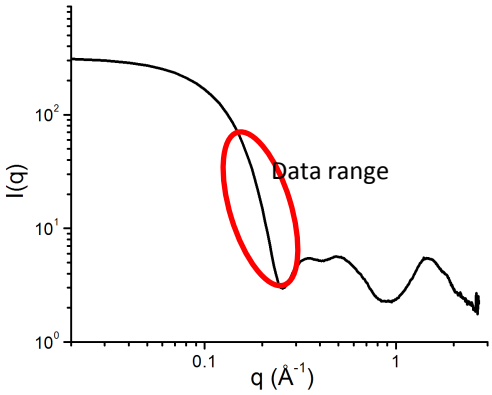



Figures from: Putnam, D., et al. (2007) Quart. Rev. Biophys. 40, 191-285.

# Porod's law

Scattering at high q values in SAXS range contains molecular shape information

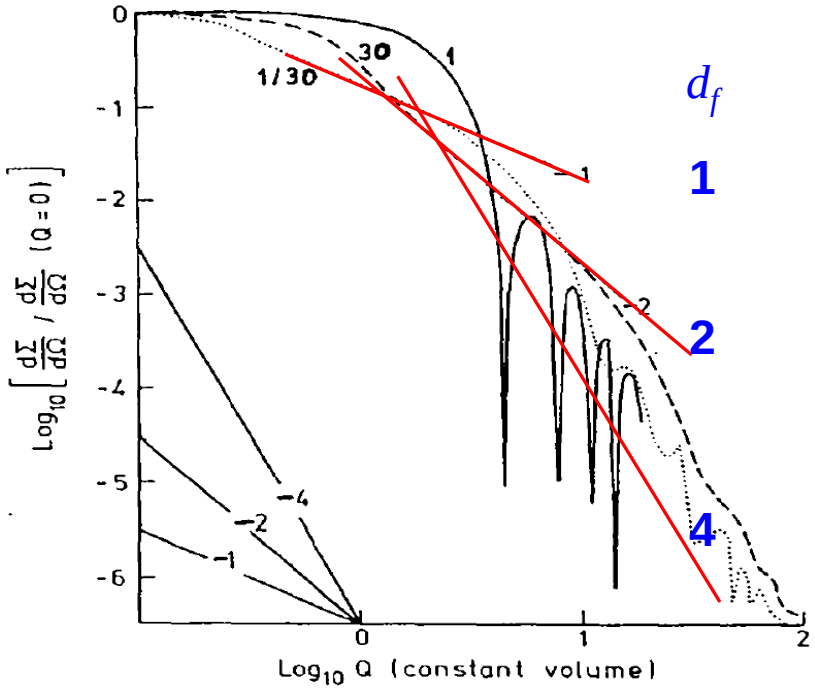



Günther Porod (1919-1984)

$$I(q) \propto q^{-d_f}$$

$d_f$ degree of freedom

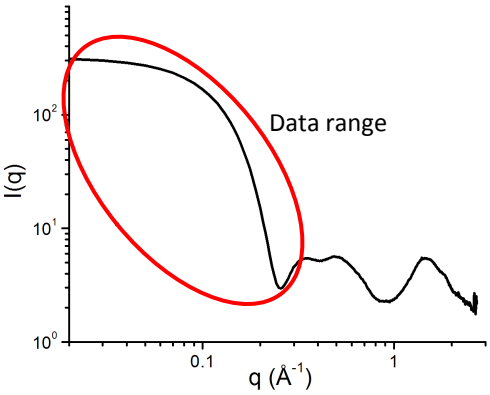
$d_f$ = 1 rod-like
2 lamellar
4 sphere




- Assumes uniform density in the particle
- Break down when atomic resolution information contribute significantly

Figure adopted from C. Windsor, *J. Appl. Cryst.* 1988, (21) 285

# Porod Invariant and Porod volume




Under the uniform electron density assumption, the integration of $q^2I(q)$ over all q range (SAXS range) is an invariant(Q) that only depends on the property of the molecule under study.

Porod Invariant Q:

$$Q \equiv \int_0^\infty q^2 I(q) dq = 2\pi^2 (\Delta\rho)^2 V$$

excess electron density

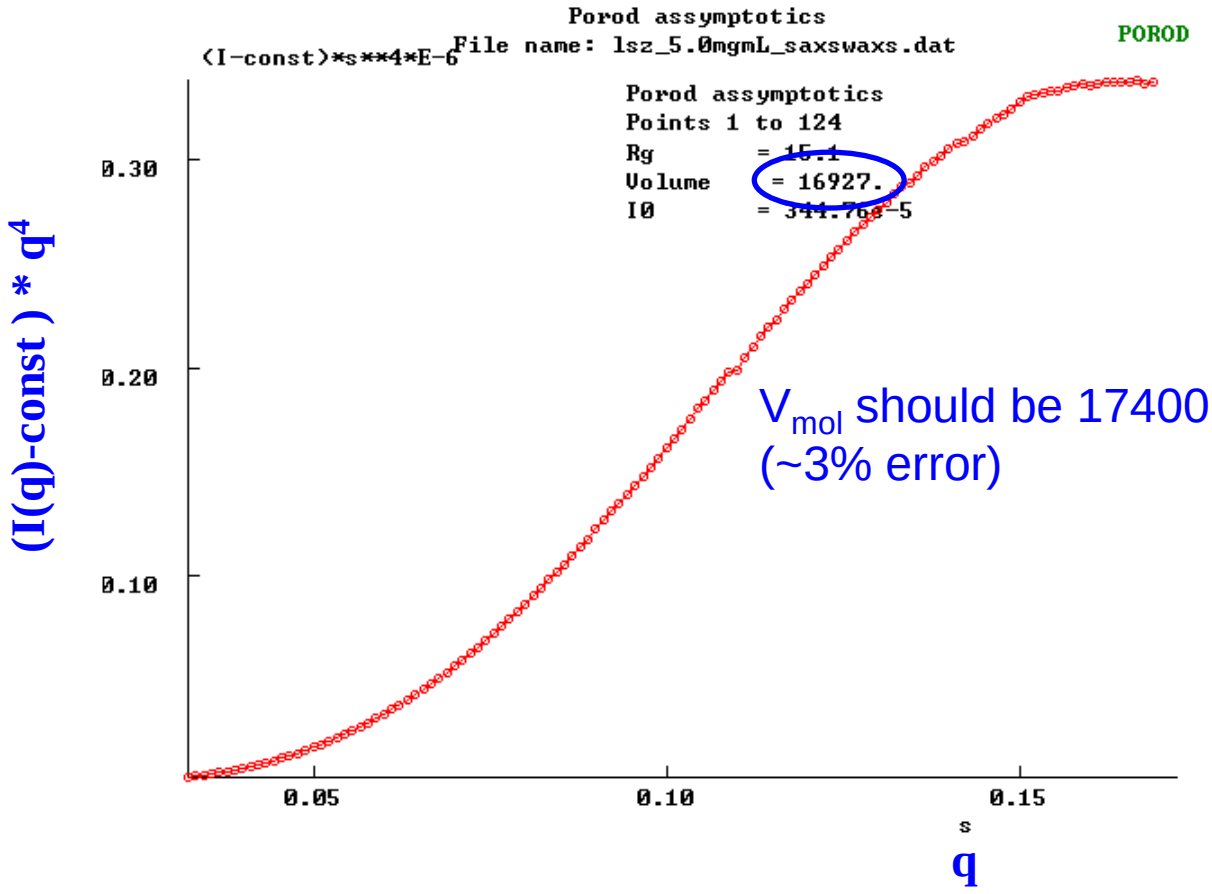
$$I(0) = (\Delta\rho V)^2$$

Porod volume: $$V = \frac{2\pi^2 I(0)}{Q}$$

- Calculation of particle volume does not require absolute data scaling.
- The volume of a molecule can be estimated solely from scattering data.
- The accuracy of the derived volume varies depending on the shape and s/n of the data.
- Inaccurate for highly asymmetric particles.
- Quality deteriorates above $q_{max}$ ~0.2 Å$^{-1}$.

# Porod volume calculation with Primus

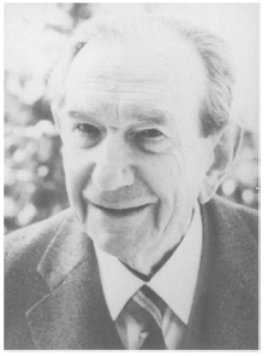
For a globular macromolecule, in Porod regime: $I(q) = a*q^{-4} + const$



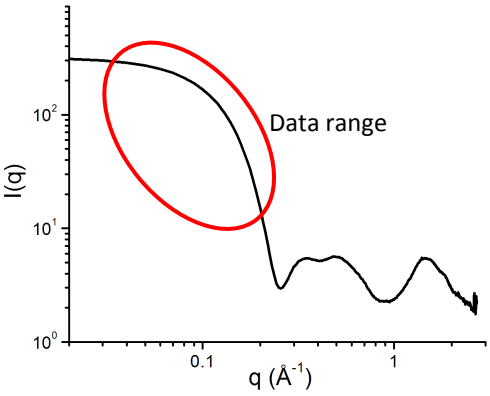

$(I(q)-const) * q^4$ vs $q$

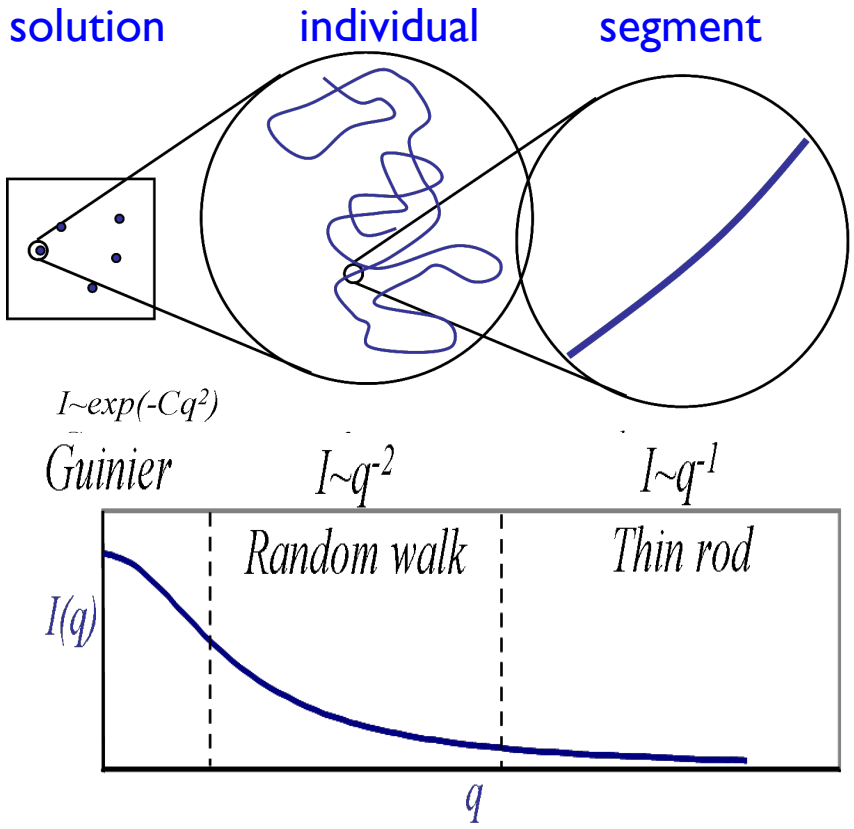
$V_{mol}$ should be 17400 $A^3$
(~3% error)

# Kratky plot: a quick check for conformation

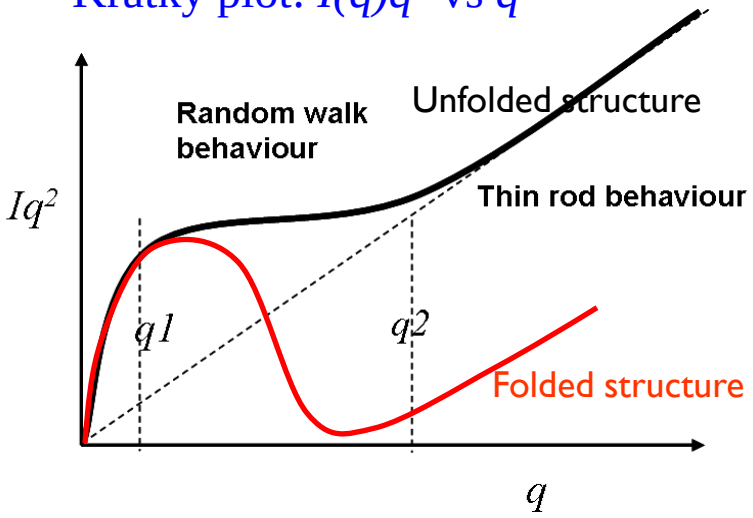


Otto Kratky (1902-1995)



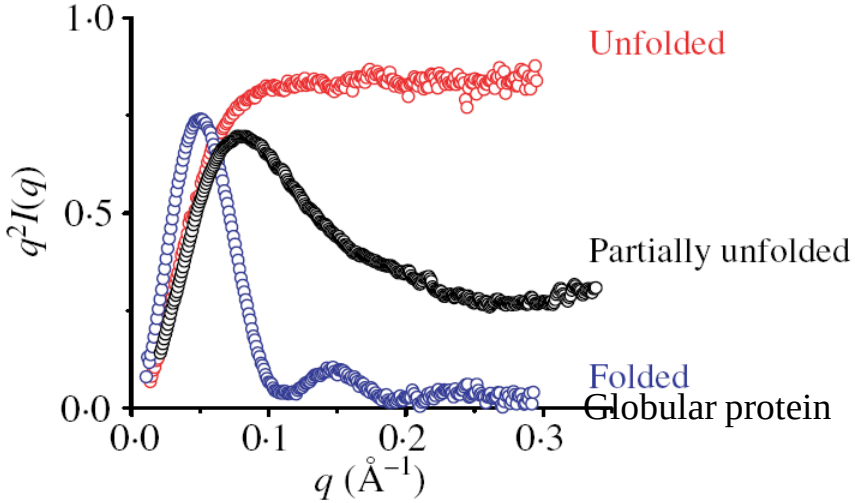




Kratky plot: $I(q)q^2$ vs $q$

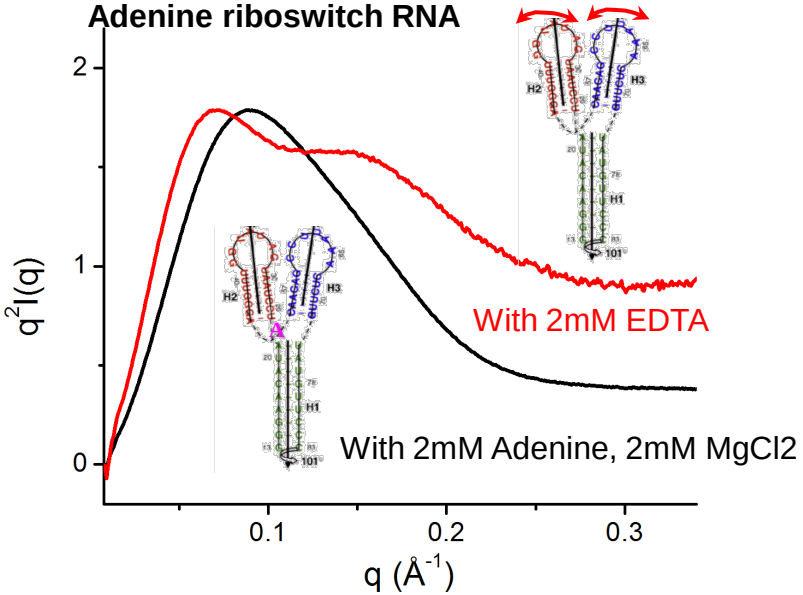



Scattering profile for a random coil type sample.

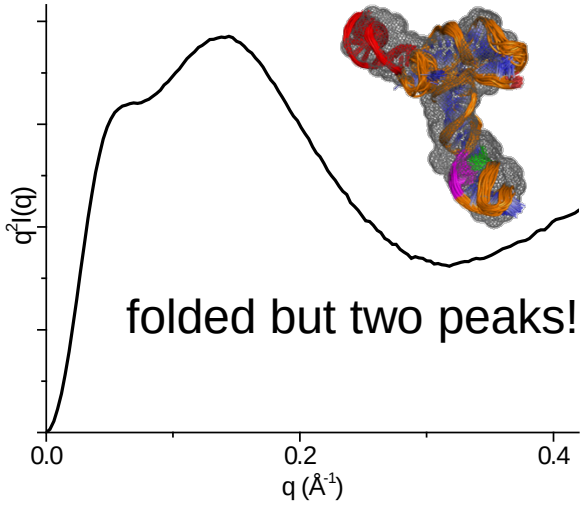
# Examples of Kratky plots for molecular conformation




Putnam, D., et al. (2007) Quart. Rev. Biophys. 40, 191-285.




Wang, J., et al. (2009) J. mol. Biol. 393, 717-734.




Folded vs compactly packed!

Zuo, X., et al. (2010) PNAS, 107, 1385-1390.

# Pair distance distribution function (PDDF)

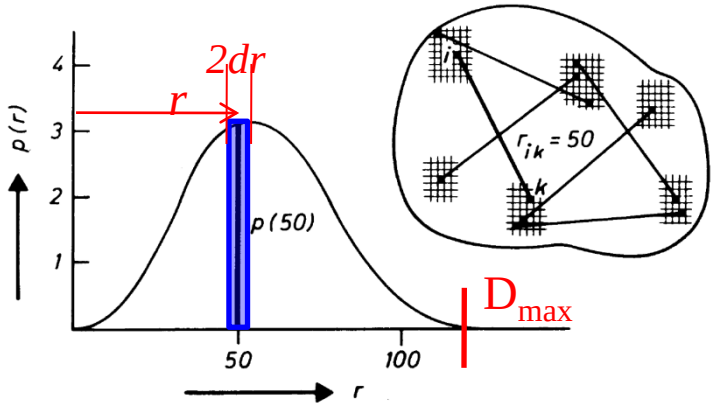
Fourier transform of I(q) (the reciprocal space measurement) is a real space function, $p(r)$:

$$p(r) = \frac{r^2}{2\pi^2}\int_0^\infty q^2 I(q)\frac{\sin qr}{qr}dq \quad \Longleftrightarrow \quad I(q) = 4\pi\int_0^\infty p(r)\frac{\sin qr}{qr}dr$$

Pair distance distribution function (PDDF, $p(r)$) is a weighted (with distance and net charge) atom-pair distance histogram/population in real space.

$$p(r) = r^2\left\langle\int_V \Delta\rho(\mathbf{u})\Delta\rho(\mathbf{u}+\mathbf{r})d\mathbf{u}\right\rangle_\Omega$$

$$I(0) = 4\pi\int_0^{D\max} p(r)dr$$



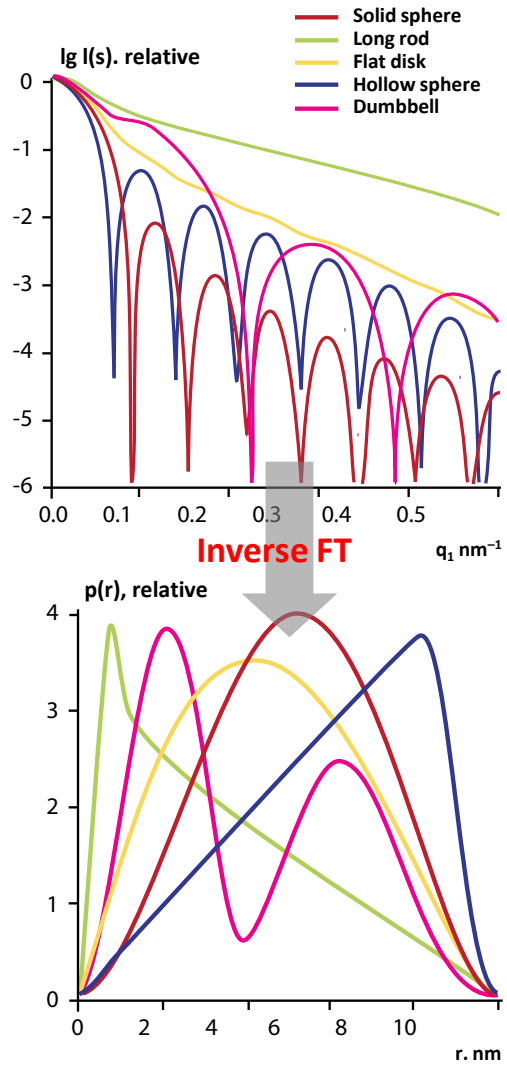

$$p(r) \sim \sum_{|\mathbf{r_j}-\mathbf{r_k}|<r+dr} 1\times\Delta n(\mathbf{r_j})\times\Delta n(\mathbf{r_k})\times r^2$$

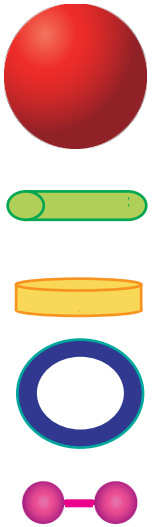
excess electrons of atom *j* over solvent

- Real space information: distance distribution, $D_{max}$, MW (I(0)), shape, …

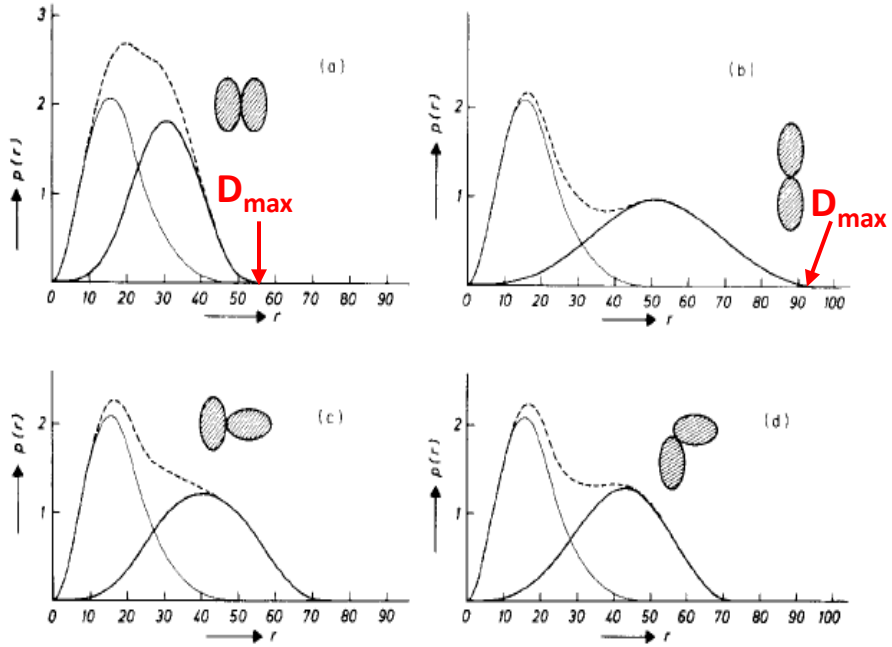
# PDDF for shape/conformation determination

Both scattering profile and PDDF may be characteristic for shape/conformation determination.
Objects with similar views in the reciprocal space may be very different in the view of real space, visa verse.



dimer with various conformations




The various dimeric conformations change p(r) and $D_{max}$ as well.

Svergun, D., Koch, M. (2003) Rep. Prog. Phys. 66, 1735-1782.

Glatter, O., & Kratky, O., Small Angle X-ray Scattering, 1982, p179

# Calculate pair distance distribution function / p(r) from SAXS

Obtaining PDDF from SAXS is an ill-condition problem because limited q range. p(r) calculated from direct integration is severely distorted by the q truncation.

$$p(r) = \frac{r^2}{2\pi^2}\int_0^\infty q^2 I(q)\frac{\sin qr}{qr}dq$$

Program GNOM is an indirect Fourier transform program with perceptual criteria, for example: oscillations, systematic deviations, discrepancy, stability, positivity, central shape, etc.

target function: $$T_\alpha[p(r)] = \left|I(q) - \mathbf{F}[p(r)]\right|^2 + \alpha * \Omega[p(r)]$$

The default parameters optimized for globular shape.

# GNOM operation

Principles of operation:

*1. Minimal oscillations in the fitted P(r)*

*2. Non-negativity of the P(r).*

*3. Proper shape for the central part of P(r).*

*4. p(0)=0 and p($D_{max}$)~0.*

*5. Highest stability of the regularized solution*

*6. Good data fit quality*

*7. Minimal systematic deviations of the I(q) fit*

*(highest number of sequential residuals changing sign)*

Information obtained from fit:

*1. PDDF*

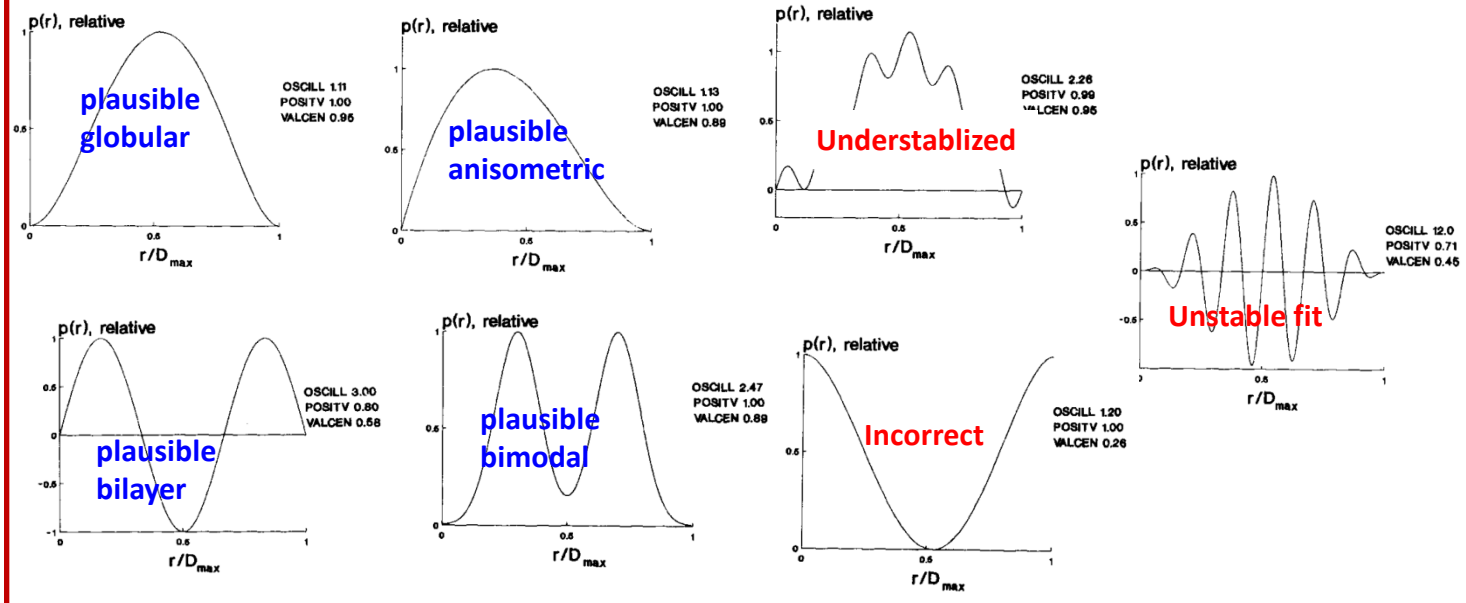
*2. $D_{max}$*

*3. $R_g$*

*4. I(0)*

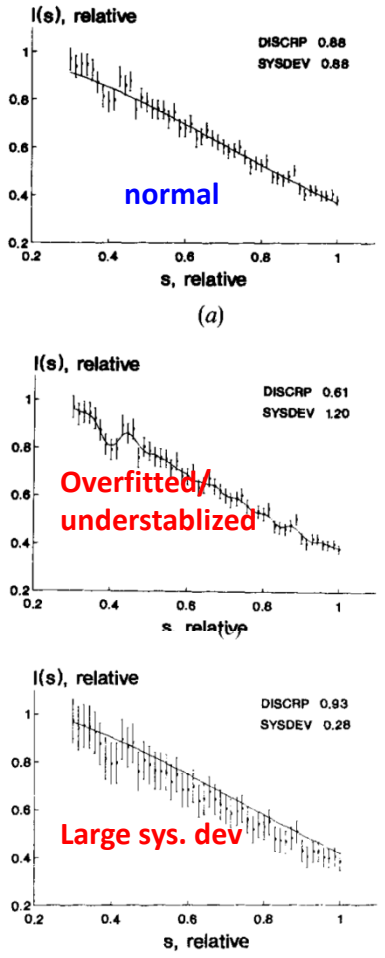
*5. Data extrapolation q → 0*

## PDDF profiles




Svergun, D. (1992) J. Appl. Cryst. 25, 495-503.

## SAXS fitting



# GNOM input & output

```
          ***  PLEASE SELECT THE FIRST DATA FILE NAME  ***

 Working directory: C:\Documents and Settings\Zuox\Desktop\workshop\demo_data\g
nom\
 File to be opened: test_30.dat
 Output file                          [ gnom.out    ] : gtest30.out
 No of start points to skip           [    0        ] :
 Run title:  02.16440.01489
 Number of points in the run is  100
 Input data, second file              [ none        ] :
 No of end points to omit             [    0        ] :
 Total number of input data points read is  100
 Angular range as read: from  0.00300   to  0.30000
 Angular scale (1/2/3/4)              [    1        ] :
 Kernel already calculated      (Y/N) [   No        ] :
 Type of system       (0/1/2/3/4/5/6) [    0        ] :
 Zero condition at r=rmin       (Y/N) [   Yes       ] :
 Zero condition at r=rmax       (Y/N) [   Yes       ] :
   --  Arbitrary monodisperse system  --
  Rmin=0.  Rmax is maximum particle diameter
 Rmax for evaluating p(r)                          : 150
 Kernel-storage file name             [ kern.bin    ] :
 Experimental setup           (0/1/2) [    0        ] :
 Evaluating design matrix. Please wait...

 Evaluating stabilizer matrix. Please wait ...
  The measure of inconsistency AN1 equals to    0.1657E+00
     Alpha    Discrp  Oscill  Stabil  Sysdev  Positv  Valcen    Total
  0.1597E+02  0.6934  1.1855  0.0086  0.3434  1.0000  0.8706  0.71410

 Parameter    DISCRP    OSCILL    STABIL    SYSDEV    POSITV    VALCEN
 Weight        1.000     3.000     3.000     3.000     1.000     1.000
 Sigma         0.300     0.600     0.120     0.120     0.120     0.120
 Ideal         0.700     1.100     0.000     1.000     1.000     0.950
 Current       0.693     1.186     0.009     0.343     1.000     0.871

 Estimate      1.000     0.980     0.995     0.000     1.000     0.646

  Angular   range    :     from    0.0030   to    0.3000
  Real space range    :     from      0.00   to    150.00

  Highest ALPHA (theor) :   0.416E+04                JOB = 0
  Current ALPHA         :   0.160E+02   Rg :  0.518E+02   I(0) :   0.211E+01

         Total  estimate : 0.714  which is  A REASONABLE  solution

               ===   Select one of the following options  ===
                   - - - - - - - - - - - - - - - - - - -
                  CR     ---  to accept the solution and   EXIT
             -(NewAlpha) ---  to manually change           ALPHA
           1,2,3,4,5,6   ---  to change weight/sigma of PARAMETERS
                  7      ---  to maximize a new  total   ESTIMATE
                  8      ---  to replot the              SOLUTION
 Your choice :
```

SAXS data input
GNOM output file

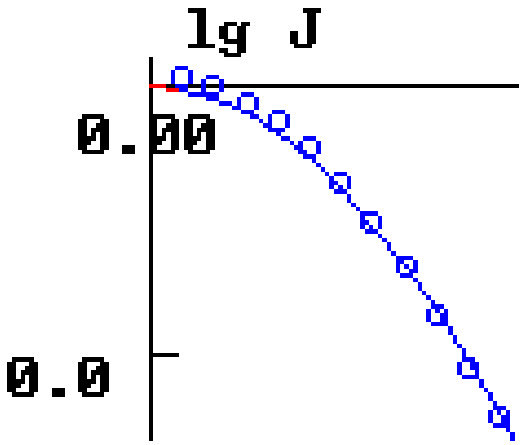
Using 0 for Synchrotron data

$D_{max}$

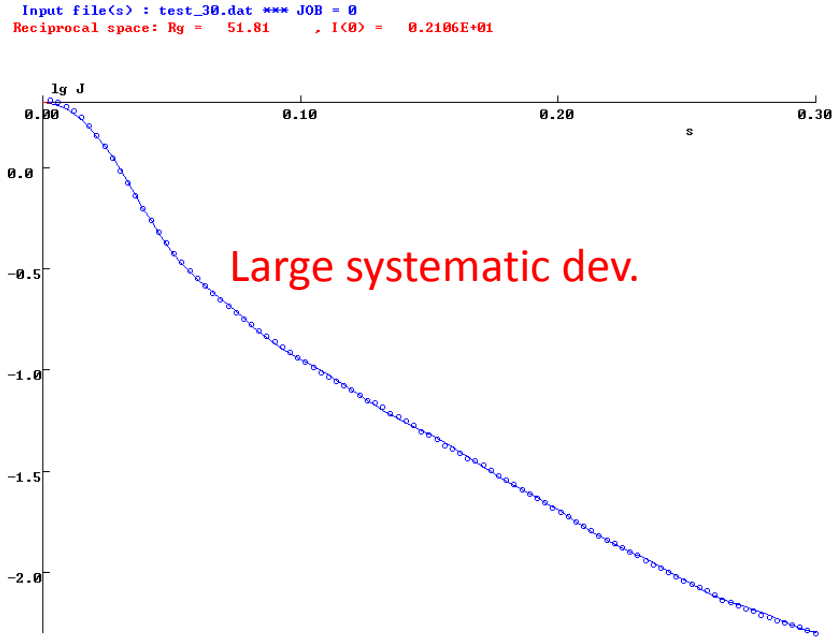
Default parameters are optimized for globular shape.

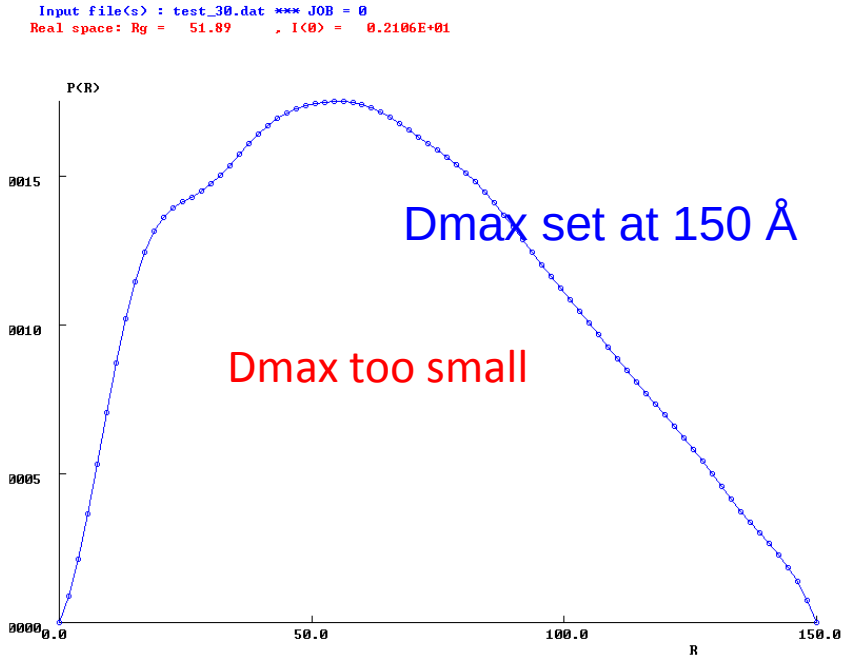
Screen capture from GNOM program

# Incorrect PDDF fit (1)

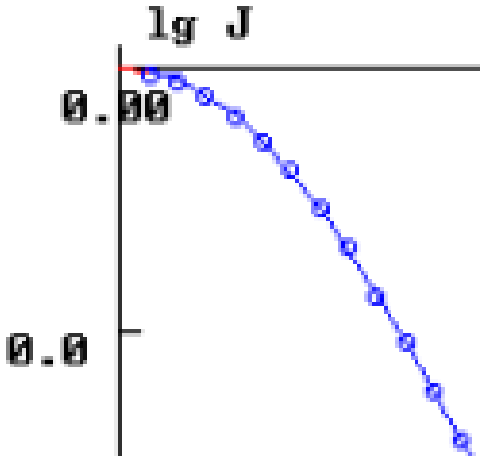



$D_{max}$ is under-estimated

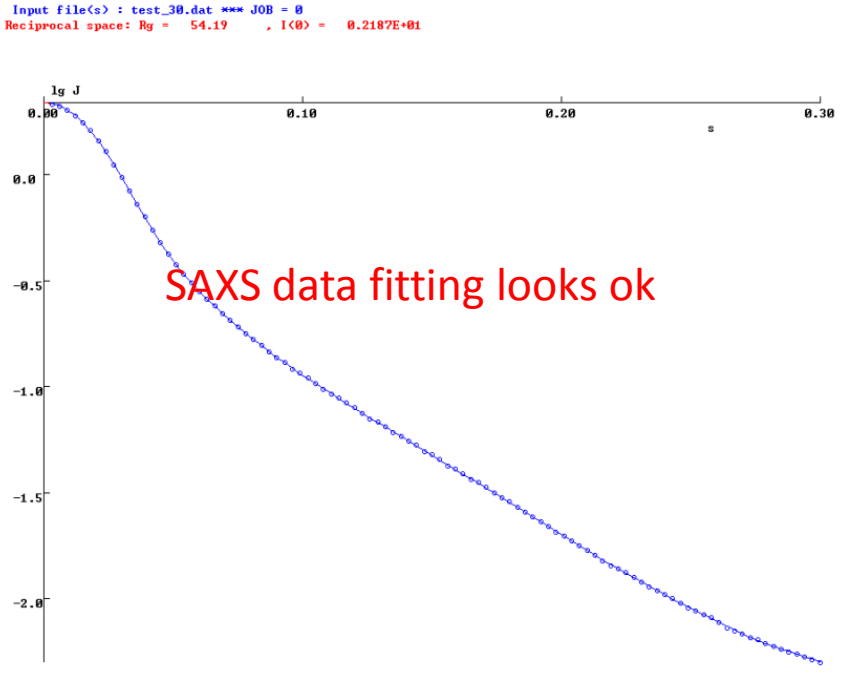


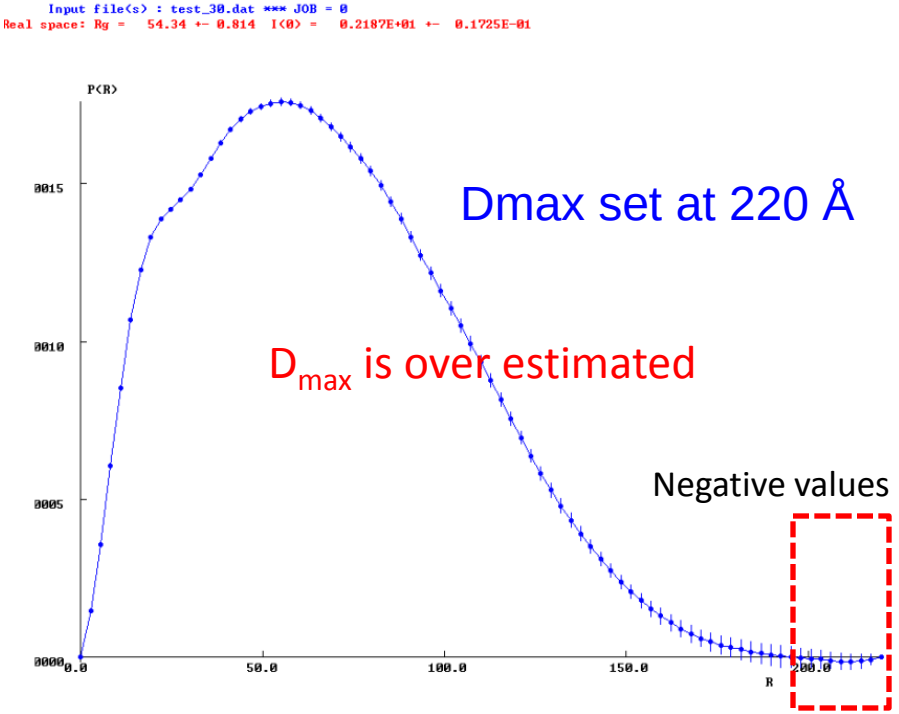
# Incorrect PDDF fit (2)




$D_{max}$ is over-estimated




SAXS data fitting looks ok



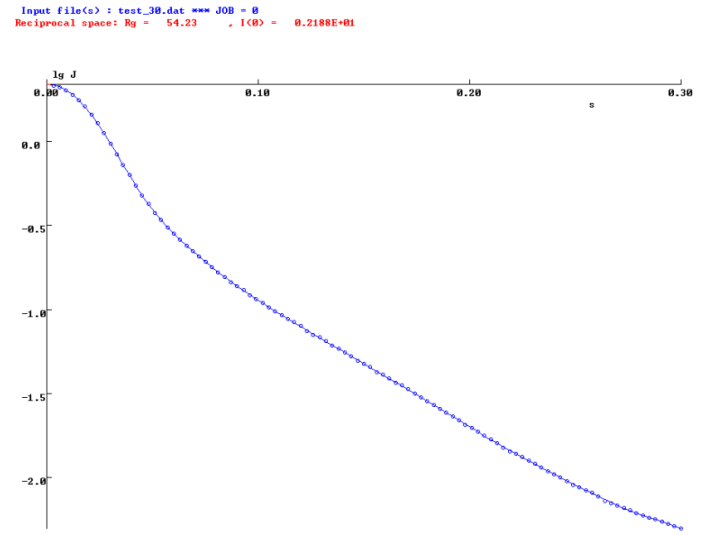

Dmax set at 220 Å

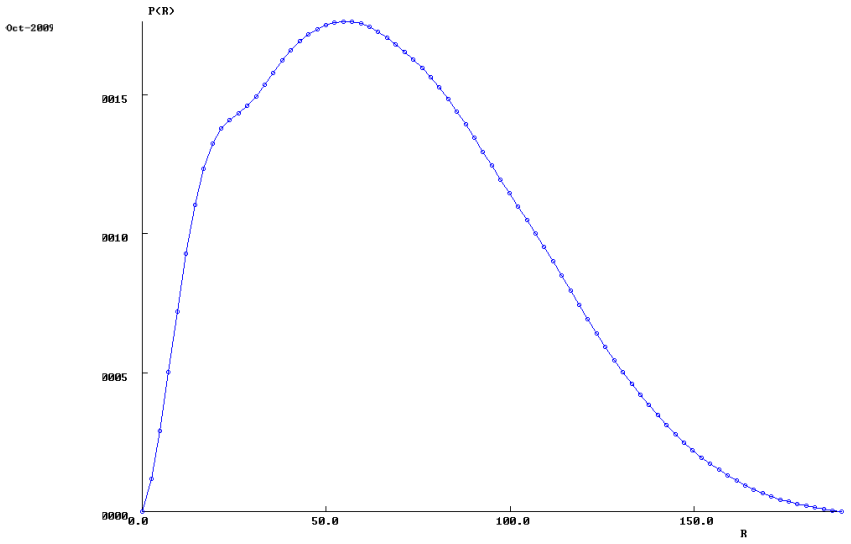
$D_{max}$ is over estimated

Negative values

# Obtaining correct PDDF fit

```
Next data set           (Yes/No/Same) [ y           ] :
Input data, first file                [ test_30.dat ] :
Output file                           [ gtest.out   ] :
No of start points to skip            [    0        ] :
Run title:  02.16440.01489
Number of points in the run is  100
Input data, second file               [ none        ] :
No of end points to omit              [    0        ] :
Total number of input data points read is  100
Angular range as read: from  0.00300   to  0.30000
Angular scale (1/2/3/4)               [    1        ] :
Kernel already calculated       (Y/N) [   No        ] :
Type of system        (0/1/2/3/4/5/6) [    0        ] :
Zero condition at r=rmin        (Y/N) [   Yes       ] :
Zero condition at r=rmax        (Y/N) [   Yes       ] :
   --  Arbitrary monodisperse system  --
 Rmin=0.  Rmax is maximum particle diameter
Rmax for evaluating p(r)              [ 190         ] :
Kernel-storage file name              [ kern.bin    ] :
Experimental setup            (0/1/2) [    0        ] :
Evaluating design matrix. Please wait...


Parameter    DISCRP    OSCILL    STABIL    SYSDEV    POSITV    VALCEN
Weight        1.000     3.000     3.000     3.000     1.000     1.000
Sigma         0.300     0.600     0.120     0.120     0.120     0.120
Ideal         0.700     1.100     0.000     1.000     1.000     0.950
Current       0.203     1.447     0.001     0.889     1.000     0.835
              - - - - - - - - - - - - - - - - - - - - - - - - - - -
Estimate      0.064     0.715     1.000     0.424     1.000     0.399

 Angular   range    :     from    0.0030   to    0.3000
 Real space range    :     from      0.00   to    190.00

Highest ALPHA (theor) :   0.506E+04                JOB = 0
Current ALPHA         :   0.100E+01   Rg :  0.543E+02   I(0) :   0.219E+01

         Total  estimate : 0.657  which is  A REASONABLE  solution

              ===   Select one of the following options  ===
                  - - - - - - - - - - - - - - - - - - - -
                   CR     ---  to accept the solution and   EXIT
                -(NewAlpha) ---  to manually change          ALPHA
               1,2,3,4,5,6  ---  to change weight/sigma of PARAMETERS
                     7      ---  to maximize a new  total   ESTIMATE
                     8      ---  to replot the              SOLUTION
Your choice :
```




$D_{max}$ error: 5~10% of $D_{max}$

Put $p(r_{max})=0$ "No" first, give $r_{max}$ a number in range $2R_g$-$6R_g$, and see where $p(r_{max})$ ends. Then change $r_{max}$. Alpha is another very important parameter affects the fitting, tune alpha to have a stable fit. Alpha should not be too small (>0.1)

# $R_g$, I(0) and data extrapolation

```
Reciprocal space: Rg =   56.77     , I(0) =   0.2265E+01
Real space: Rg =   57.07 +- 0.862  I(0) =   0.2266E+01 +-  0.2212E-01
```

## Reciprocal space parameters:

Directely obtained from Guinier plot

## Real space parameters:

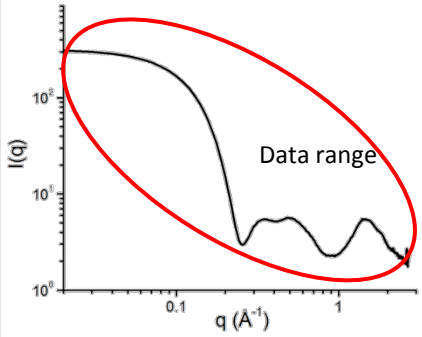
$$I(0) = 4\pi \int_0^{D\max} p(r)dr$$

$$R_G^2 = \int_0^{D_{max}} r^2 P(r)\mathrm{d}r \Big/ \int_0^{D_{max}} P(r)\mathrm{d}r$$

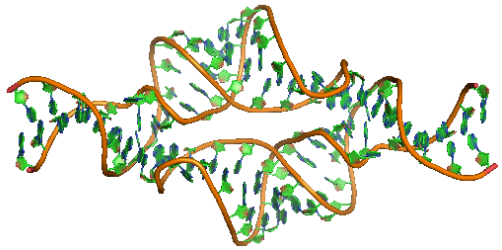
## data extrapolation

| S | J EXP | ERROR | J REG | I REG |
|---|---|---|---|---|
| 0.0000E+00 | | | | 0.2266E+01 |
| 0.5569E-03 | | | | 0.2265E+01 |
| 0.1114E-02 | | | | 0.2263E+01 |
| 0.1671E-02 | | | | 0.2259E+01 |
| 0.2228E-02 | | | | 0.2253E+01 |
| 0.2784E-02 | | | | 0.2247E+01 |
| ... | | | | |
| 0.1058E-01 | | | | 0.2010E+01 |
| 0.1114E-01 | | | | 0.1985E+01 |
| 0.1169E-01 | | | | 0.1959E+01 |
| 0.1225E-01 | | | | 0.1932E+01 |
| 0.1281E-01 | | | | 0.1904E+01 |
| 0.1337E-01 | | | | 0.1875E+01 |
| 0.1392E-01 | | | | 0.1846E+01 |
| 0.1448E-01 | | | | 0.1817E+01 |
| 0.1504E-01 | 0.1797E+01 | 0.2309E-01 | 0.1786E+01 | 0.1786E+01 |
| 0.1559E-01 | 0.1752E+01 | 0.2324E-01 | 0.1756E+01 | 0.1756E+01 |
| 0.1615E-01 | 0.1729E+01 | 0.2264E-01 | 0.1725E+01 | 0.1725E+01 |
| 0.1671E-01 | 0.1693E+01 | 0.1664E-01 | 0.1693E+01 | 0.1693E+01 |
| 0.1726E-01 | 0.1658E+01 | 0.1873E-01 | 0.1662E+01 | 0.1662E+01 |
| ... | | | | |

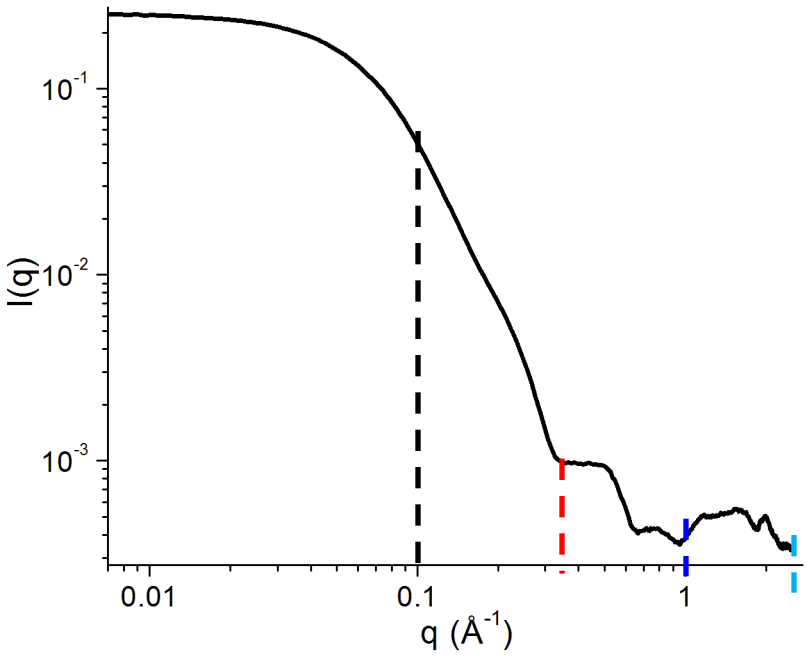
Extrapolated data

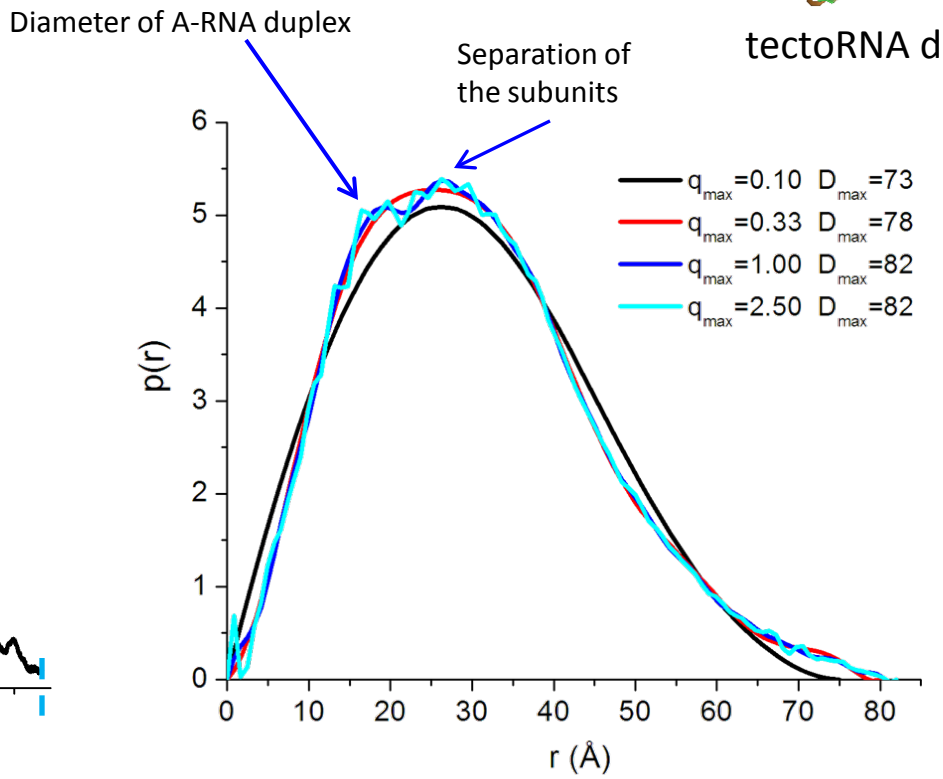



# PDDF at various resolutions



tectoRNA dimer



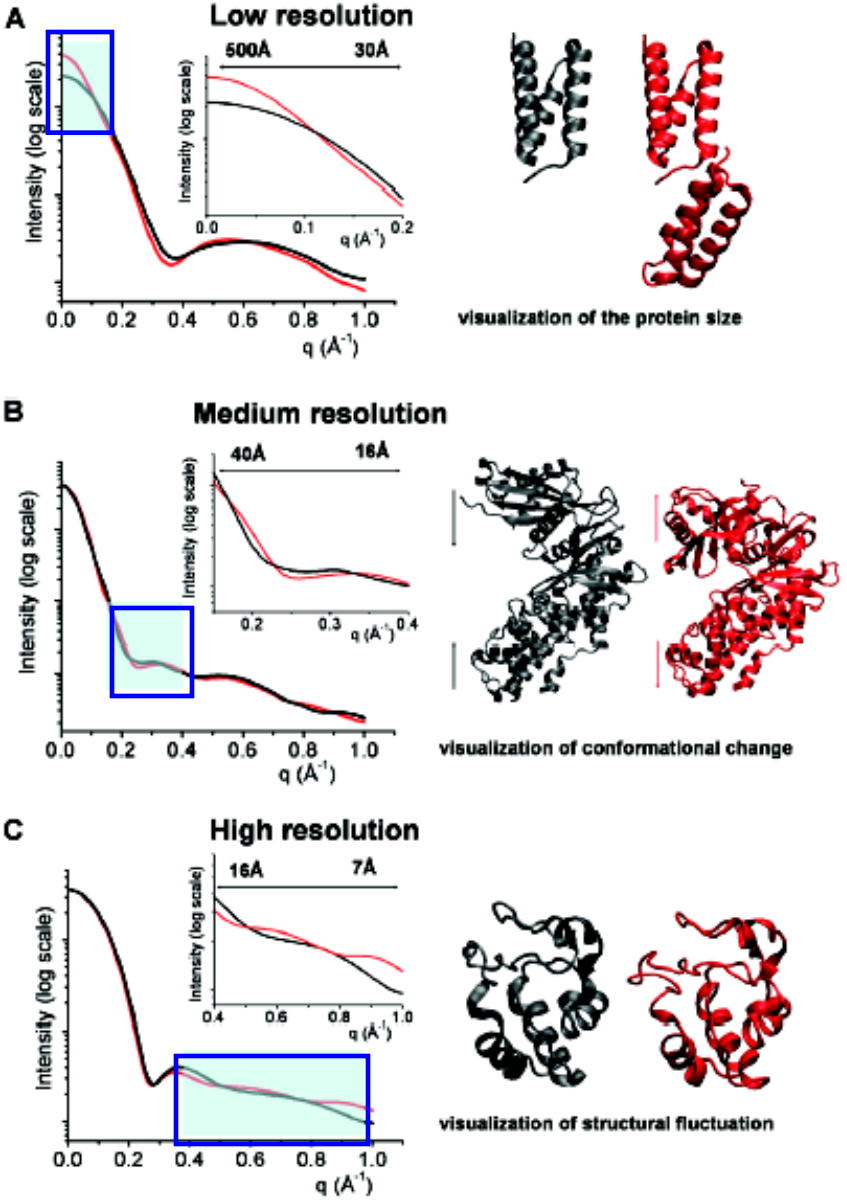

- PDDF/p(r) obtained from certain q-range reflects the resolution defined by $q_{max}$.
- Dmax may also be affected.
- High enough $q_{max}$ needed for PDDF used in structure/conformation analysis.

Zuo, X., Wang, J., Foster, T. R., Schwieters, C. D., Tiede, D. M., Butcher, S. E. & Wang, Y. X. (2008). J Am Chem Soc 130, 3292-3

# 2. Structural analysis from SAXS data

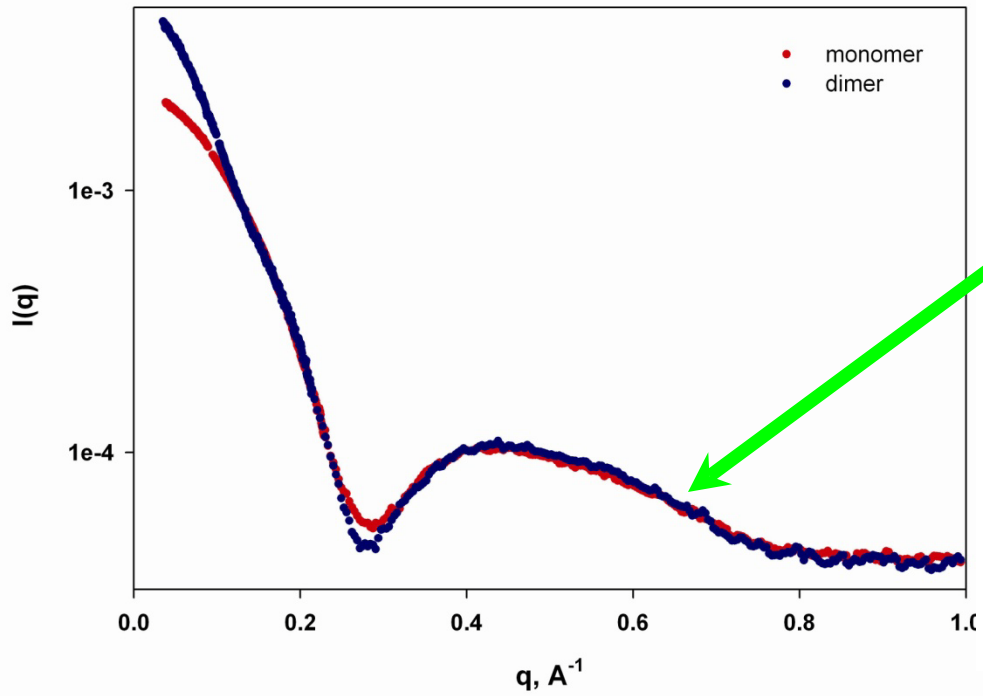
- Structure/data relationship
- Low-resolution model/shape reconstruction
- Multi-subunit systems: rigid body refinement
- Flexible and unfolded systems
- Multiple components: micelle-embedded systems
- Hybrid NMR/SAXS approaches

# What can SAS data tell us about molecular structure? Actually, a lot...




A diverse range of structural characteristics can be captured by solution scattering. **Resolution range**, **signal/noise**, and **prior structural information** will determine how much it reveals. These factors define the information content of the solution scattering data.

# Molecular size from the low-angle scattering data: monomer vs dimer

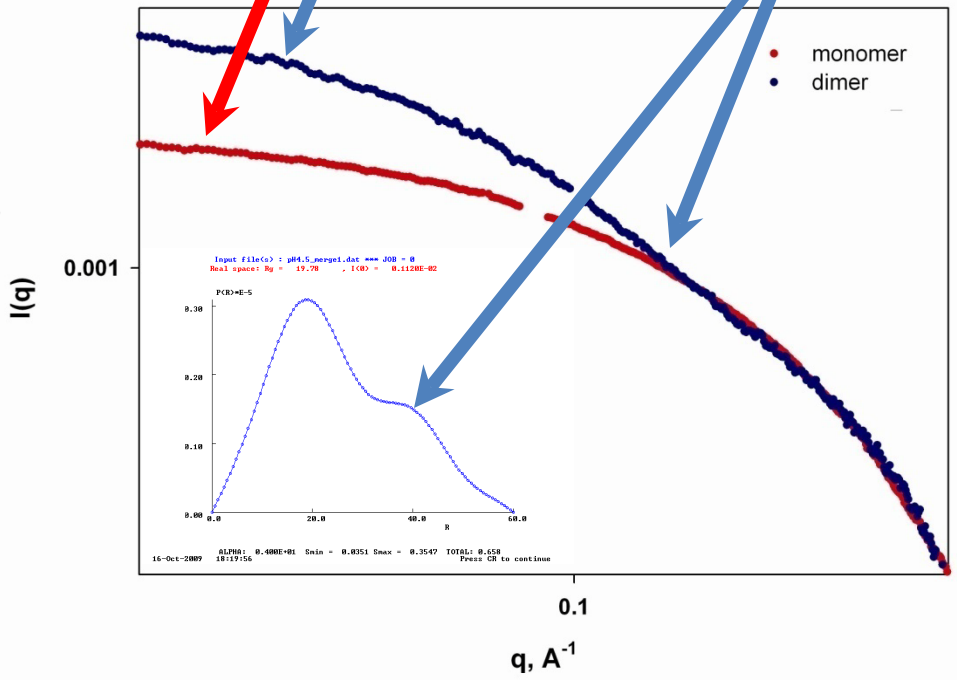



Individual domain structure
is identical for the monomer and dimer

Rg = 15.3 Å for the monomer

Rg = 19.3 Å for the dimer

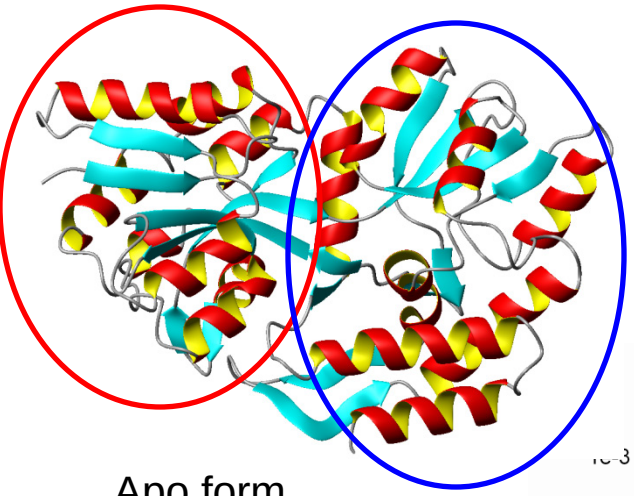
Dimer domain separation
is ~ 35 Å



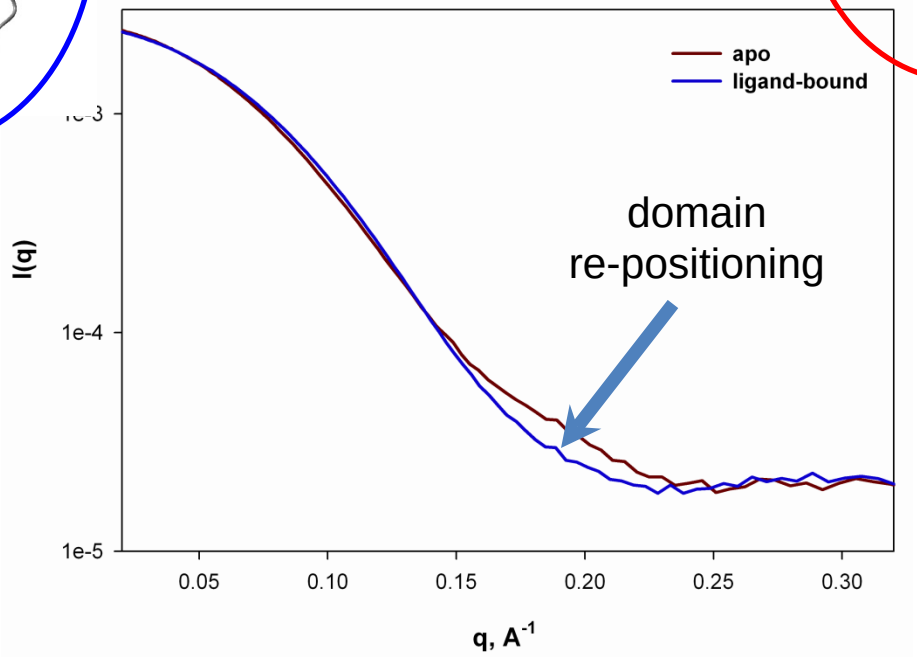

Very little prior information in needed for this P(r)- and Guinier-based analyses!

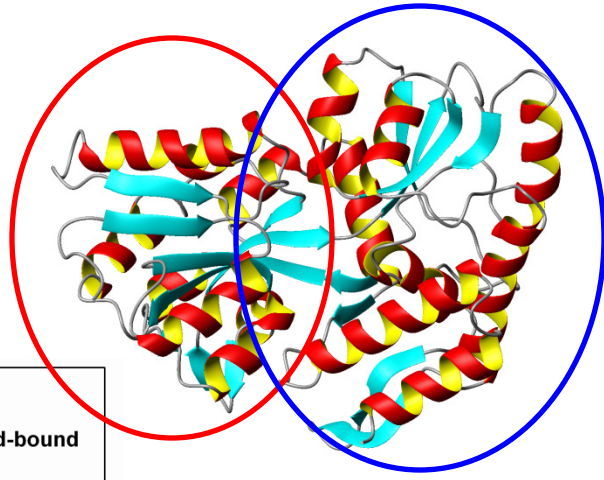
Low-res density reconstructions operate comfortably in this regime.

# Conformational change from SAS data: domain rearrangement



Apo form
Rg = 22.6 Å



Ligand-bound form
Rg = 21.5 Å

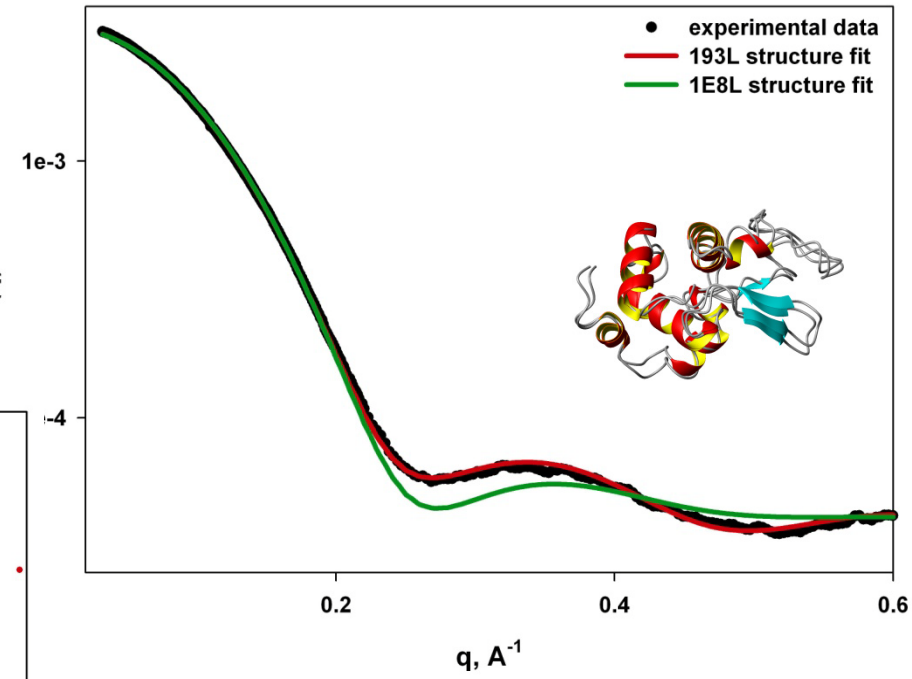
Precise prior structural information and additional restraints are very useful here!

This situation is more challenging for low-res density reconstructions.

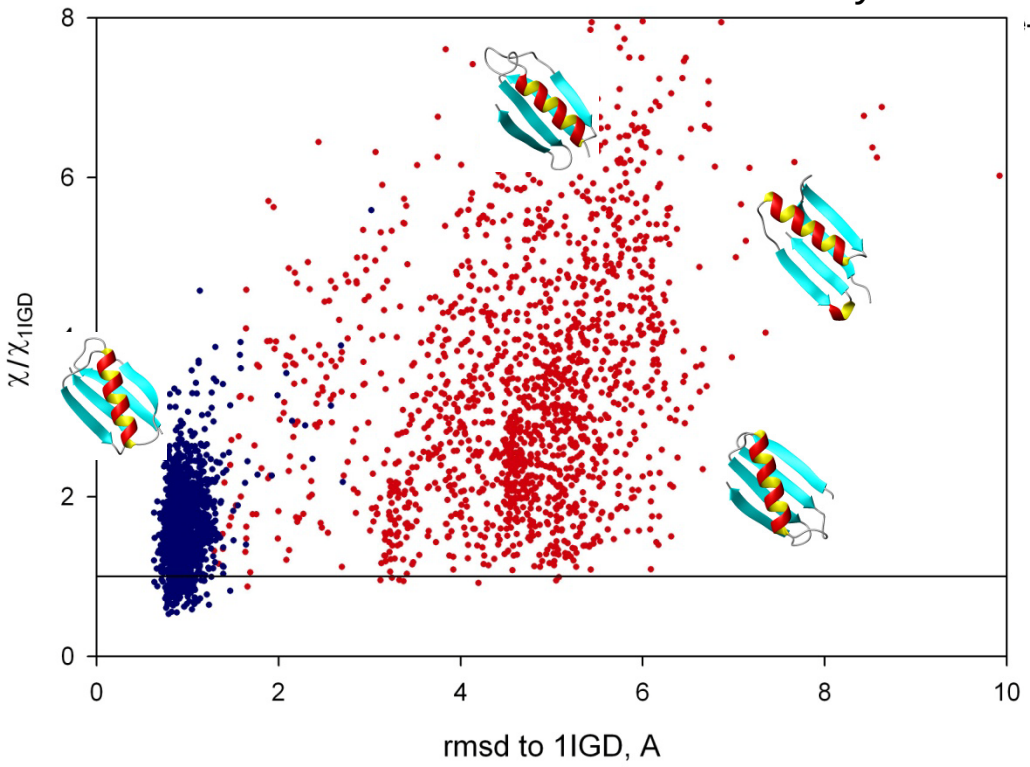
NMR-SAXS refinement commonly operates in this regime.

# SAXS data can be sensitive to subtle structural differences

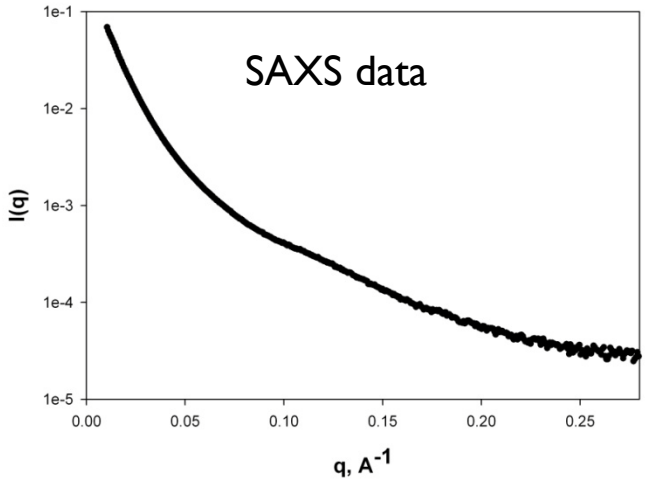
Hen egg white lysozyme:
193L and 1E8L models (1.5 Å bb rmsd)

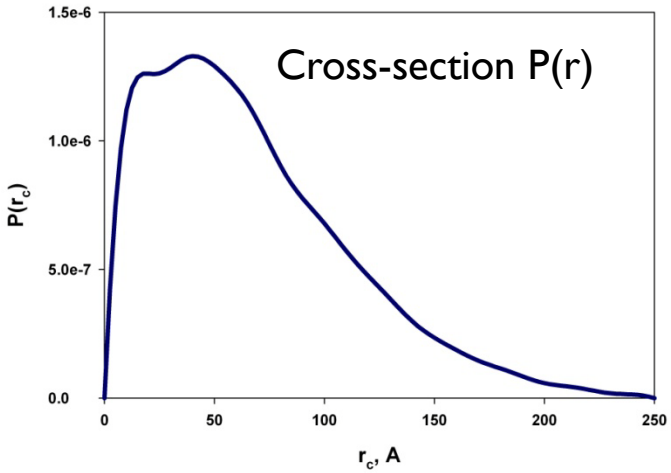


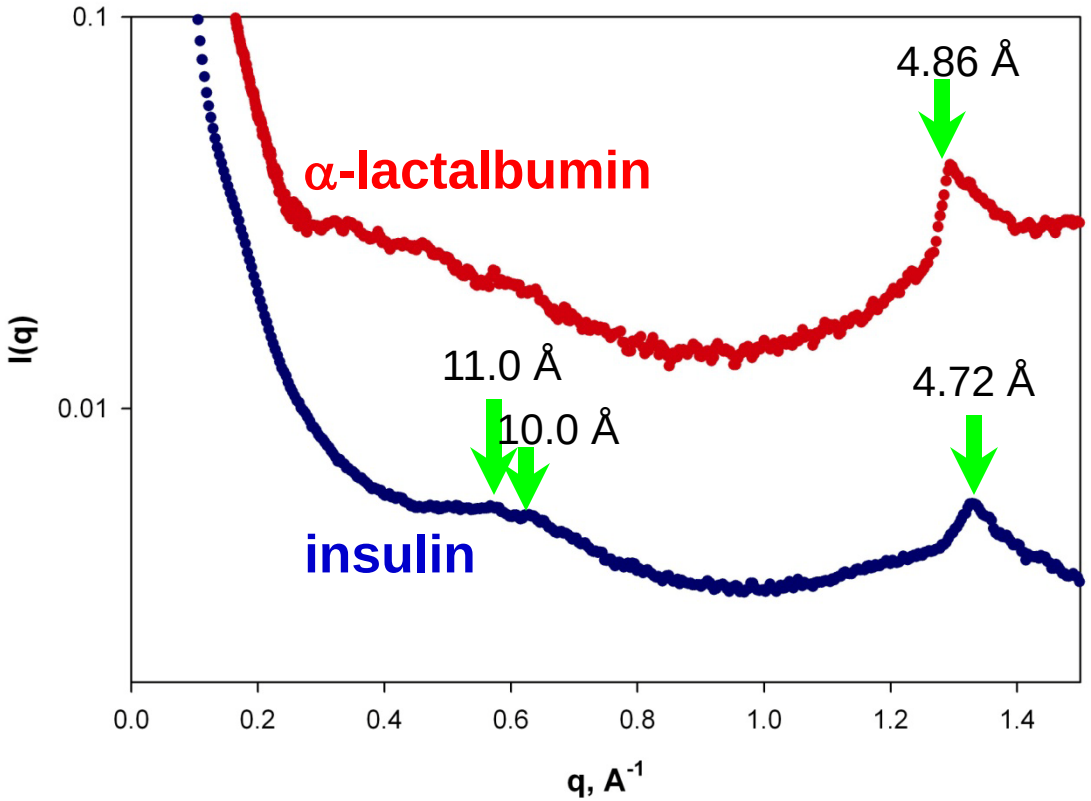

Protein G:
Rosetta and chemical shifts-Rosetta decoy sets



# Amyloid fibrils: SAXS data define cross-section shape and WAXS data are similar to fibre diffraction



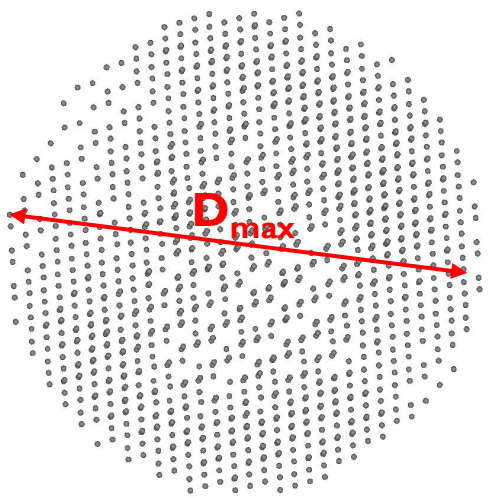



# 3D shape reconstructions from SAXS data: a general idea

Obtaining 3D shapes from 1D SAXS data is an ill-defined problem that can be solved by regularizing the fitted models.

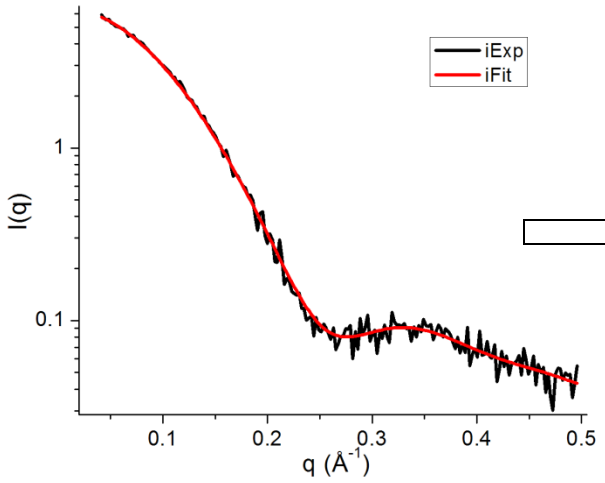
Imposing prior restraints on the fitted models such as non-negativity and compactness/connectivity greatly increases solution stability.



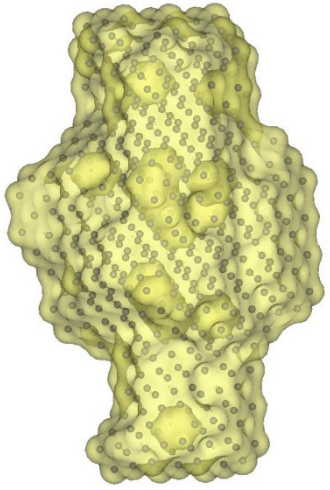

Model Space

MC

SA




Fitted Data



Low-resolution model

## Available Programs:

- Genetic Algorithm: DALAI_GA (1998)
- Simulated Annealing: DAMMIN (1999), GASBOR (1999)
- Monte Carlo: saxs3d (1999)
- Monte Carlo: LORES (2005)

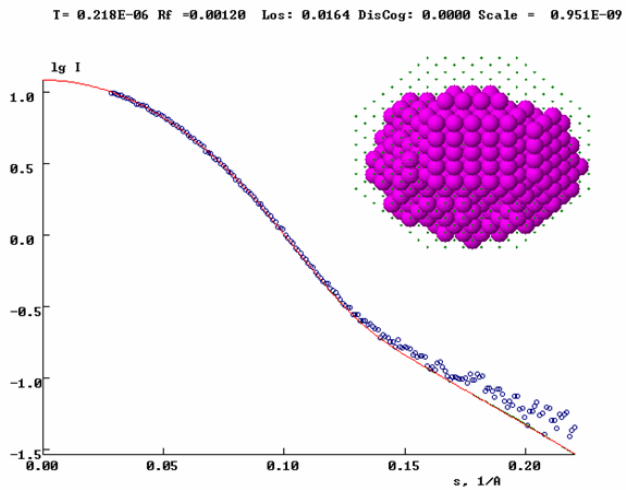
# Molecular shape reconstructions with DAMMIN

Penalty function: $f(X) = \chi^2 + \alpha P(X)$

Regularizer P(X) penalizes loose structures

$$P(X) = 1 - \langle C(N_c) \rangle$$

$$C(N_c) = 1 - [\exp(-0.5N_c) - \exp(-0.5N_c)]$$



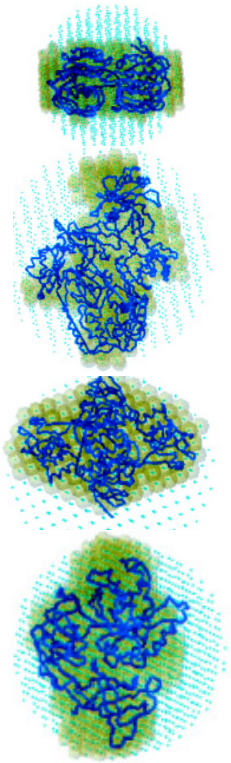

Fits *ab initio* low-resolution molecular shapes from SAXS/SANS data only.

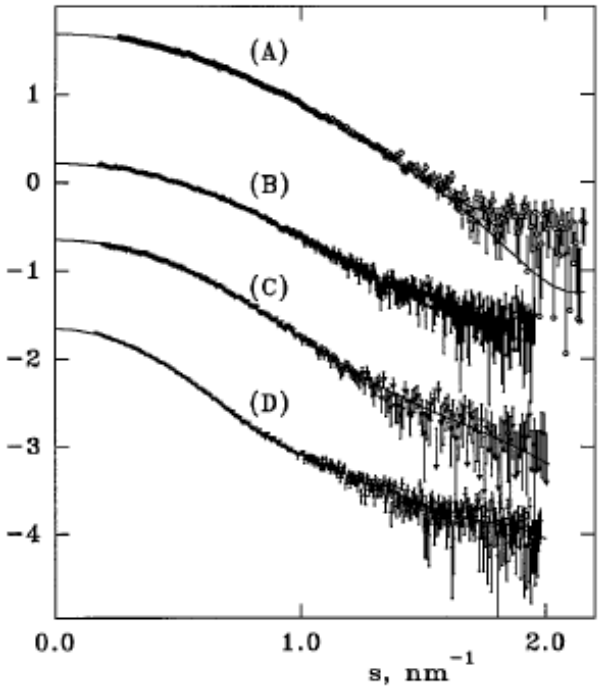
.

$q_{max}$ < ~ 0.2-0.3 $A^{-1}$

Not applicable to RNA/DNA/protein complexes, natively unfolded proteins.

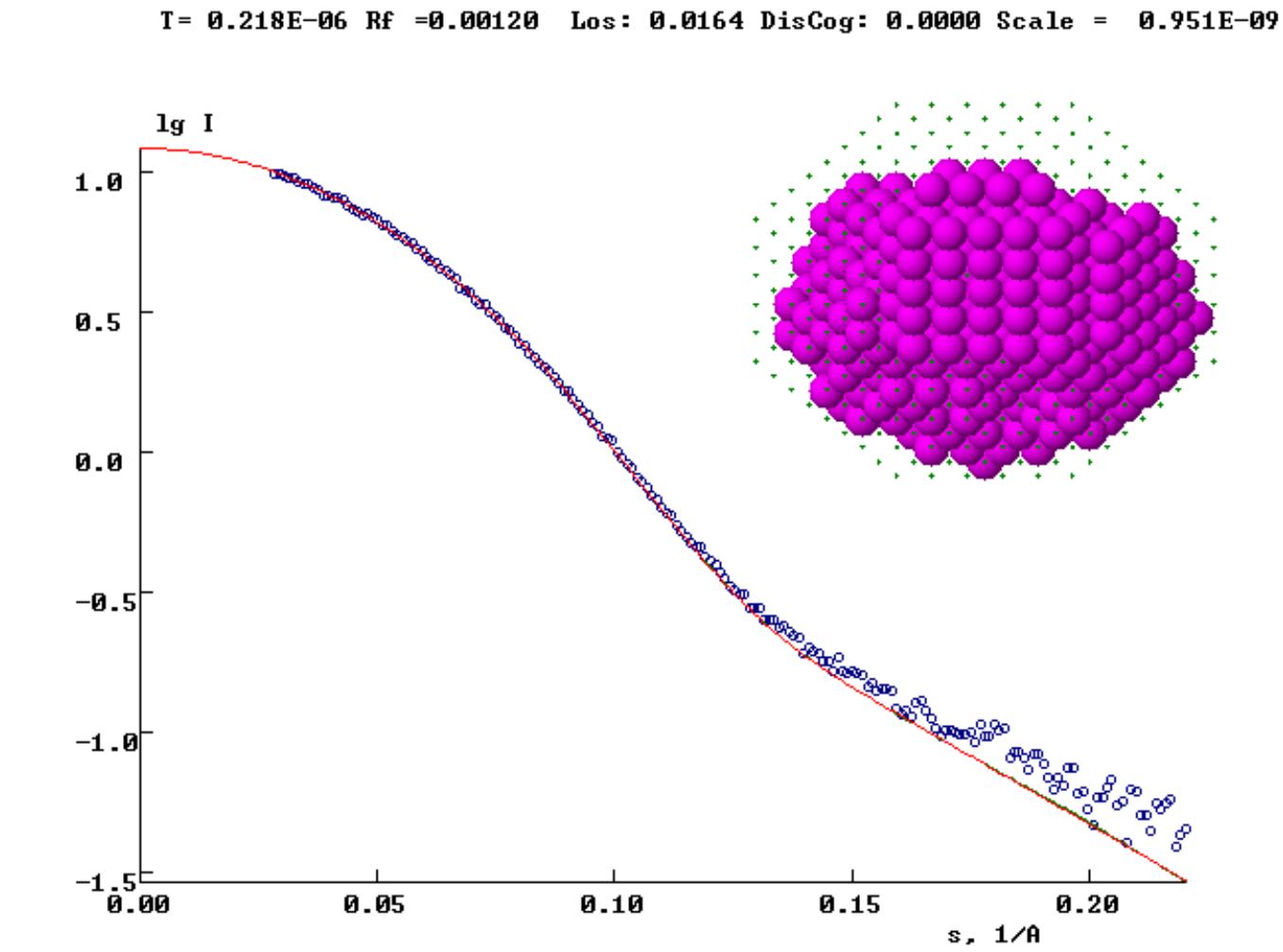
Point symmetry/particle anisometry can (*and should*) be specified when known!



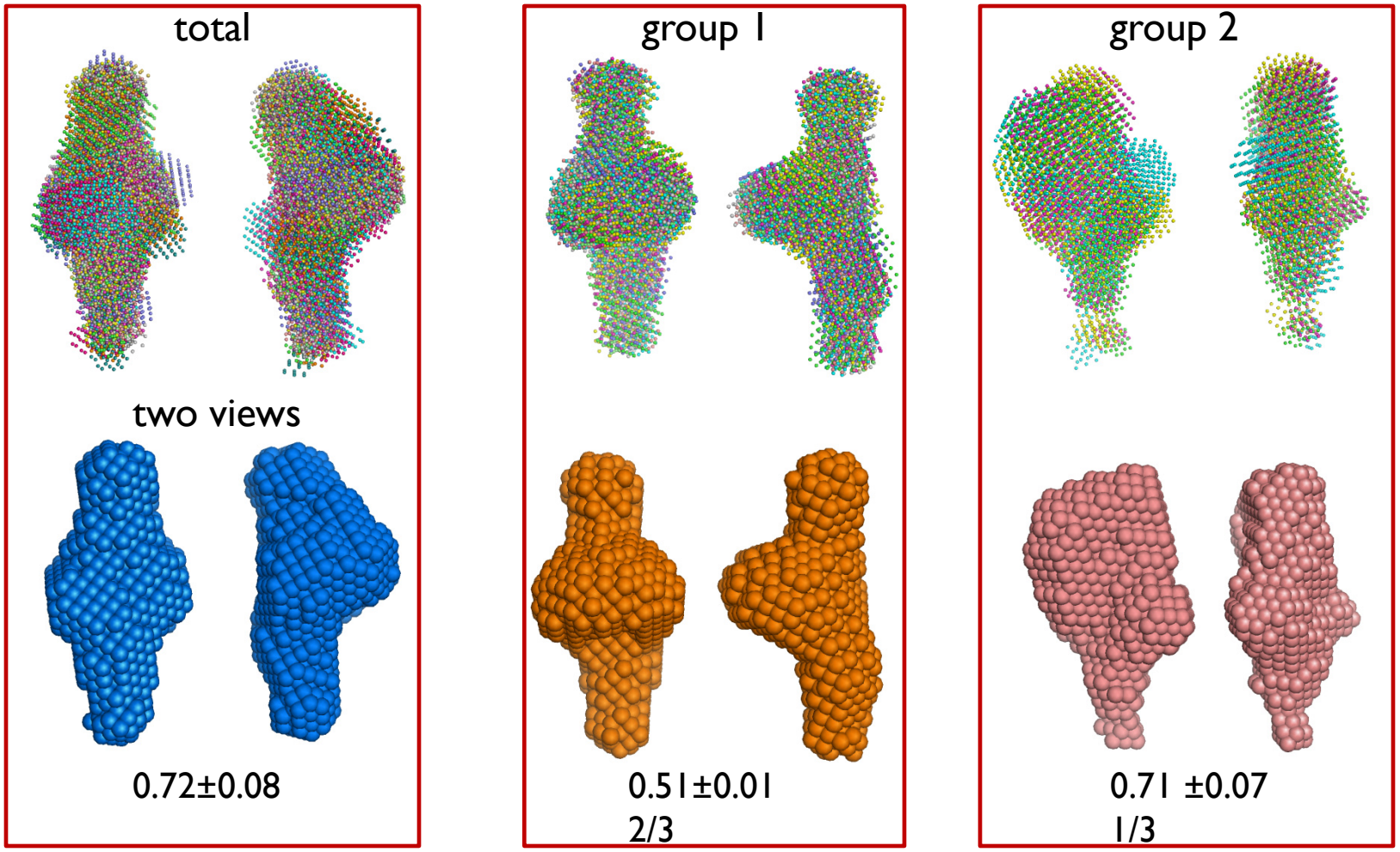



Svergun, D.et al. (1999) Biophys. J. 80, 2946-2953.

# DAMMIN example

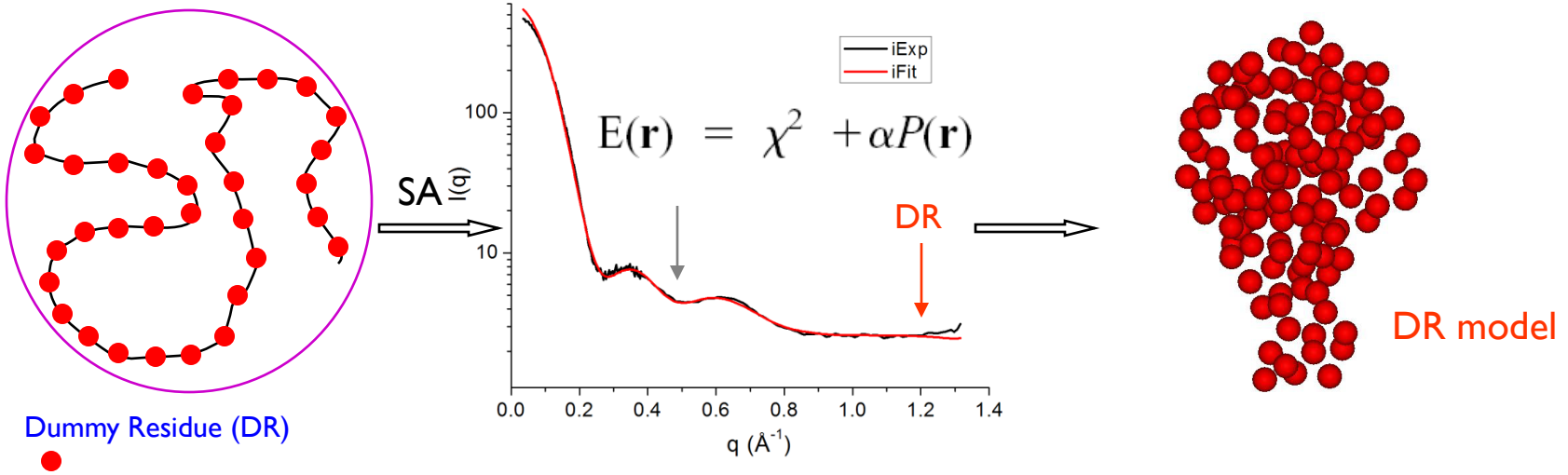


# Dammin ab initio reconstructions: multiple solutions




Solution degeneracy is quite common. Check individual solutions visually & divide into groups. Use prior information to discard those that do not agree with it.

# GASBOR: a dummy-residue approach




GASBOR uses quasi-realistic "average residue" form factor and fits to higher $q_{max}$

## Regularizer:

$$P(\mathbf{r}) = \sum_{k} [W(R_k)(N_{DR}(R_k) - \langle N(R_k)\rangle)]^2 + G(\mathbf{r})$$

$$+ [\max\{0, (|\mathbf{r}_c| - r_0)\}]^2$$

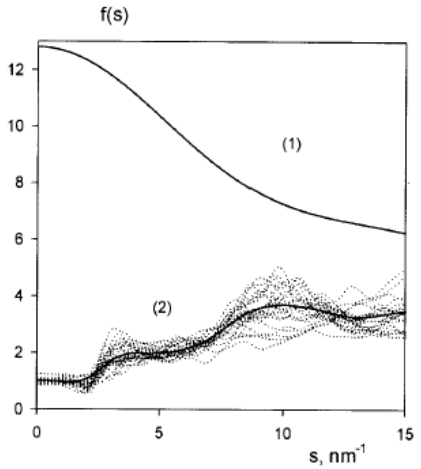



FIGURE 2 Averaged form factor of a residue (*1*) and the average correction factor (*2*). Dotted curves represent individual correction functions for the proteins in Fig. 1.

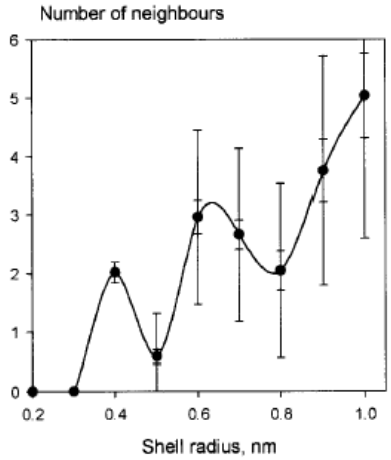



FIGURE 3 Histogram of an average number of $C_\alpha$ atoms in 0.1 nm thick spherical shells around a given $C_\alpha$ atom. Smaller error bars: variation of the averaged values over all proteins; larger error bars: averaged variation within one protein.

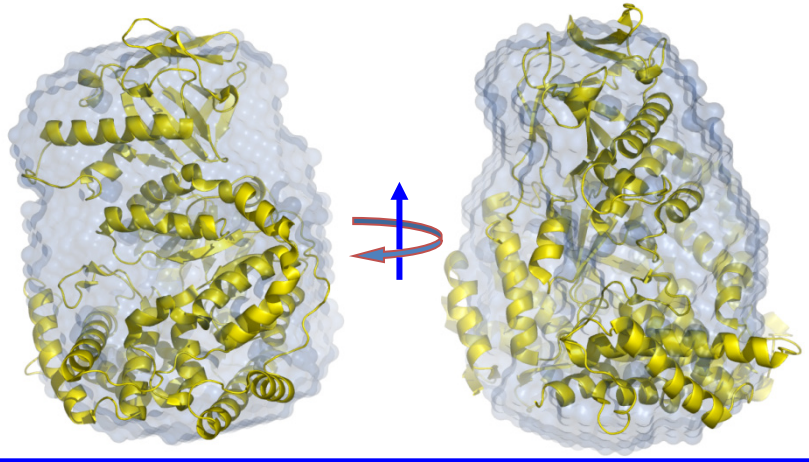
Svergun, D. et al. (1999) Biophys. J. 80, 2946-2953.

# Reconstruction accuracy – DAMMIN vs GASBOR

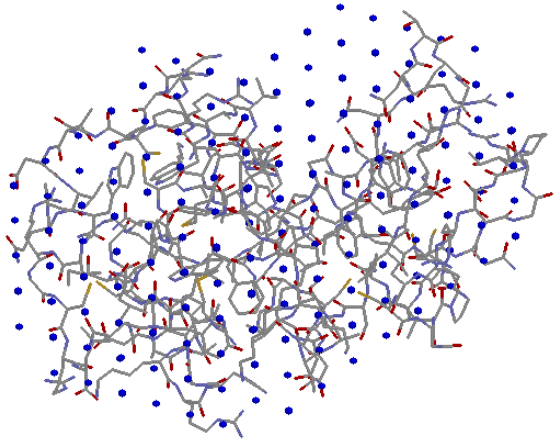
*Malate Synthase G*

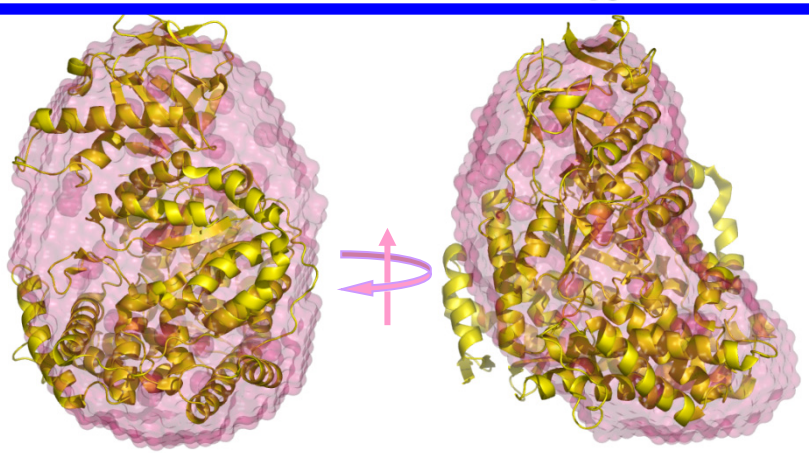
*Lysozyme*

*more accurate*



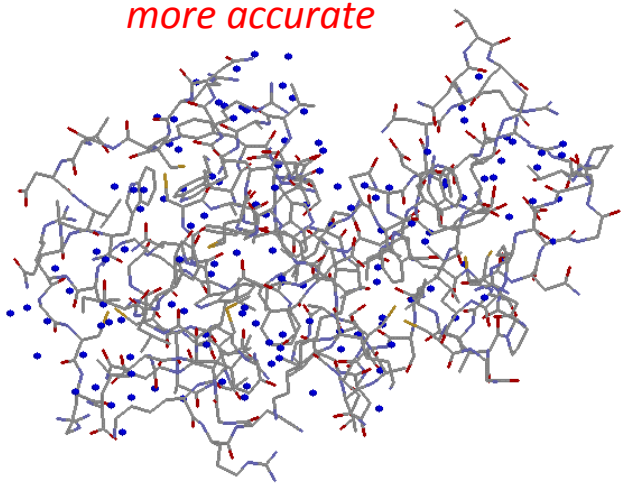
DAMMIN





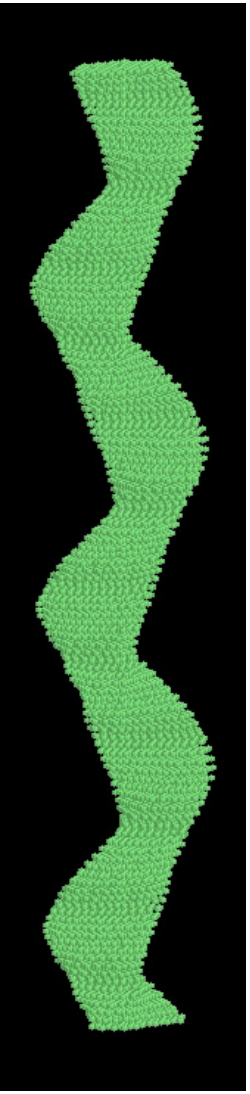
GASBOR

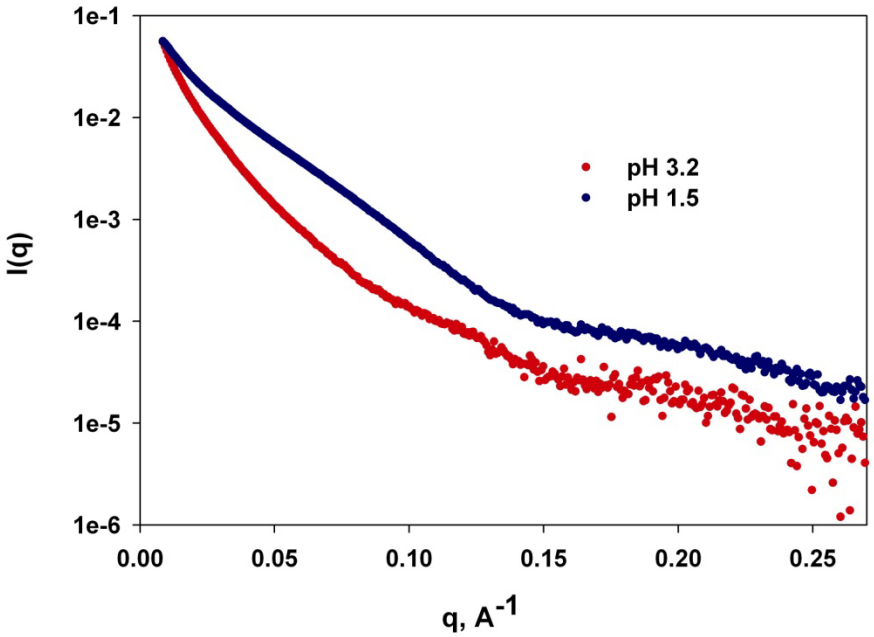
*more accurate*

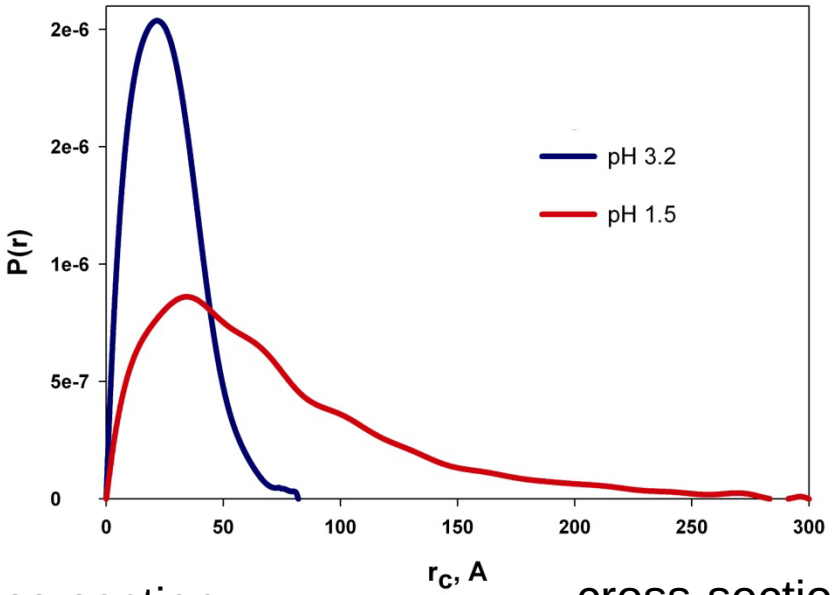


# Amyloid fibril low-resolution shape reconstructions: insulin

**pH 3.2**

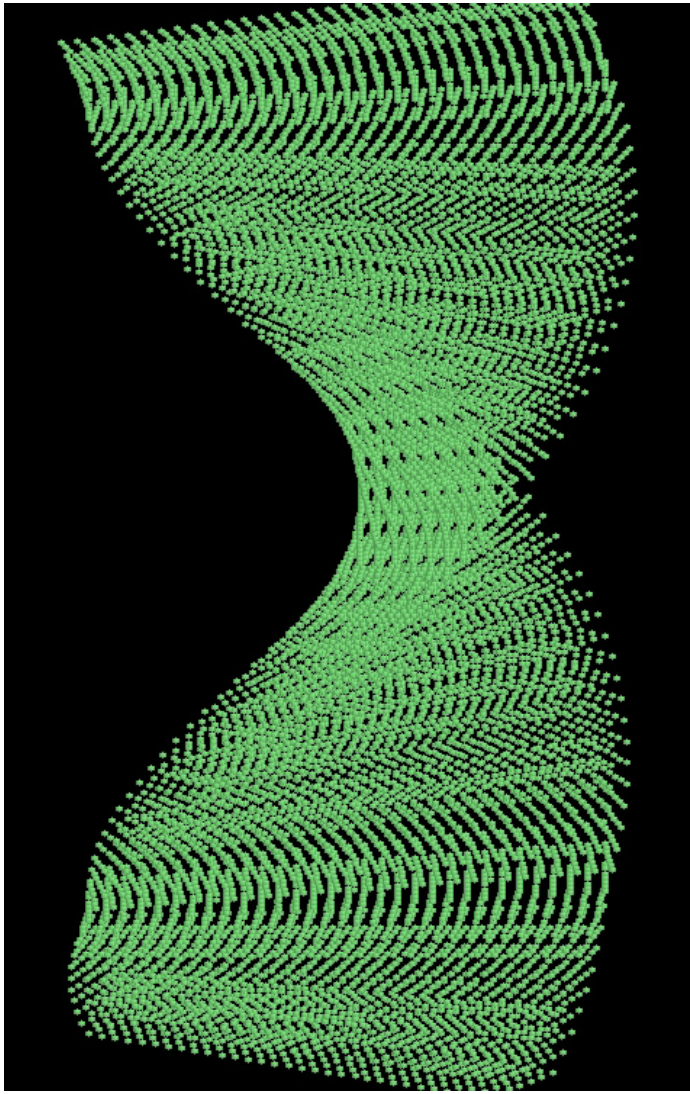





cross-section=
60 x 40 Å

**pH 1.5**



cross-section=
100 x 300 Å

# Rigid body domain positioning against solution scattering data: neuroligin/neurexin complex

Dimer for NL, 2:2 complex for NL/NX (confirmed by SAXS data), ~170 kDa

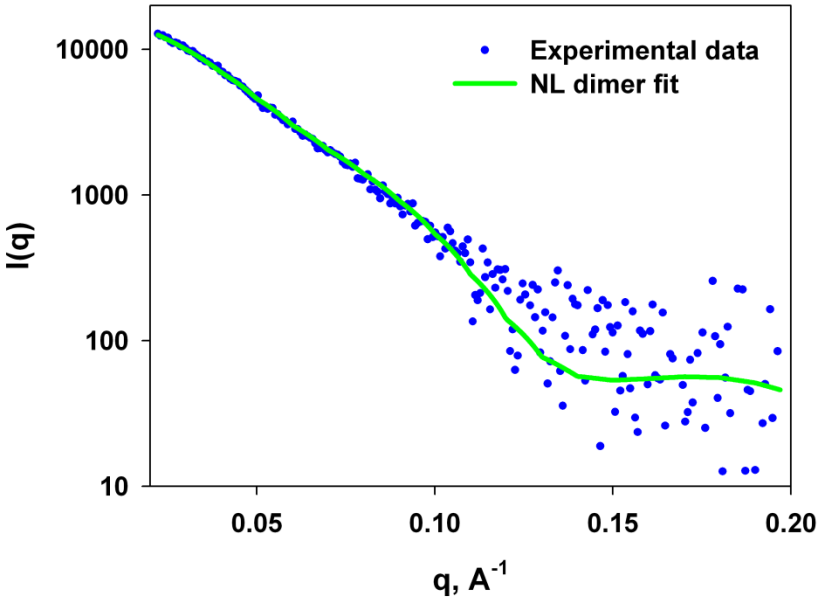
Individual domain structures were available
(3 Å accuracy for the homology model of neuroligin monomer)

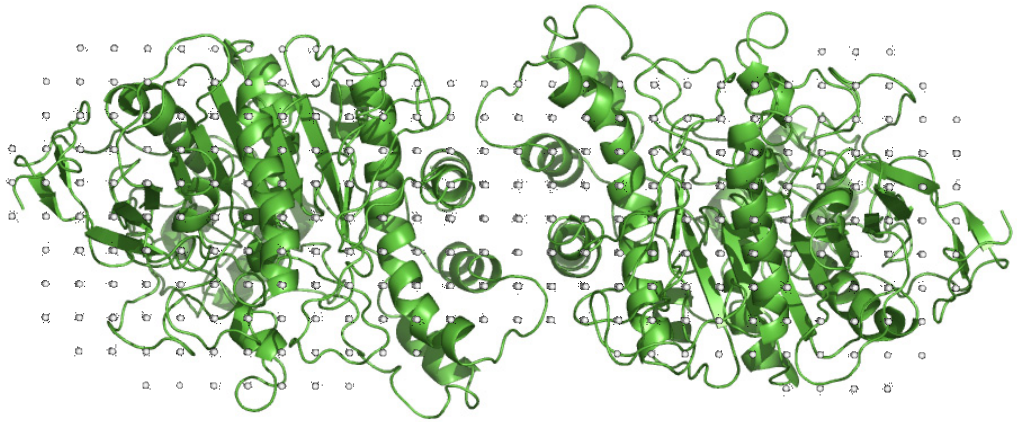
Residues involved in the NL dimer and NL/NX interfaces were established from MS H/D exchange and correlated mutation data

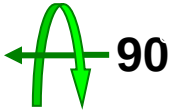
Both SAXS and SANS data were collected

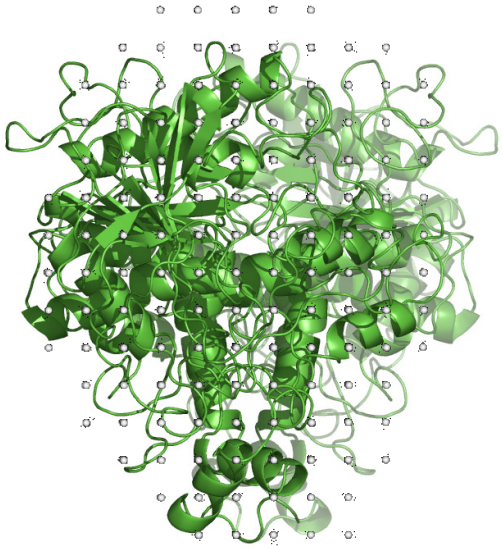
Comoletti et al. (2007). Structure 15, 693-705.

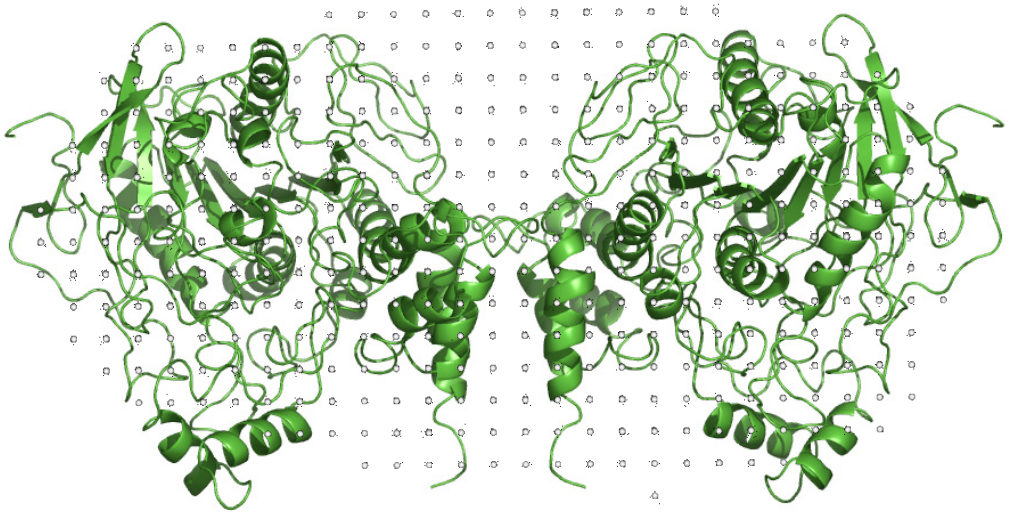
# Neuroligin dimer structure by a rigid-domain fit against the SAXS data





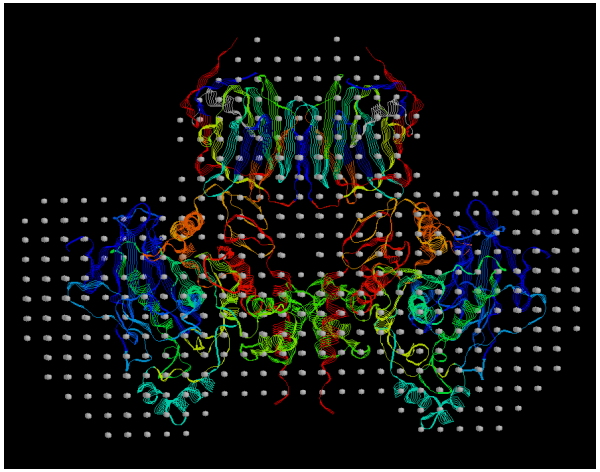
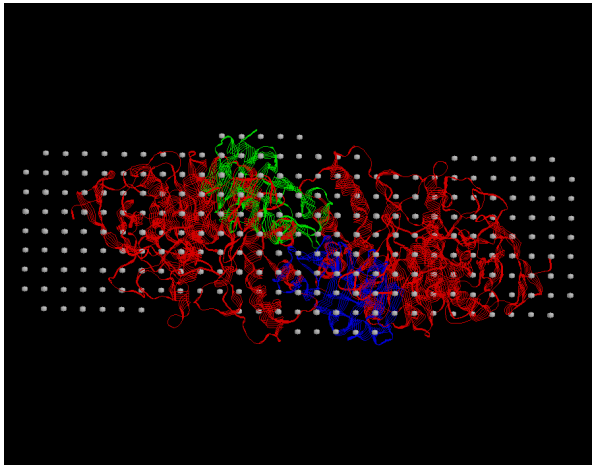
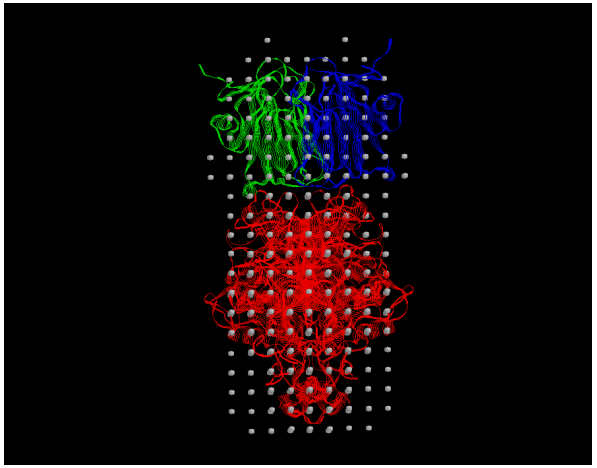



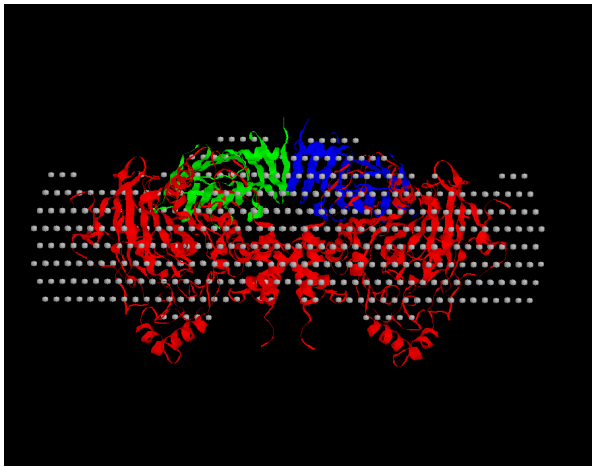
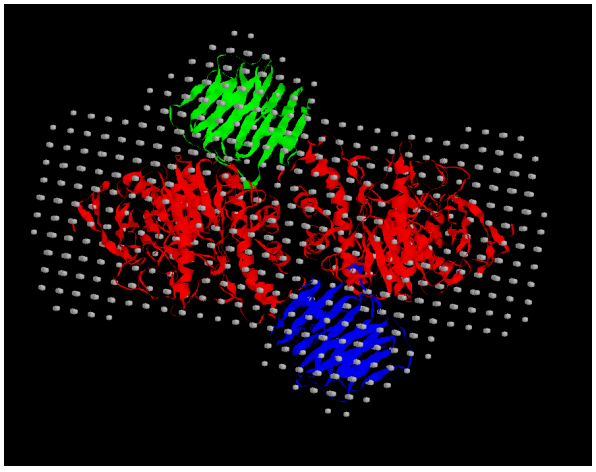
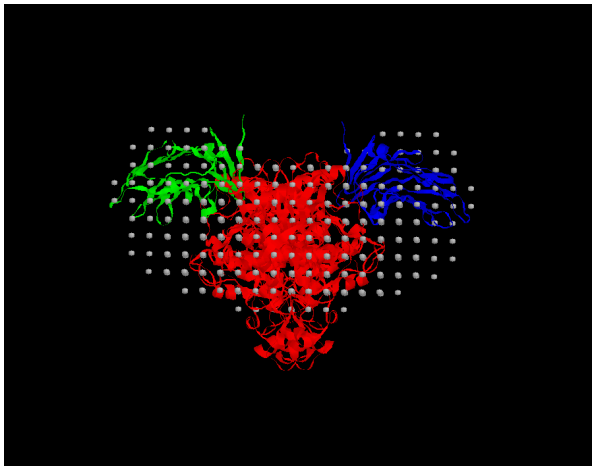
Comoletti et al. (2007). Structure 15, 693-705.

# NL/NX complex:
# shape reconstructions from the SAXS data do not yield a unique answer

## 3-lobe shape

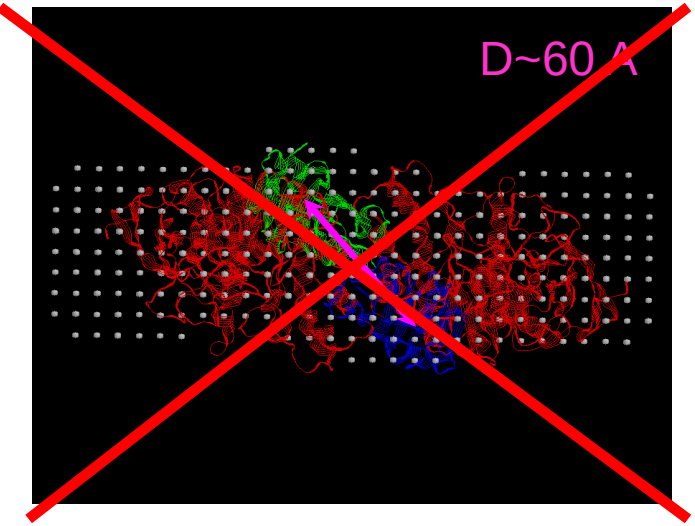




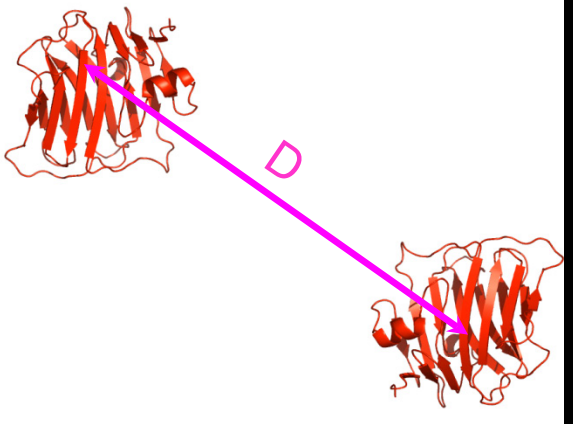
## 4-lobe shape

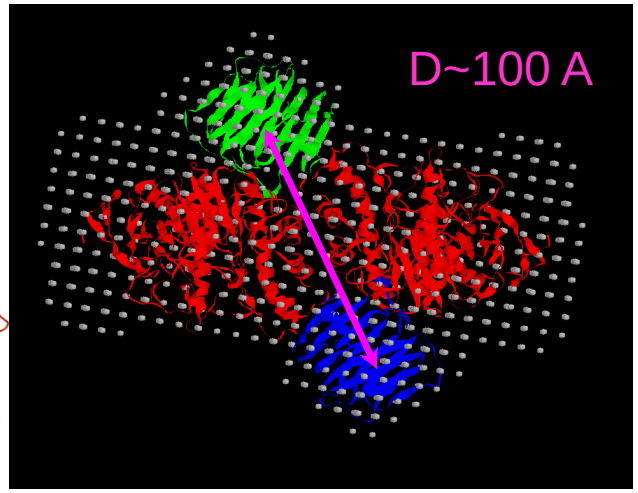




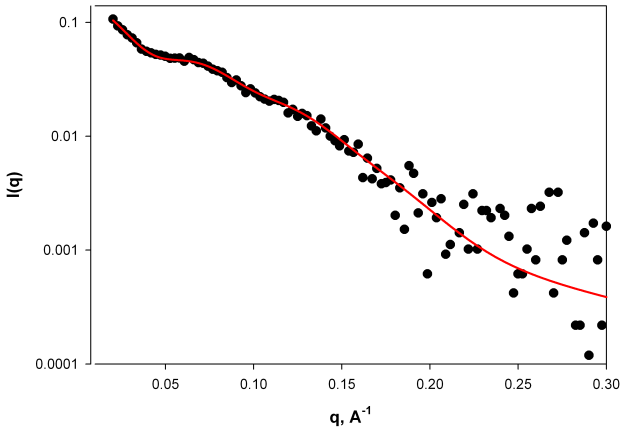
Comoletti et al. (2007). Structure 15, 693-705.

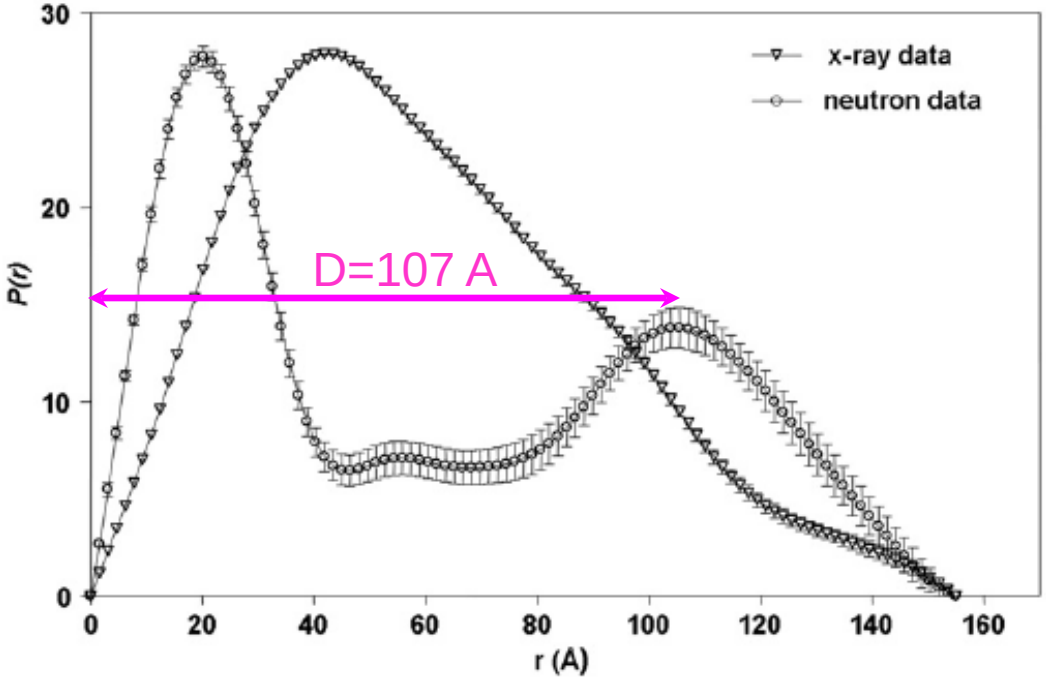
# Contrast-matched neutron scattering of $^{1}$H-NL / $^{2}$H-NX in 42% $D_2O$ allows to rule out the 3-lobe shape



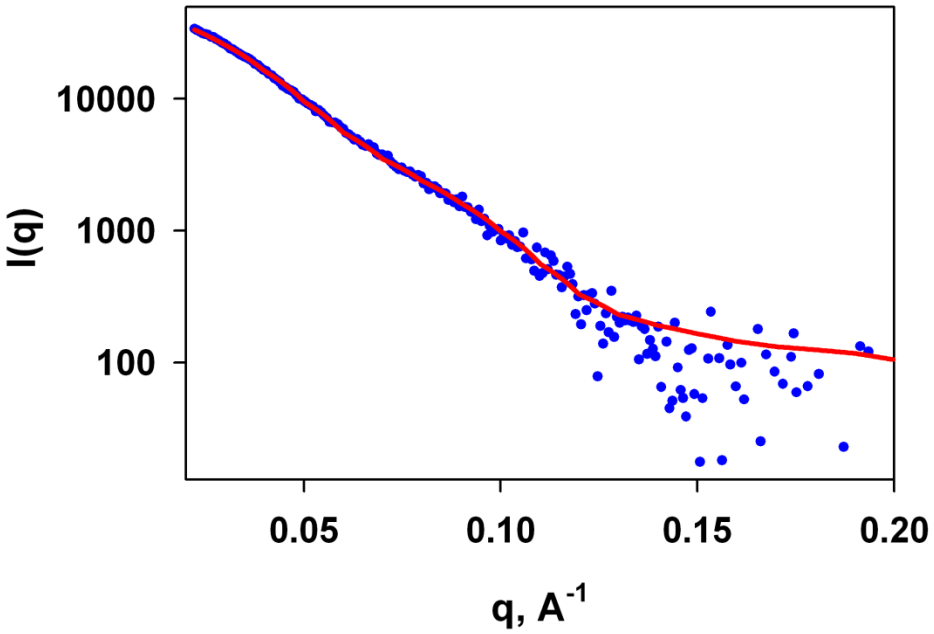



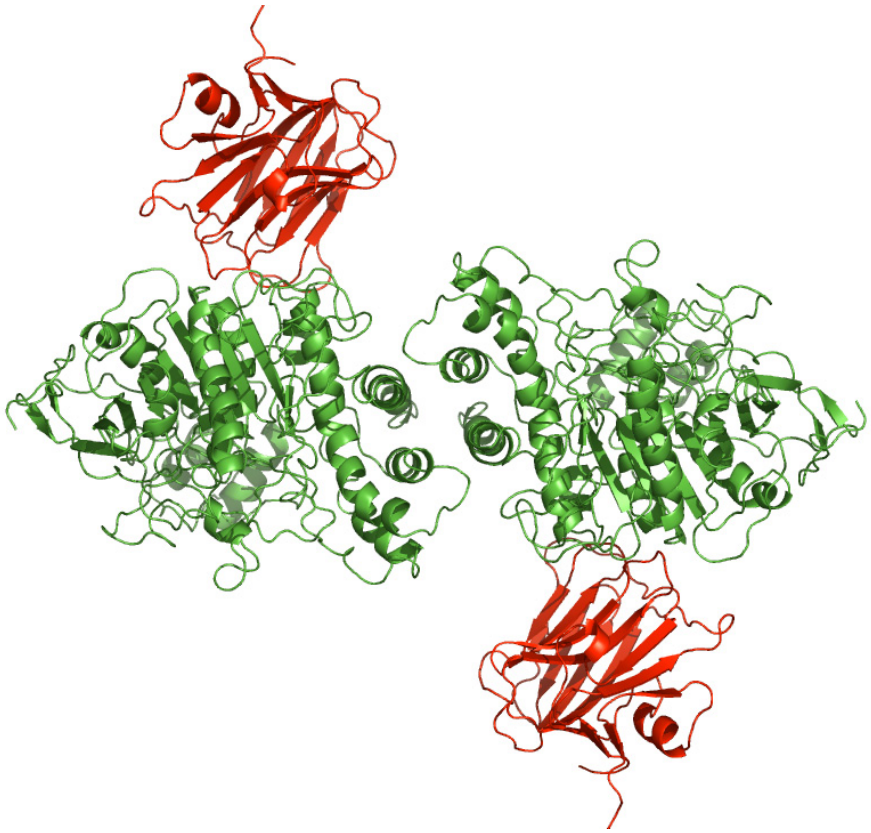




Comoletti et al. (2007). Structure 15, 693-705.

# NL/NX1 complex by a rigid-body fit against SAXS data

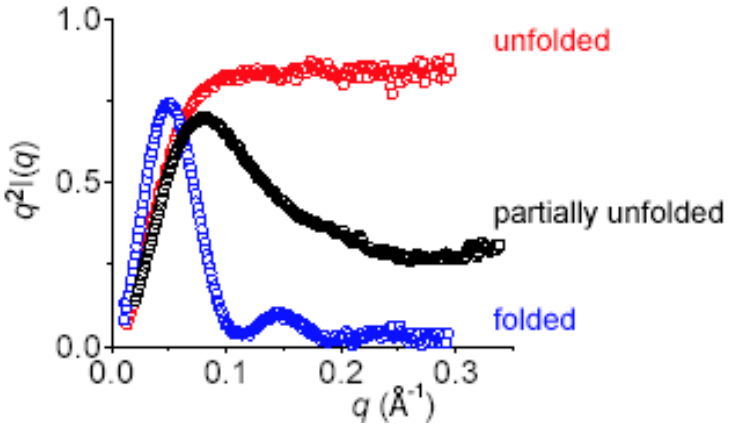


Comoletti et al. (2007). Structure 15, 693-705.

# Structural disorder/flexibility: can it be reliably detected from SAXS data?

**The Kratky plot identifies unfolded samples.** Globular macromolecules follow Porod's law and have bell-shaped curves. Extended molecules, such as unfolded peptides, lack this peak and have a plateau or are slightly increasing in the larger *q* range.




Features in Kratky plot argue against disorder.

When flexibility is present $D_{max}$ from P(r) transforms is underestimated.

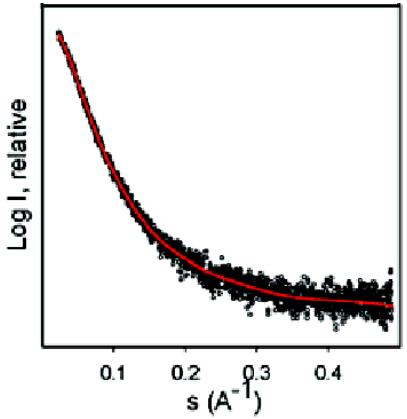
Significant polydispersity of sizes leads to extremely narrow Guinier region at very low q.

EOM is a useful tool for interpreting data affected by disorder.


Heller, W. (2004) Acta Cryst. D61, 33-44.
Bernado et al. (2007) J. Am. Chem. Soc. 129, 5656-5664.
Wang, Y et al. (2008) J. Mol. Biol. 377, 1576-1592.

# Ensemble Optimization Method (EOM): Ensemble fitting for flexible systems

N-member ensemble is reconstructed that reproduces the observed scattering data.




$$I(s) = \frac{1}{N}\sum_{n=1}^{N} I_n(s)$$

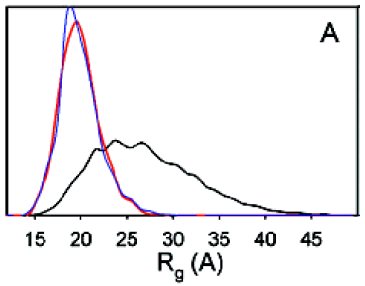
Inputs: I(q) data, protein sequence, individual rigid domain structures (if any).

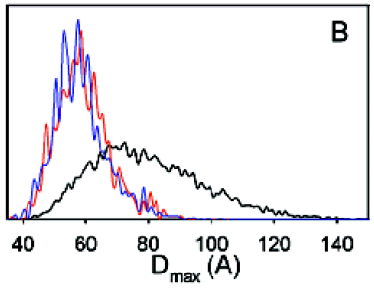
Starting pool of conformers can be either program-generated or user-specified.

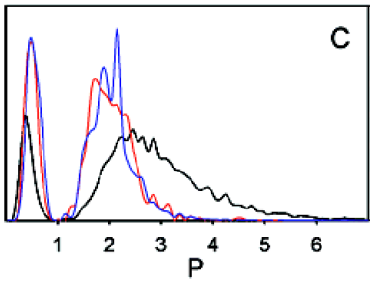
Applicable to both natively unfolded systems or rigidly held domains connected by flexible linkers.

Multiple scattering curves from deletion mutants can be fitted simultaneously.

Program outputs: best-fitting ensemble members and distributions of $R_g$, $D_{max}$, and anisotropy parameters.

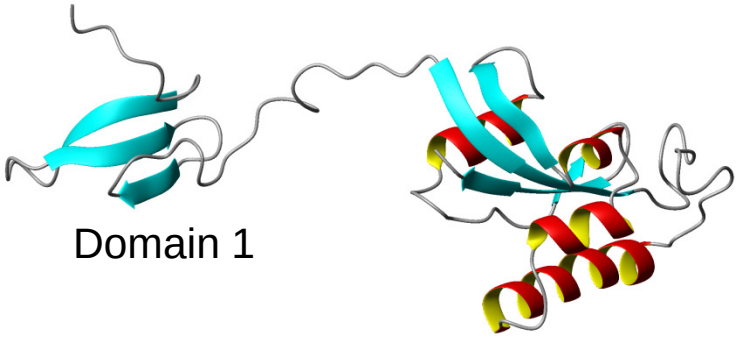


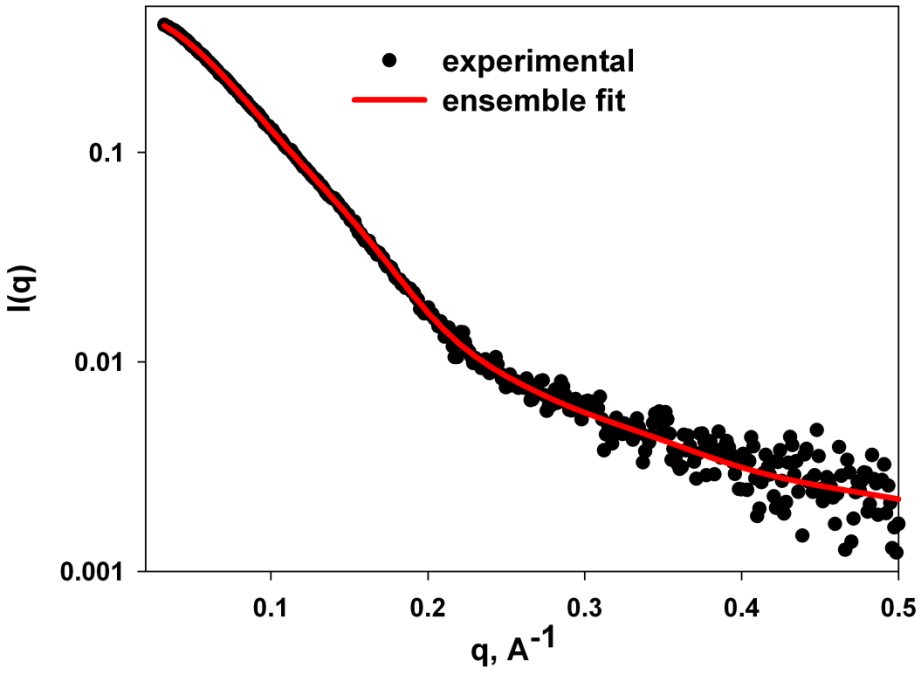

Bernado et al. (2007) J. Am. Chem. Soc. 129, 5656-5664.

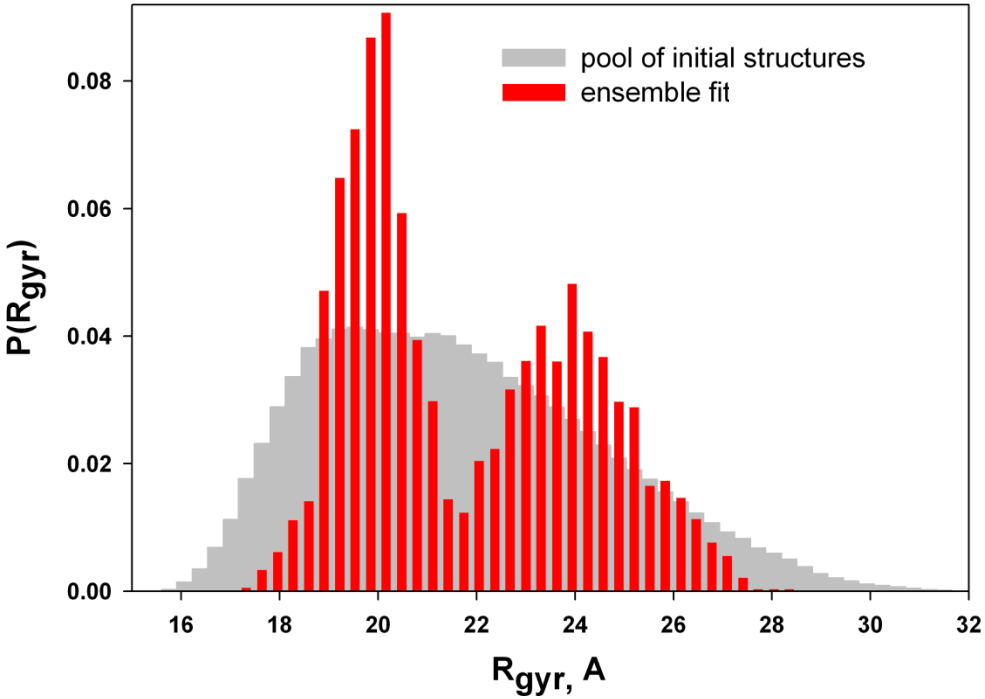
# EOM fit example




Domain 1:
$t_c$=8.8 ns, Pf1 Da= -7.9 Hz, R=0.64

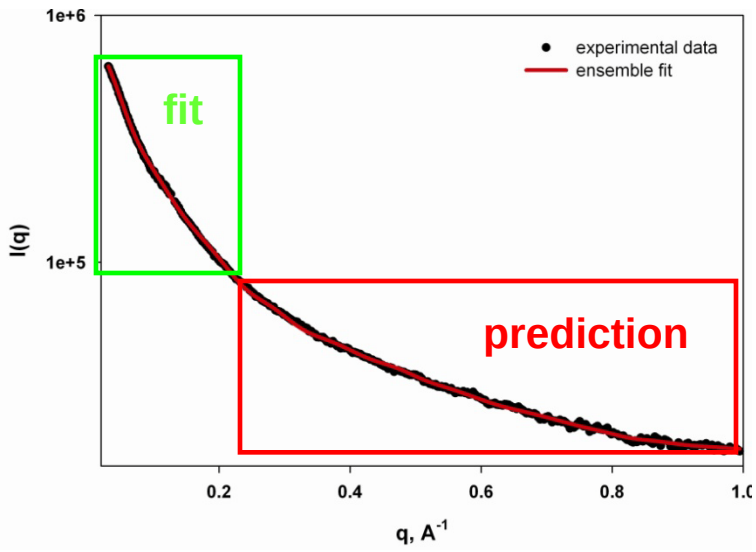
Domain 2:
$t_c$=11ns, Pf1 Da = -14.5 Hz, R=0.54



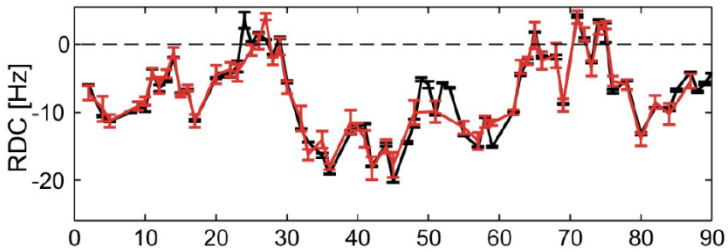



# Fitting an unfolded structural ensemble: Sic1 kinase inhibitor

## SAXS

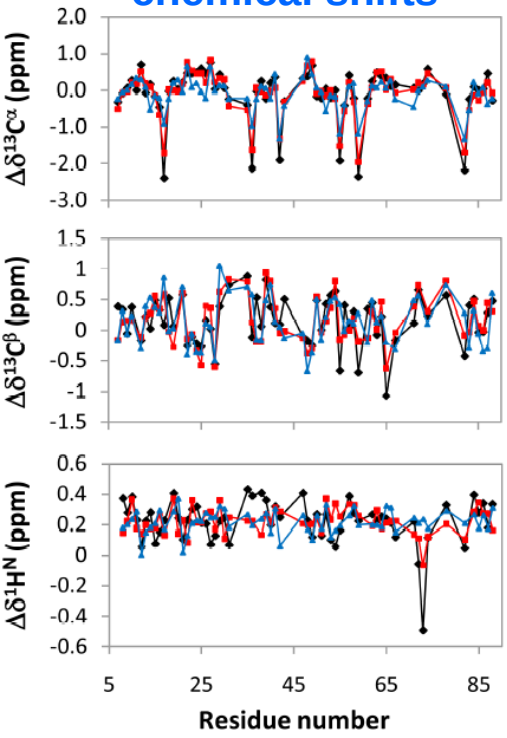



## RDCs

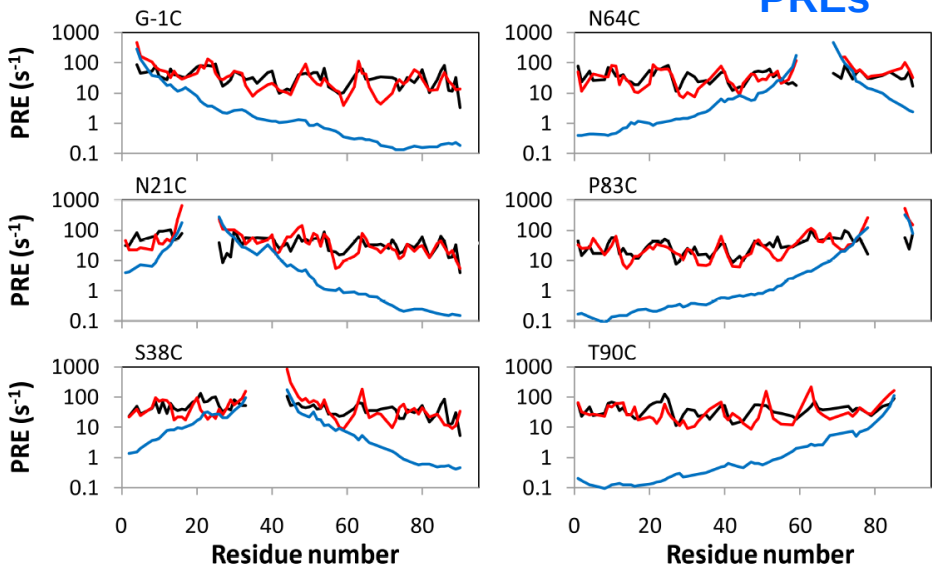



| Restraint type | # of restraints | RMSD |
|---|---|---|
| $^{13}C^{\alpha}$ chemical shifts | 66 | 0.31 ppm |
| $^{13}C^{\beta}$ chemical shifts | 62 | 0.34 ppm |
| $^{13}H^{N}$ chemical shifts | 59 | 0.15 ppm |
| PRE | 381 | 0.98 Å |
| SAXS [a] | 38 | 0.035 |
| $^{15}N$ $R_2$ | 70 | 0.61 |

## chemical shifts

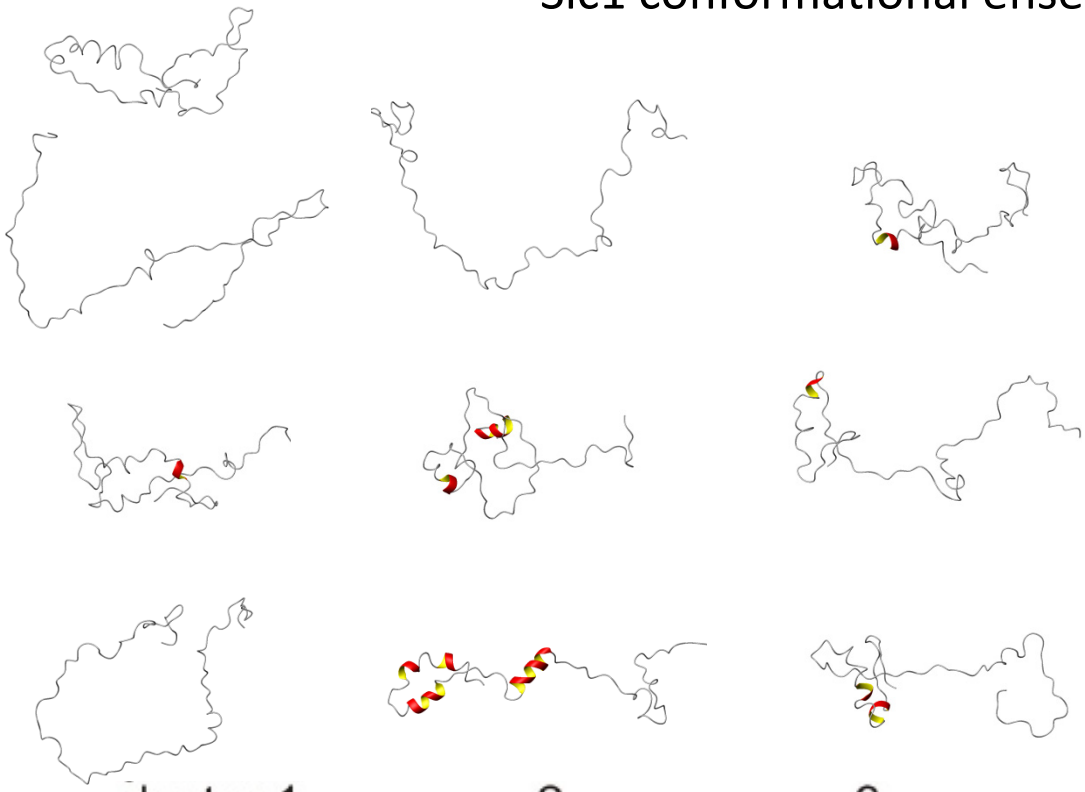



## PREs

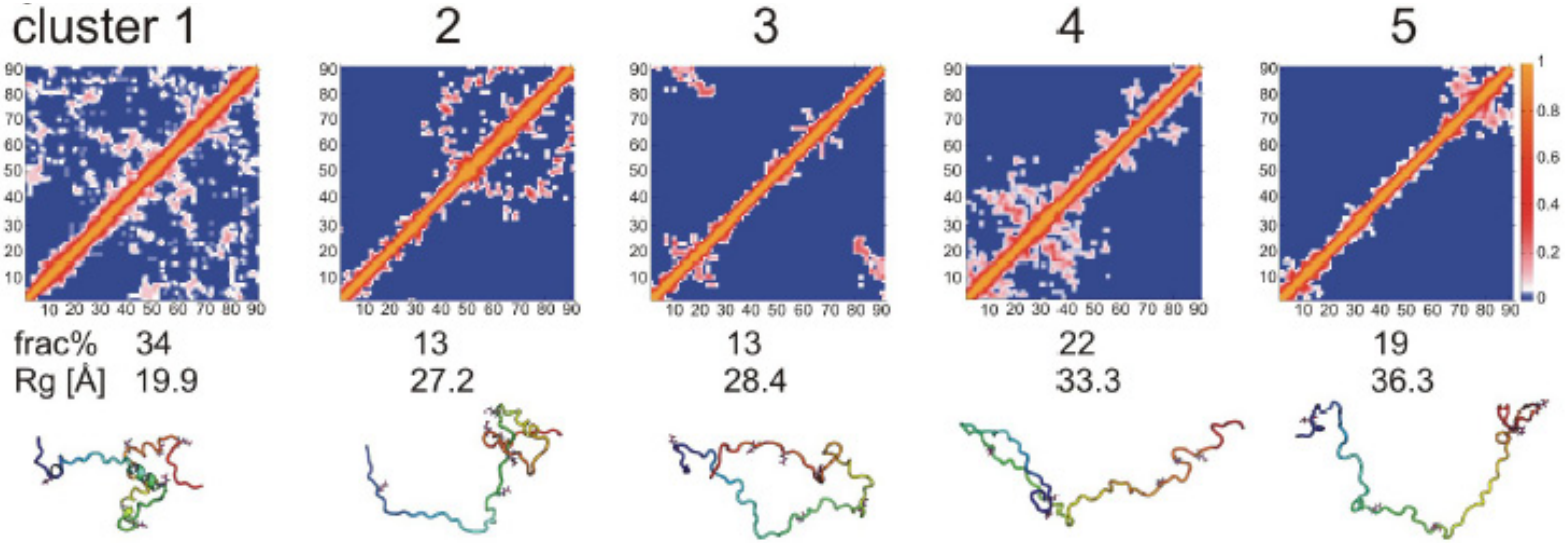



Mittag et al. (2010). Structure 18, 494-506.

# Sic1 conformational ensemble

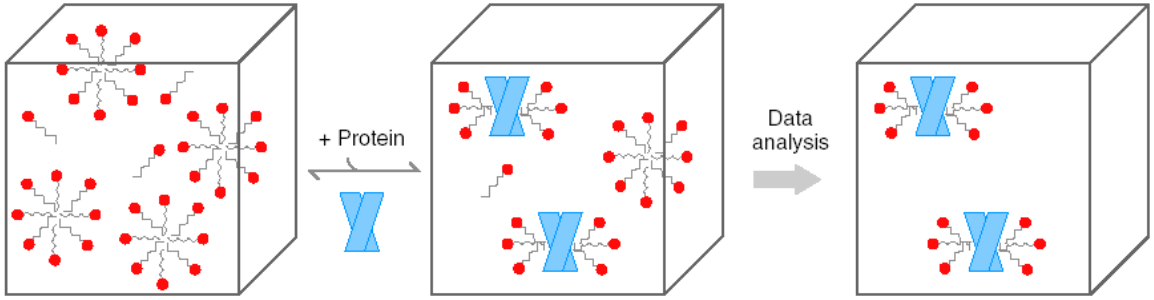


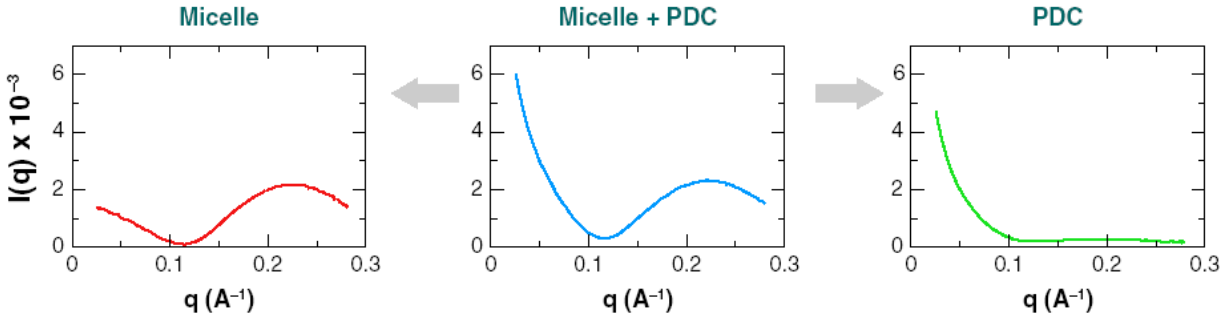
| | |
|---|---|
| Number of structures | 10.7 ± 0.5 |
| Radius of gyration (Å) | 28.6 ± 0.5 |
| Fraction in broad α-region | 0.340 ± 0.003 |
| Fraction α-helix (STRIDE) | 0.046 ± 0.004 |
| Fraction in right β-region (PPII) | 0.264 ± 0.011 |
| Fraction in left β-region (β–strand) | 0.253 ± 0.012 |

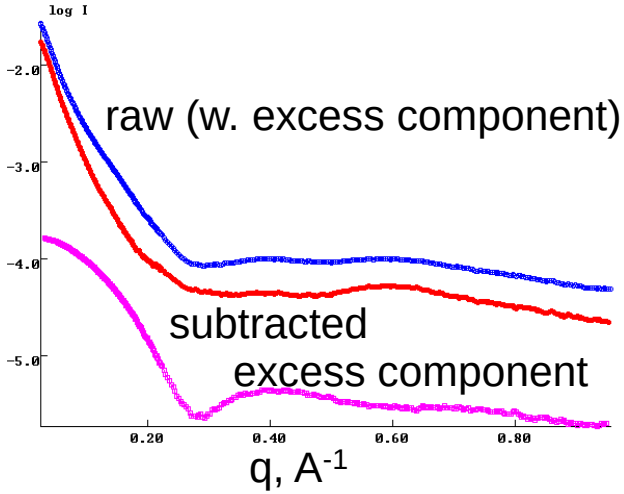



Mittag et al. (2010). Structure 18, 494-506.

# Multi-component and micelle-embedded proteins




When individual concentrations are known accurately and scattering profiles are measured separately, data can be subtracted.

For micelle-embedded systems several techniques can be used to separate the contributions:

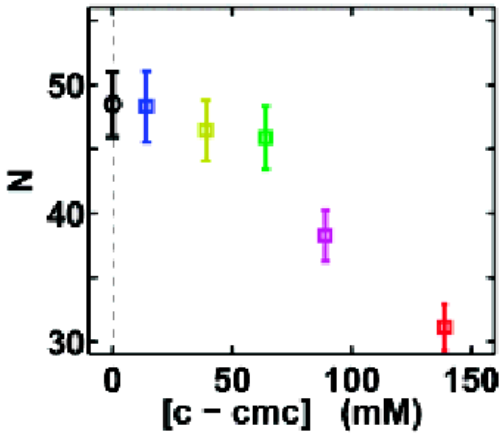
(1) contrast matching
(2) extensive dialysis against the buffer with known micelle concentration
(3) Singular value decomposition analysis ( 6-10 protein/detergent stoichiometries). The number of independent components is difficult to establish *a priori*.

Lipfert et al. (2007) Ann. Rev. Biophys. Biomol. Str. 36, 307-327.

# Detergent micelles: what can be learned from SAXS

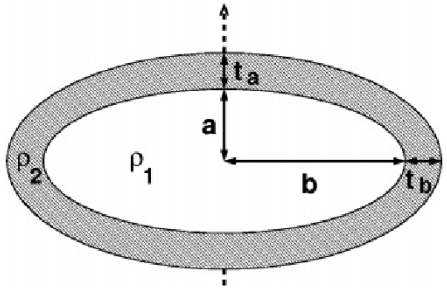
solvent: 0.334 e/Å$^3$, hydrocarbon core: 0.27 e/Å$^3$, polar head groups: 0.49 e/Å$^3$
protein: 0.42 e/Å$^3$

Aggregation number can be obtained from I(0)

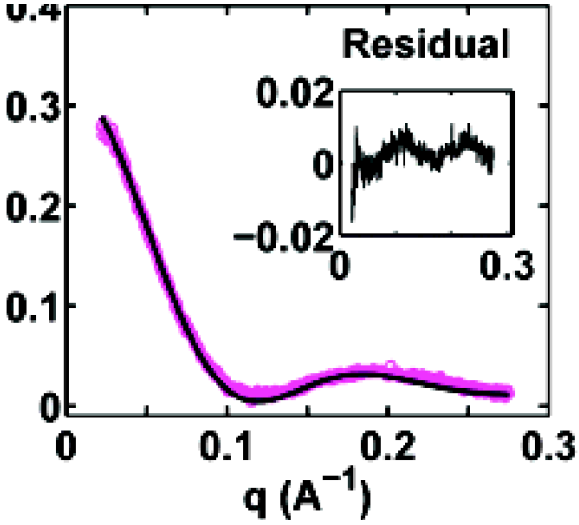



Shape can be determined from form-factor model fits:

2-shell spheroid



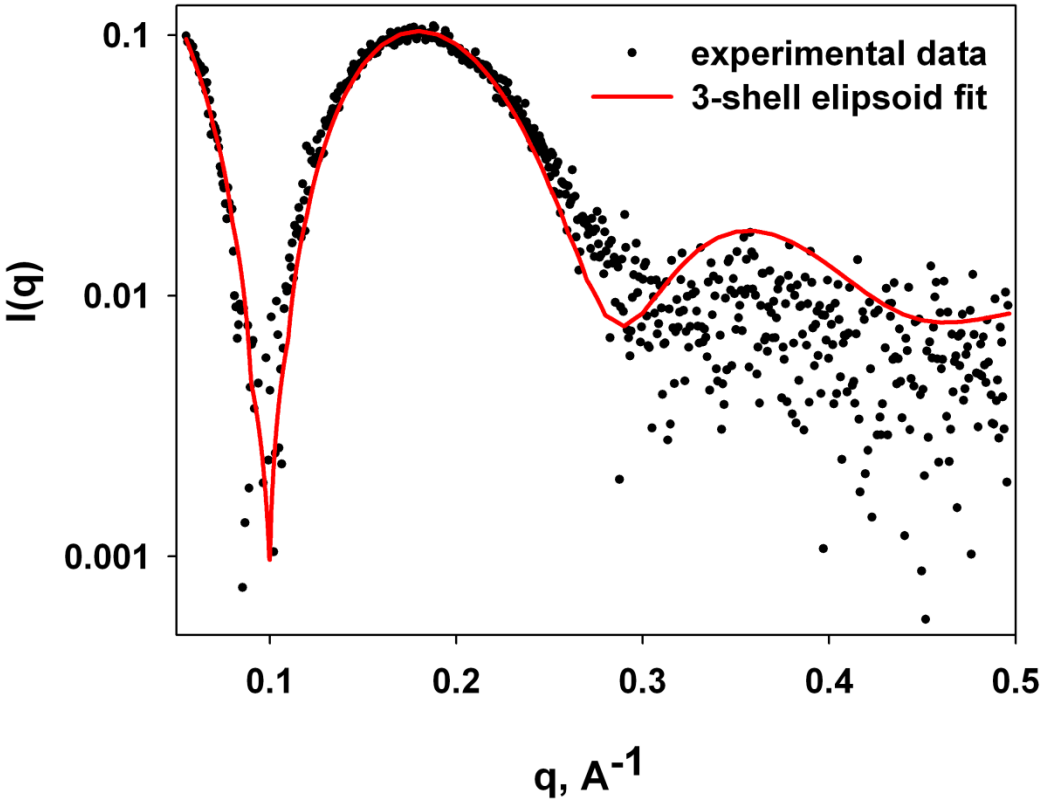

$$P(q) = \int_0^1 \left( 3V_1(\rho_1 - \rho_2)\frac{j_1(u_1)}{u_1} + 3(V_1 + V_2)(\rho_2 - \rho_s)\frac{j_1(u_2)}{u_2} \right)^2 dx$$



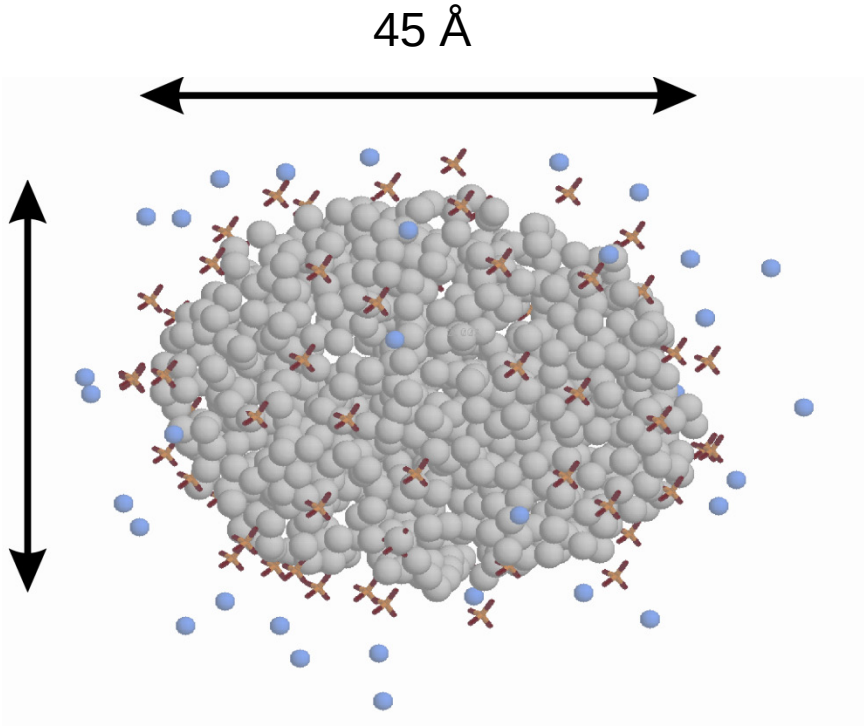

Uniqueness of the solution depends on q-range, S/N and size polydispersity

Lipfert et al. (2007) J. Phys. Chem B. 111,12427-12438.

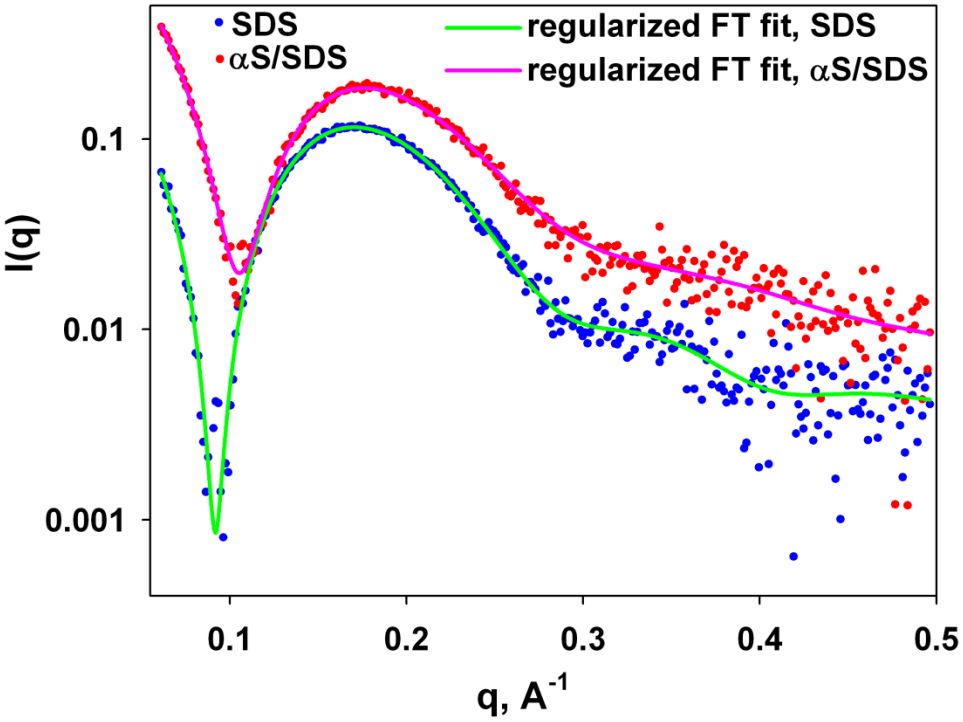
# SDS micelles shape from SAXS data



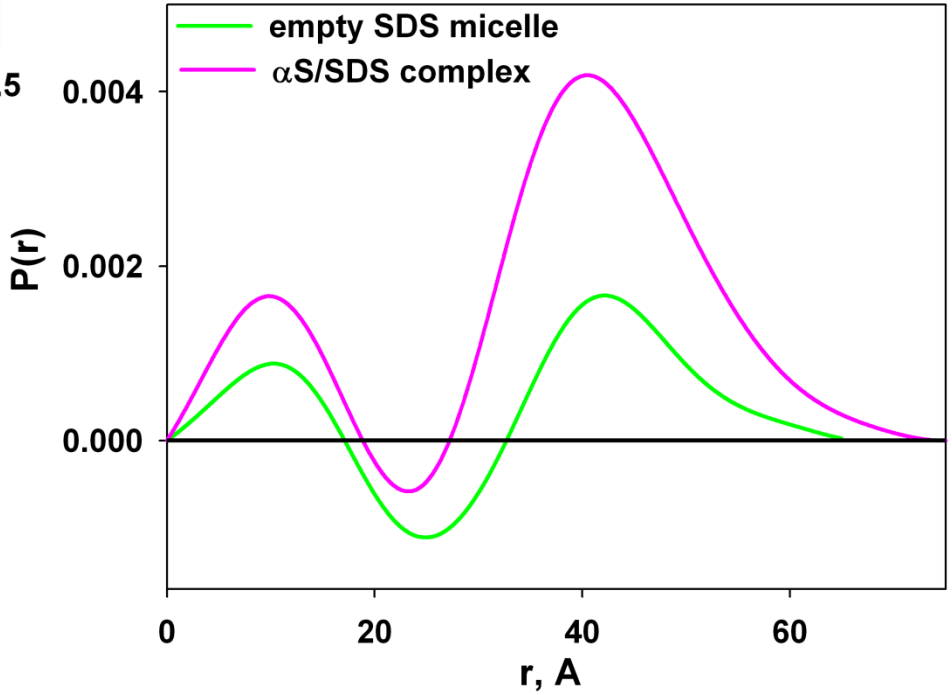

Best-fit is a prolate ellipsoid or a spherocylinder with $D_{max}/D_{min}$=1.33



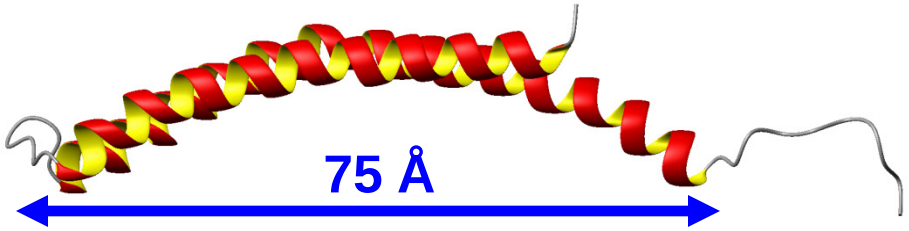
# α synuclein in SDS micelles from SAXS data

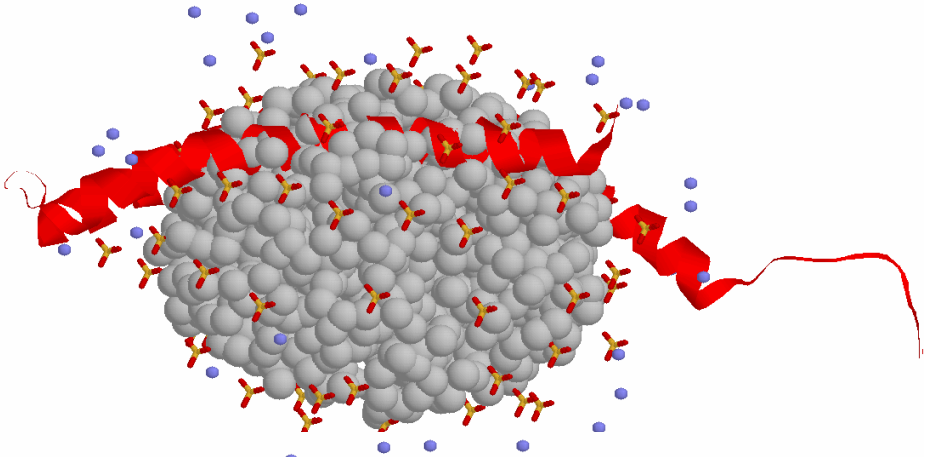


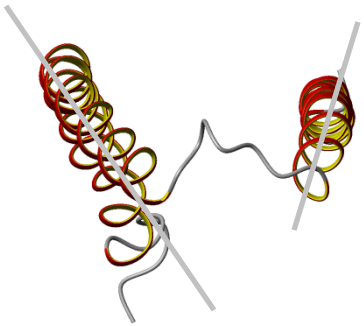

for SDS micelle, $d_{max}$ = 65 Å
for αS/SDS system, $d_{max}$ = 75 Å

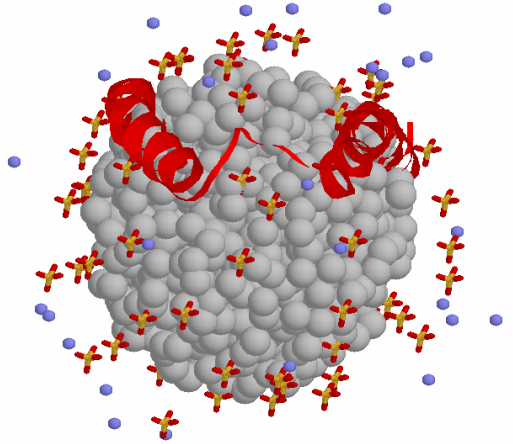


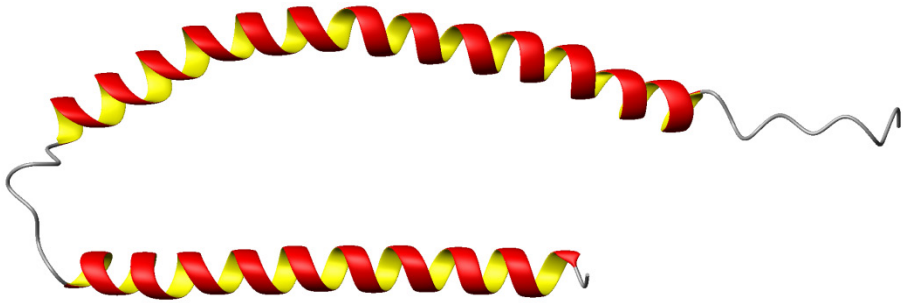
# α synuclein in SDS micelles from SAXS data

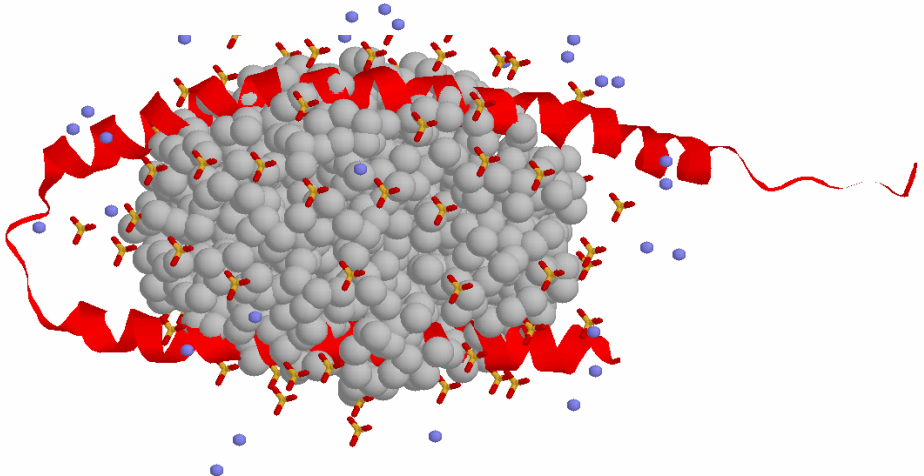




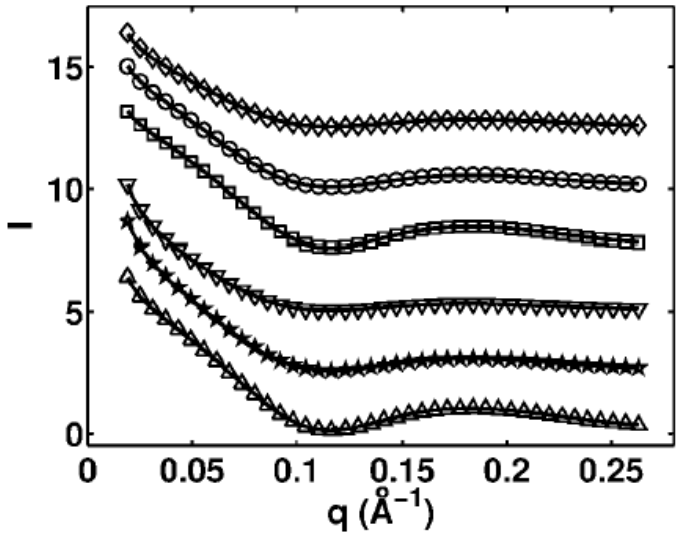






Ulmer et al. (2005) J. Biol. Chem. 280, 9595-9603.

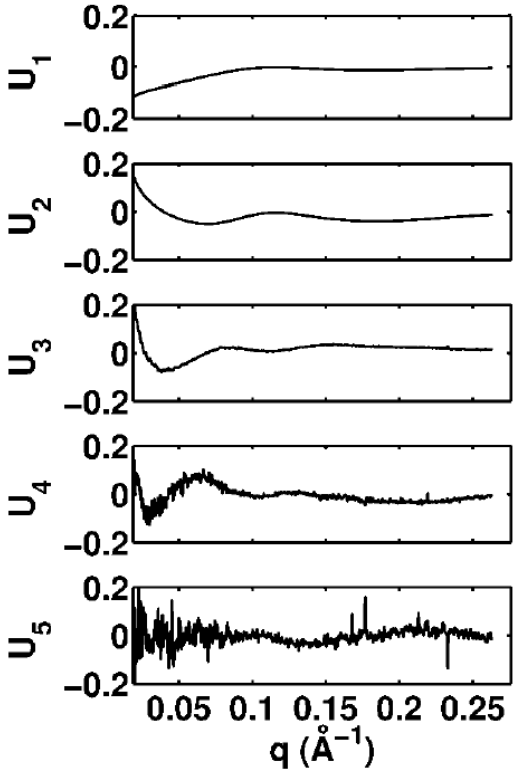
# SVD analysis of a micelle-embedded membrane protein



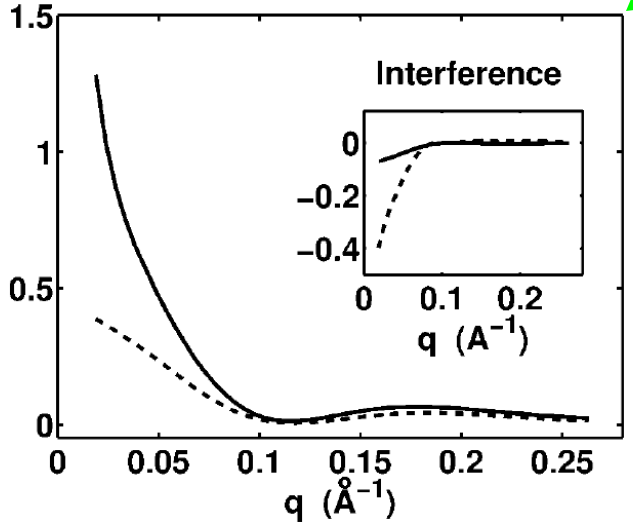

SVD

signal

noise




**model-based fit:**




$$\hat{I}(q,c) \simeq c_{\mathrm{PDC}} I_{\mathrm{PDC}}(q) + (c_{\mathrm{PDC}})^2 I_{\mathrm{int,PDC-PDC}}(q) + c'_{\mathrm{mic}} I_{\mathrm{mic}}(q) + (c'_{\mathrm{mic}})^2 I_{\mathrm{int,mic-mic}}(q) + c'_{\mathrm{mic}} c_{\mathrm{PDC}} I_{\mathrm{int,mic-PDC}}.$$

$$I_{\mathrm{mic}}(q) = \sum_{i=1}^{L} w_i b_i^{\mathrm{mic}} U_i(q) \qquad I_{\mathrm{int,mic-mic}}(q) = \sum_{i=1}^{L} w_i b_i^{\mathrm{int,mic-mic}} U_i(q)$$

$$I_{\mathrm{PDC}}(q) = \sum_{i=1}^{L} w_i b_i^{\mathrm{PDC}} U_i(q) \qquad I_{\mathrm{int,prot}}(q) = \sum_{i=1}^{L} w_i b_i^{\mathrm{int,mic-PDC}} U_i(q)$$

Lipfert et al. (2007) J. Appl. Cryst. 40, s235-s239.

# Hybrid RDC-SAXS approaches

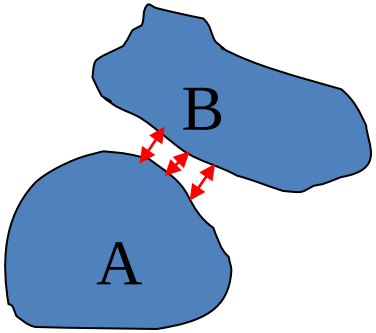
Rigid body multi-subunit positioning methods

*Ab initio* low-resolution shape/RDC combination methods

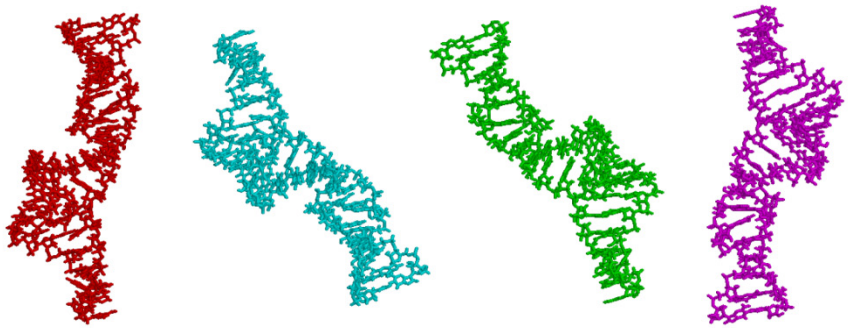
Full high-resolution structure refinement against RDC and SAXS data

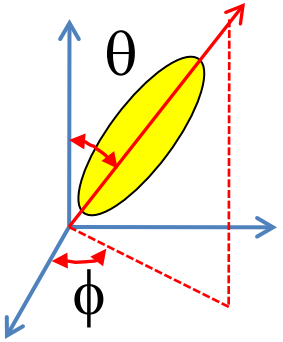
# RCD/SAXS rigid body positioning: why do it?

- Inter-subunit (often sidechain/sidechain) NOEs can be few and hard to assign correctly and unambiguously.

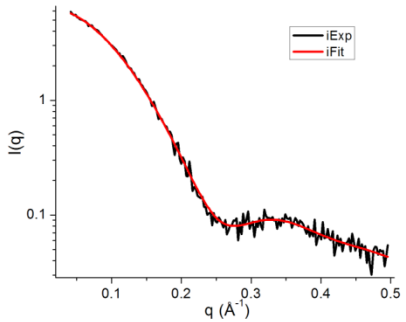


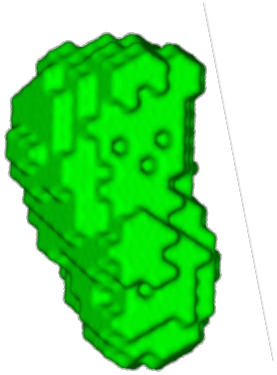

- RDCs by themselves yield a 4-fold orientational uncertainty and are completely insensitive to the translational positions.



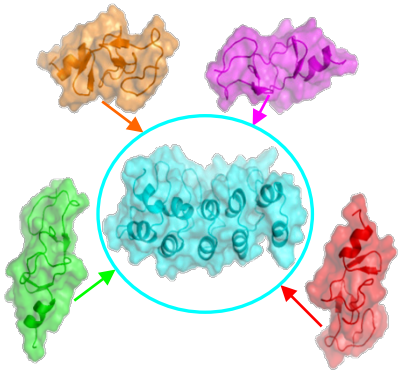



- SAXS data and shapes by themselves can be hard to interpret with certainty (less so for RNA) and can be insensitive to domain rotations.

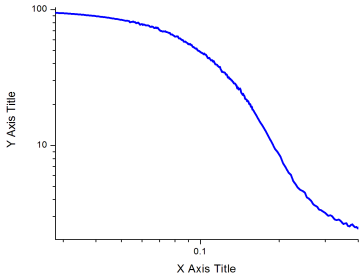


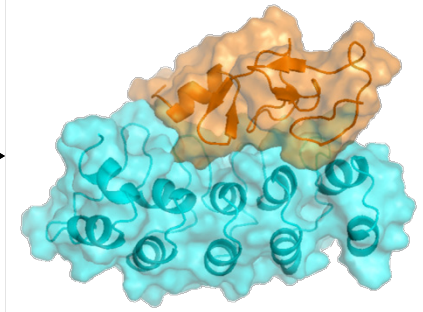
# The main idea of rigid-body refinement against RDC and SAXS data



RDC degeneracy

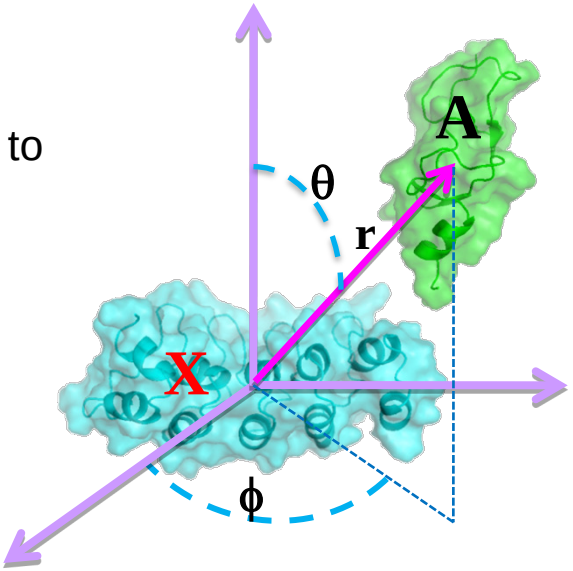


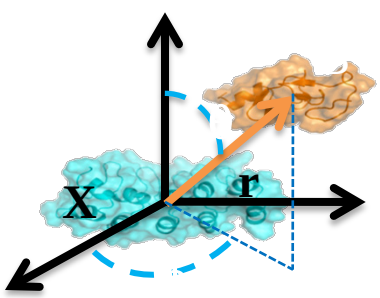

Shape degeneracy

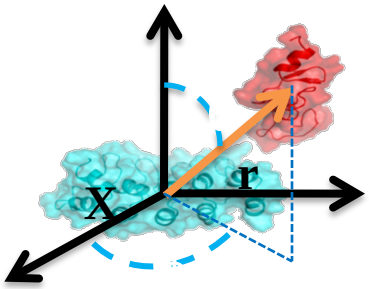


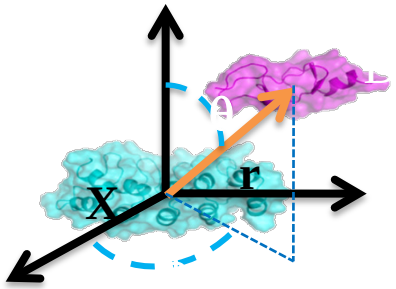
# Translational search on a grid

- Known sub-units structures are moved as rigid bodies.
- Position and orientation of subunit X are fixed at the center.
- Four possible discrete orientations of subunit A are subjected to the translational search.
- Each grid point is filtered by SAXS fit and $R_g$, $D_{max}$ (in GASR approach).

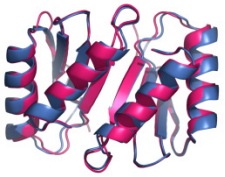


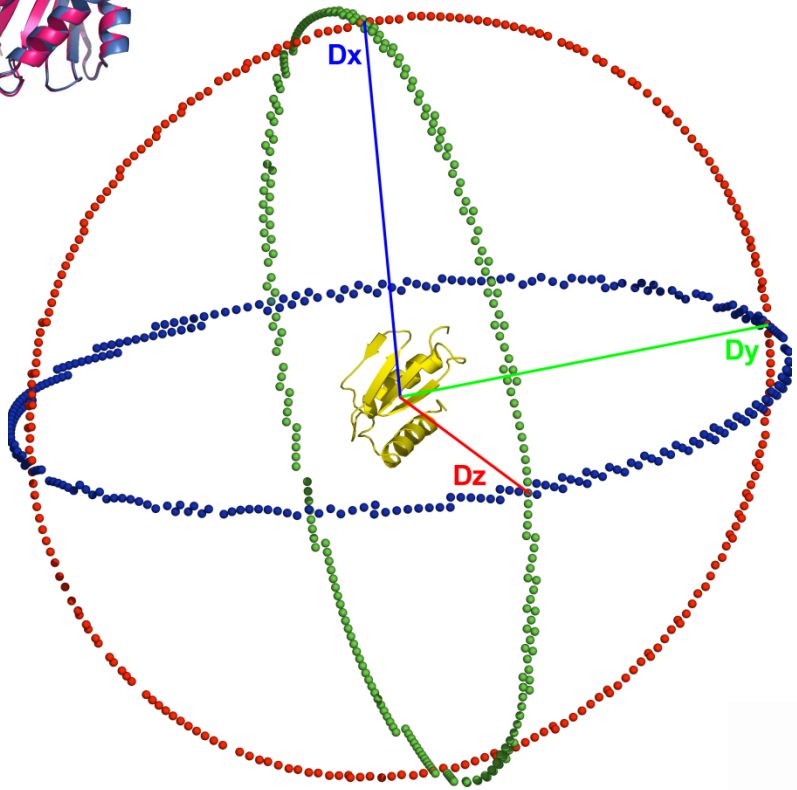



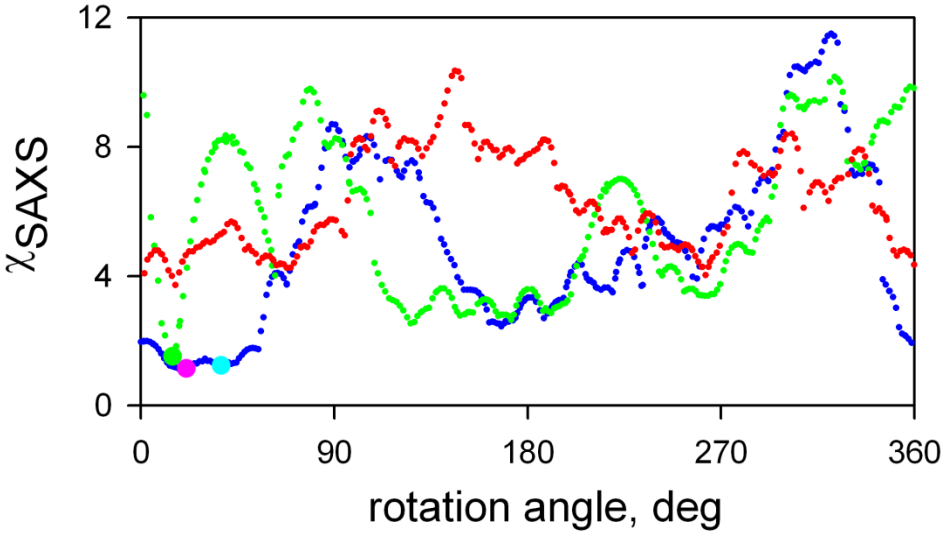



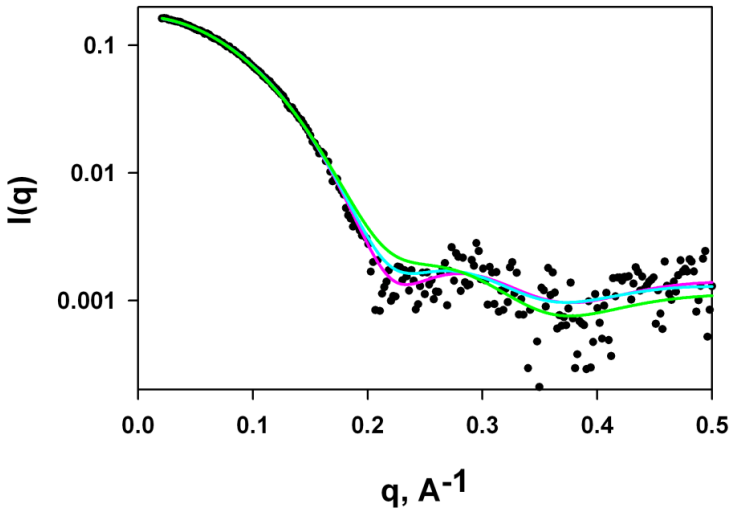

Wang, et al, (2009) JACS, 131,10507-10515.

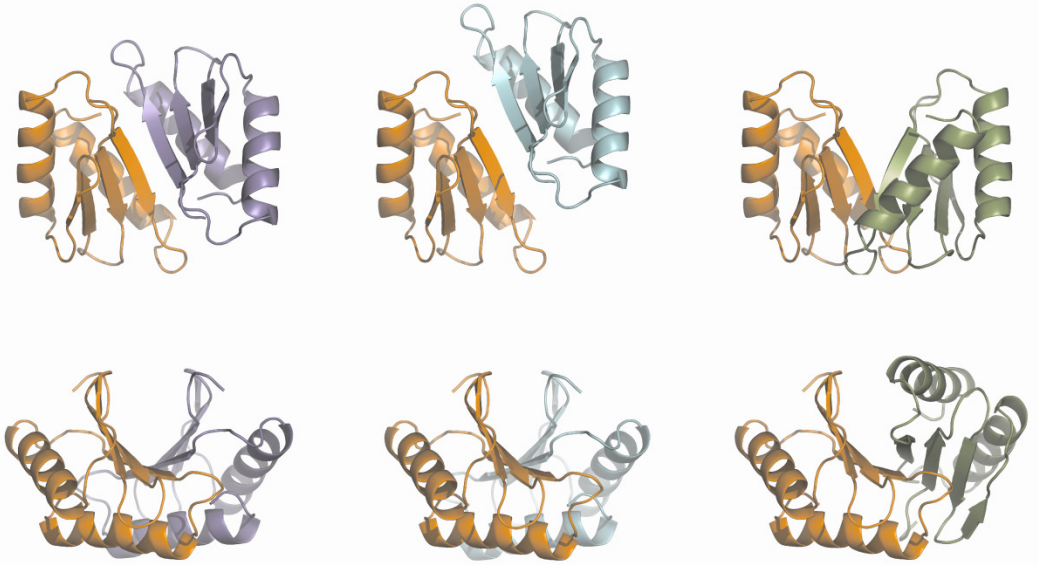


# Rigid body assembly: TolR $C_2$ dimer from RDC and SAXS data only



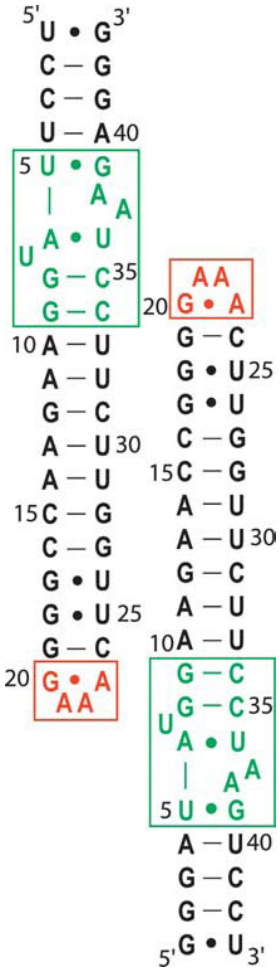



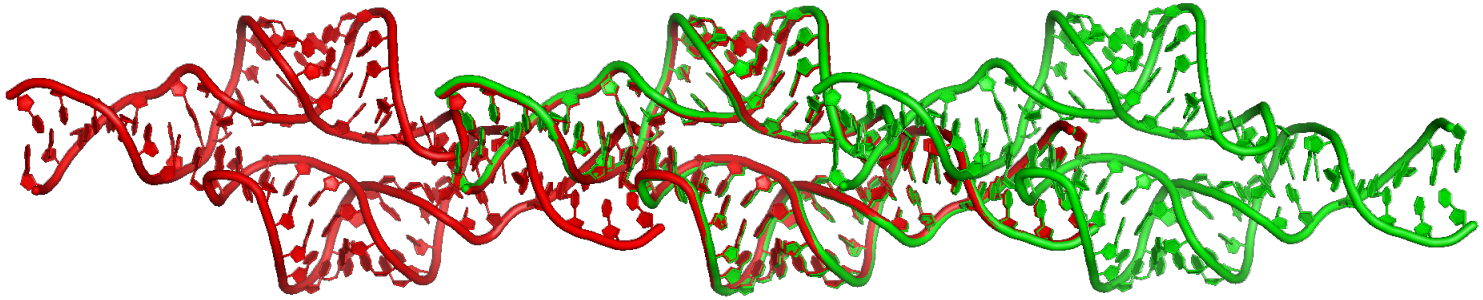



$\chi_{SAXS}$ =1.127 $\chi_{SAXS}$ =1.200 $\chi_{SAXS}$ =1.457

Parsons et al. (2008). Biochemistry 47, 3131-3142.

# GAA Tetra-loop receptor homodimeric RNA: GASR approach



NMR structure
Interface defined by
36X2 inter-NOE

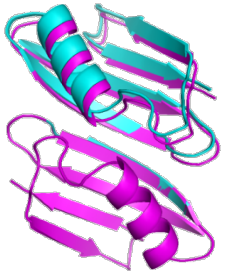
GASR structure
Interface defined by
SAXS data

rmsd: 0.4 Å

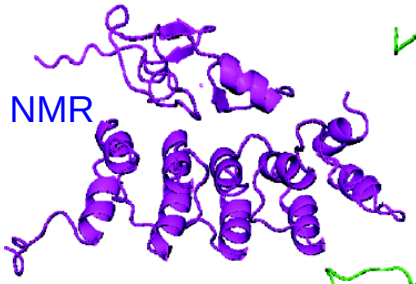
Zuo, et al (2008) JACS 130, 3292-3293.

# Dimeric and 2-domain proteins: GASR approach

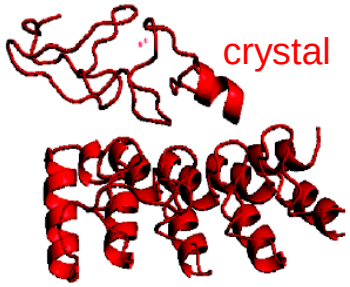
## homodimer (GB1)

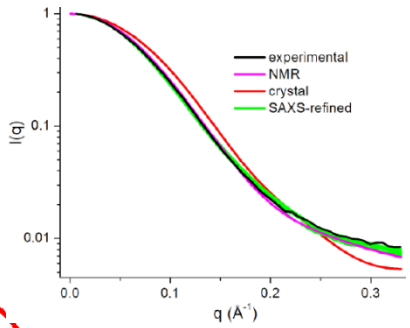


NMR
vs
NMR-SAXS



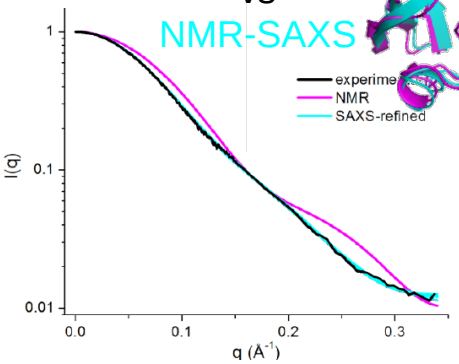

## heterodimer (ILK ARD-PINCH LIM1)

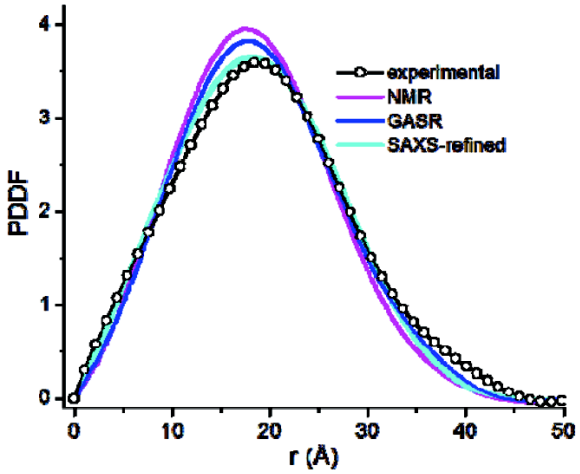



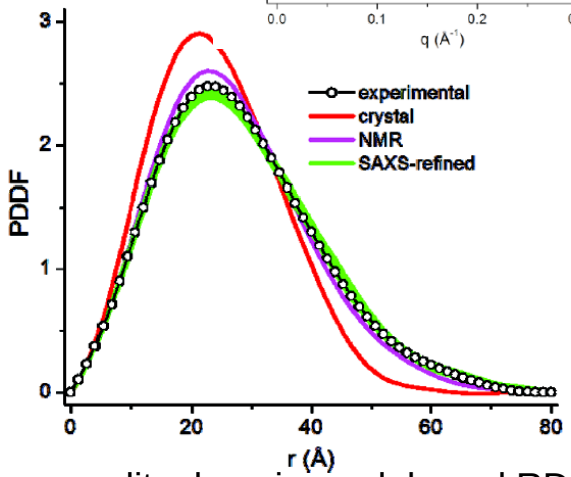

## 2-domain (L11)

NMR
vs
NMR-SAXS

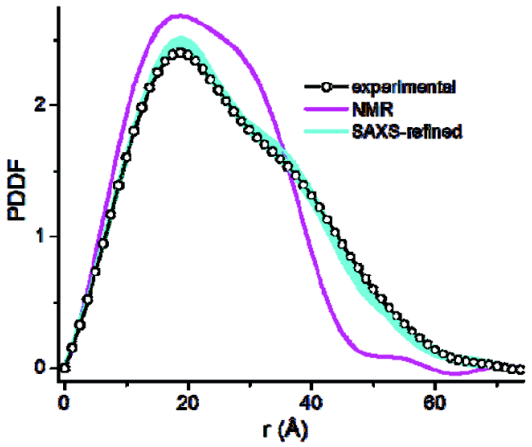



~1 Å rmsd between
NMR and NMR-SAXS



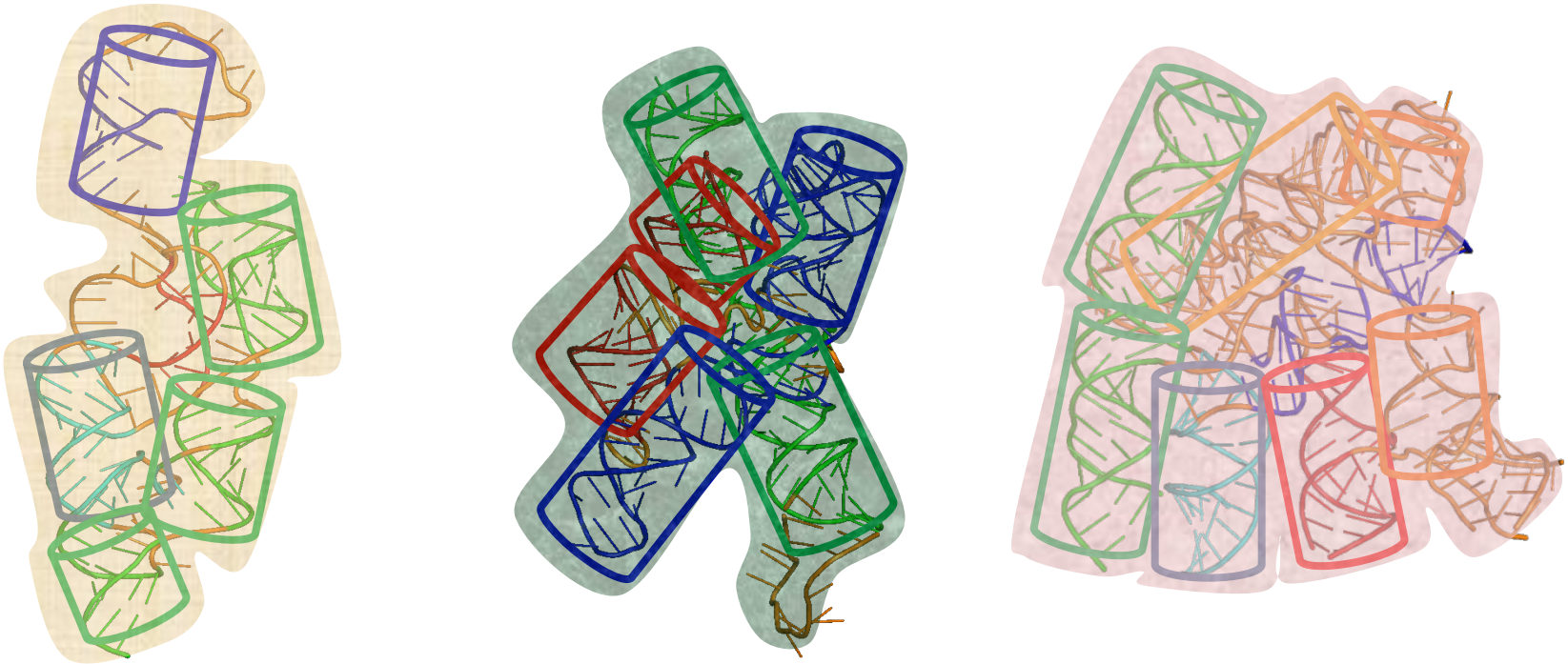

Low quality domain models and RDC data
Non-structured regions
RDC-SAXS: transitional shift w/out NOE restraint
~4 Å rmsd between NMR and NMR-SAXS




~5 Å rmsd between
NMR and NMR-SAXS

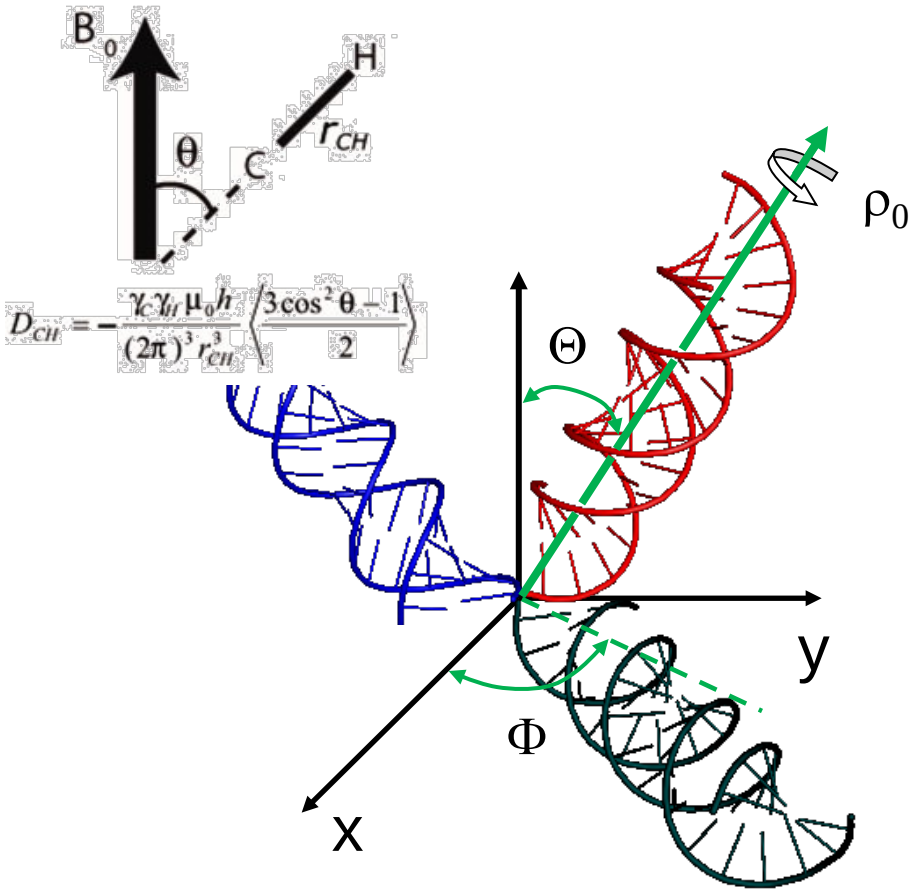
Wang, et al, (2009) JACS, 131,10507-10515.

# Global RNA structure by combination of RDC-derived relative orientations with SAXS-derived low-resolution shapes: the G2G method

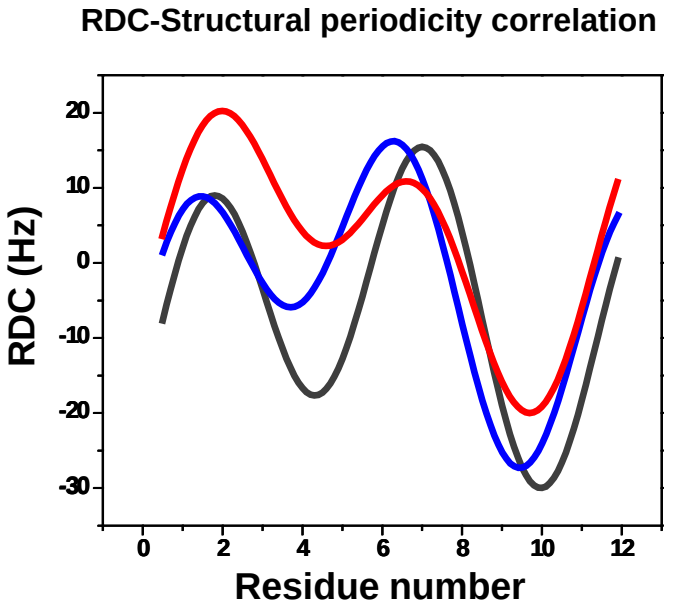


- **76% nucleotides are in duplex form.**
- **The global structure of a well-folded RNA is roughly the packing of duplexes.**
- **Duplex packing = duplex orientation + relative position.**

# Orientations of duplexes can be determined using RDC-structural periodicity correlations




$$RDC = f(Da, R, \Theta, \Phi, \rho_0, n)$$



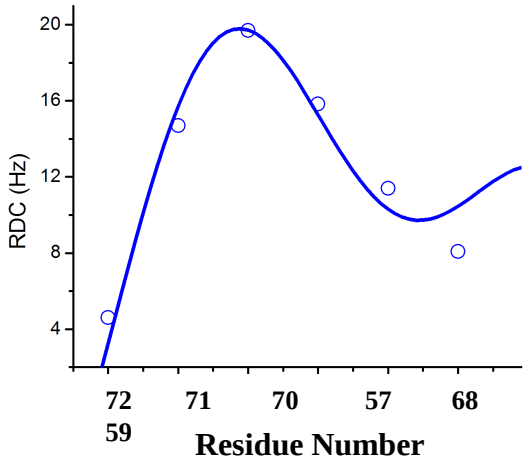

Orientation: (Θ, Φ)

Rotation/phase: $\rho_0$

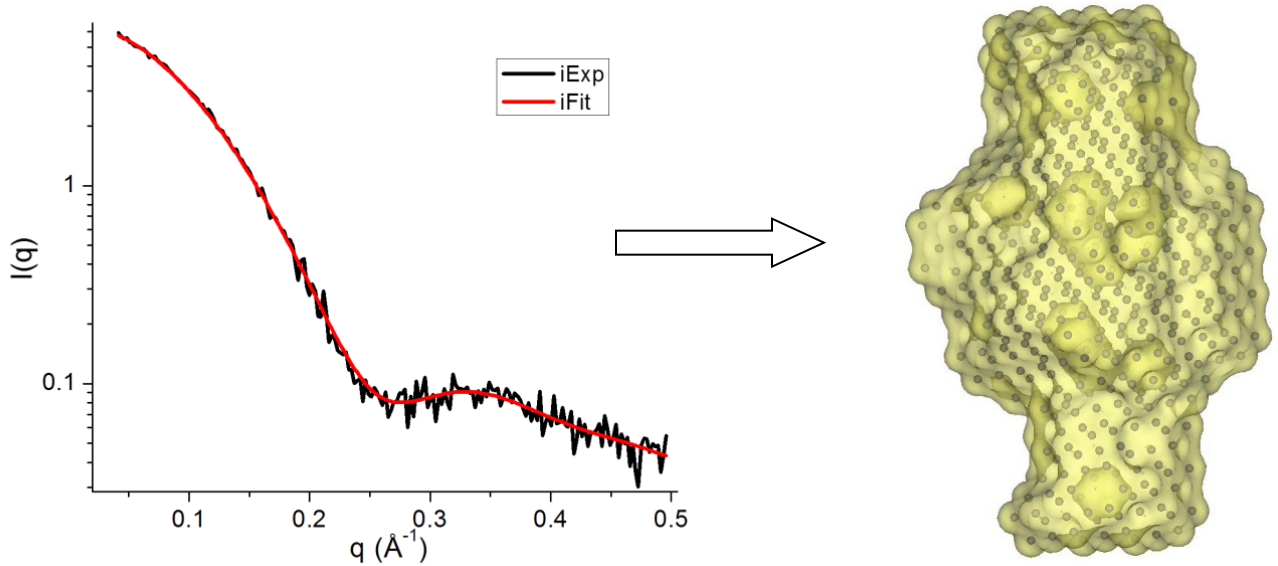
Walsh et al., JACS, 2004.

# from Global measurements to Global structures: The G2G method

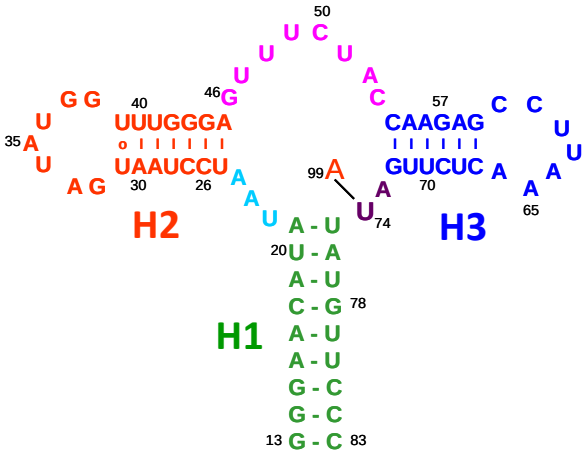
- The orientations of duplexes extracted from the RDC-structural periodicity

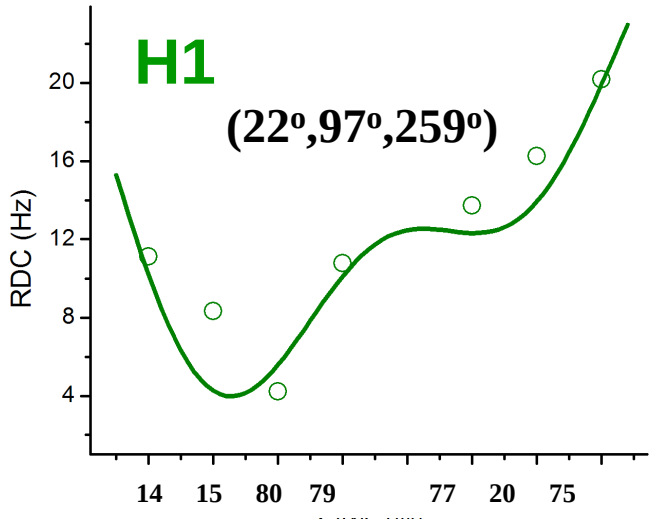


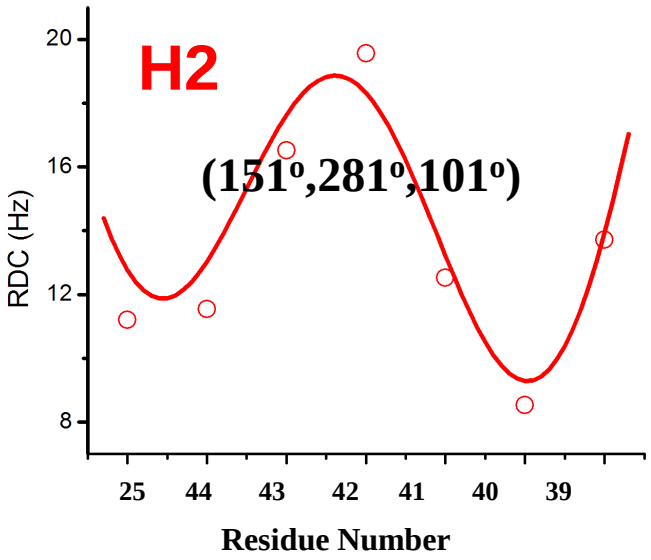

- The relative position of duplexes are defined by SAXS data / shape.

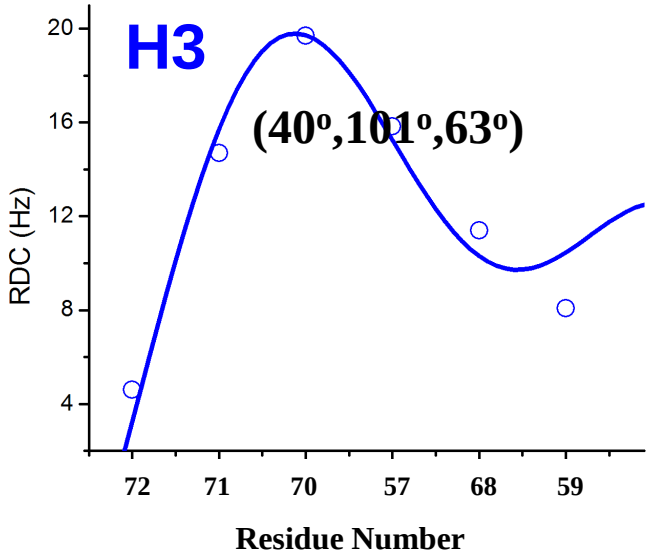


# 71-nt Adenine riboswitch (riboA) RNA: RDC structure correlation

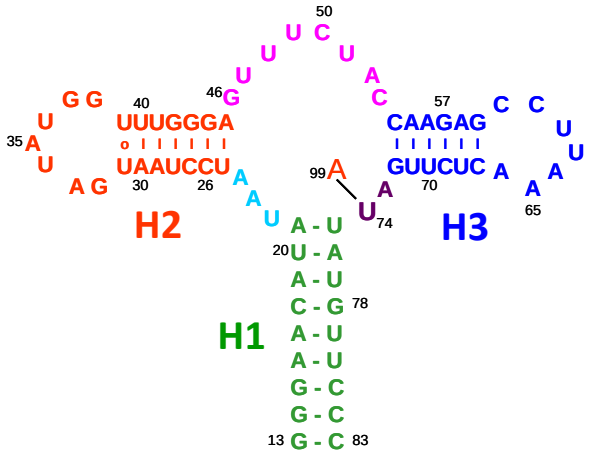


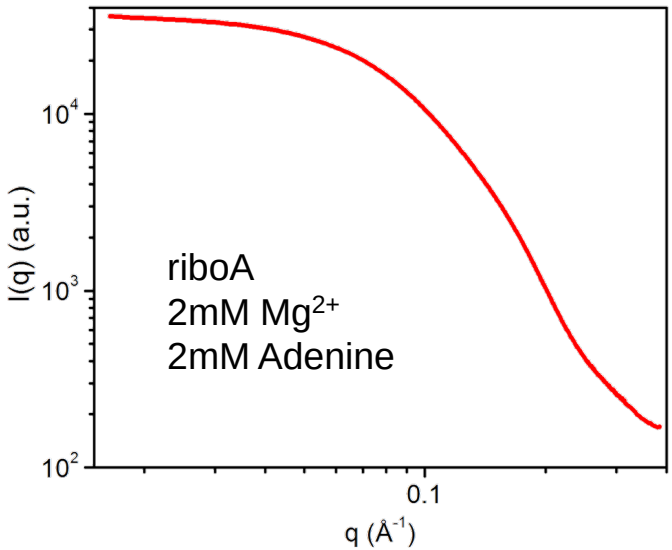



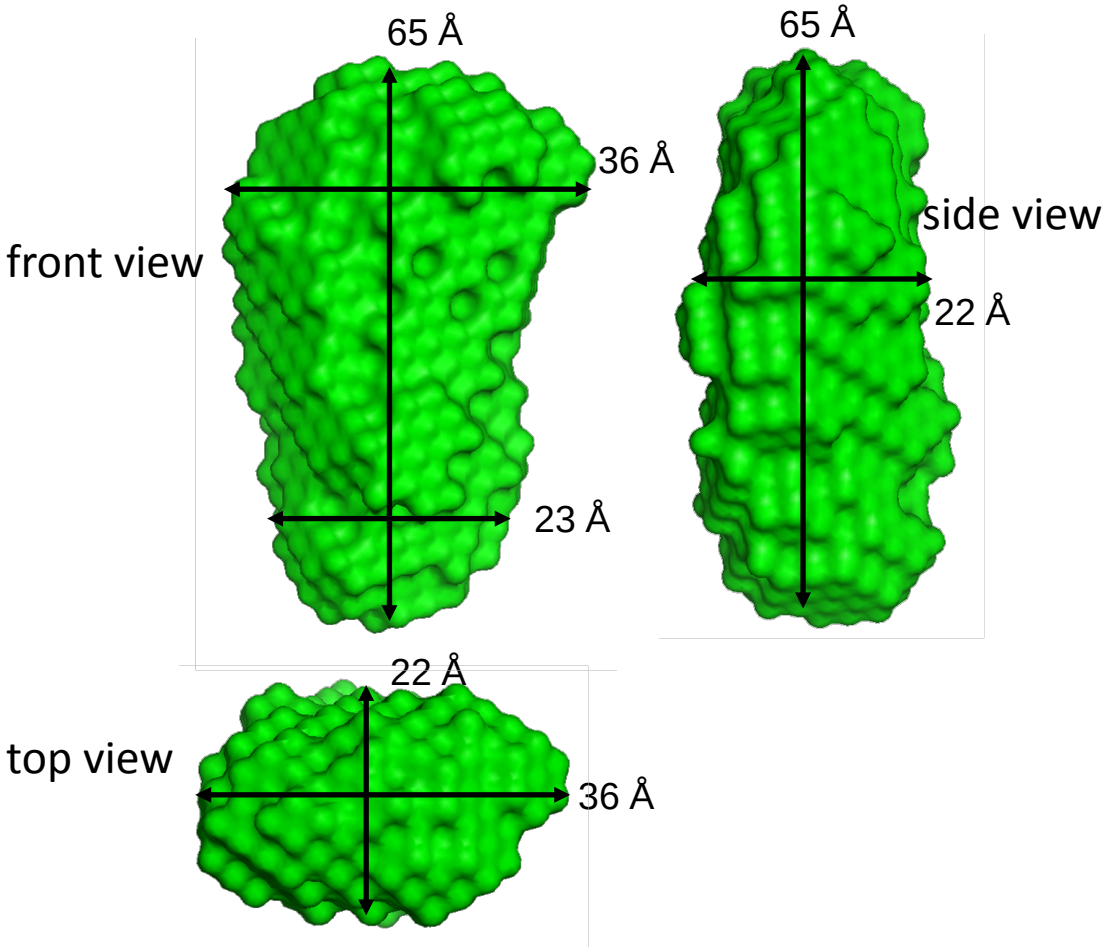



# 71-nt Adenine riboswitch (riboA) : SAXS molecular envelope and dimensions



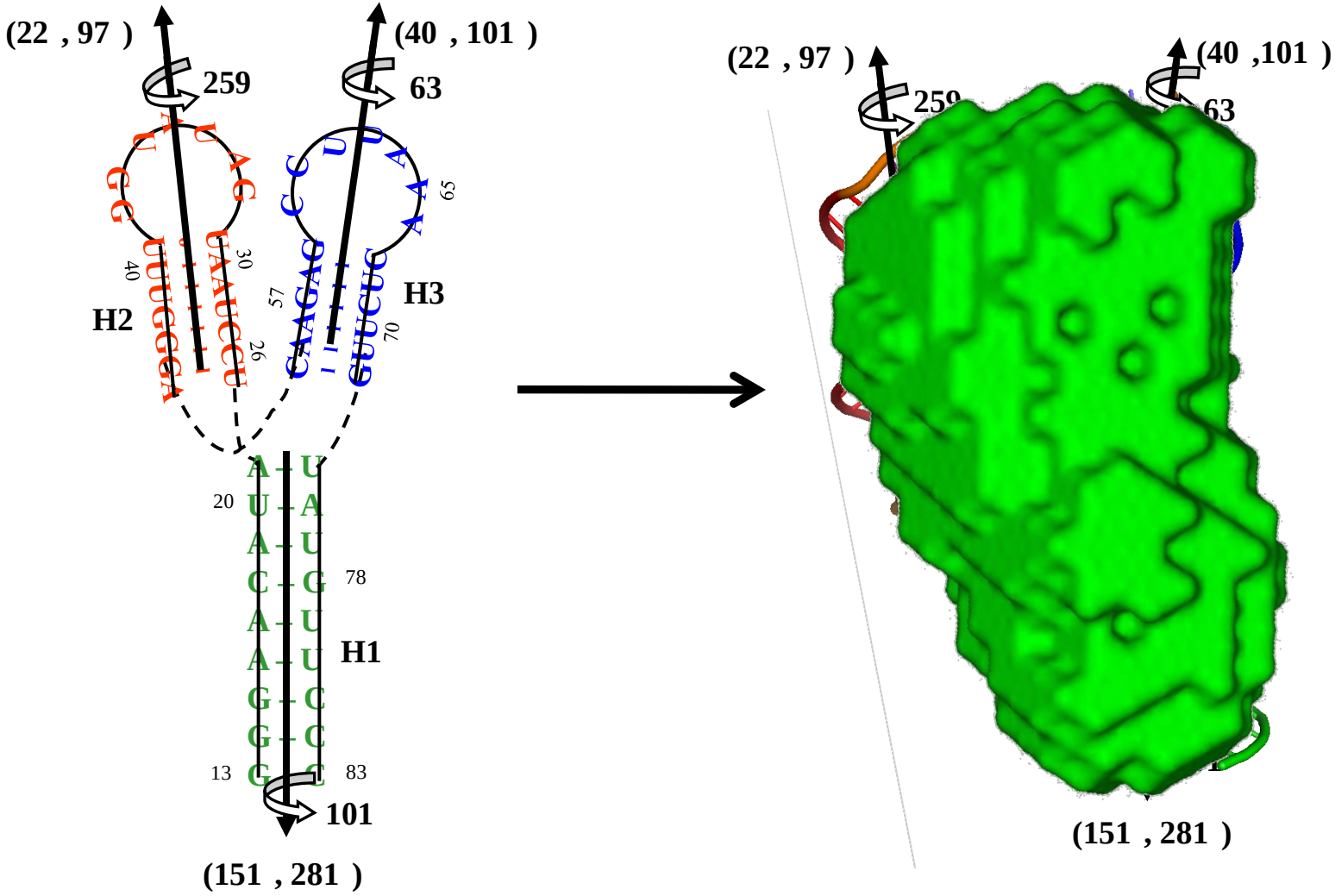




**SAXS data for riboA**
**$R_g$=19.2 Å**

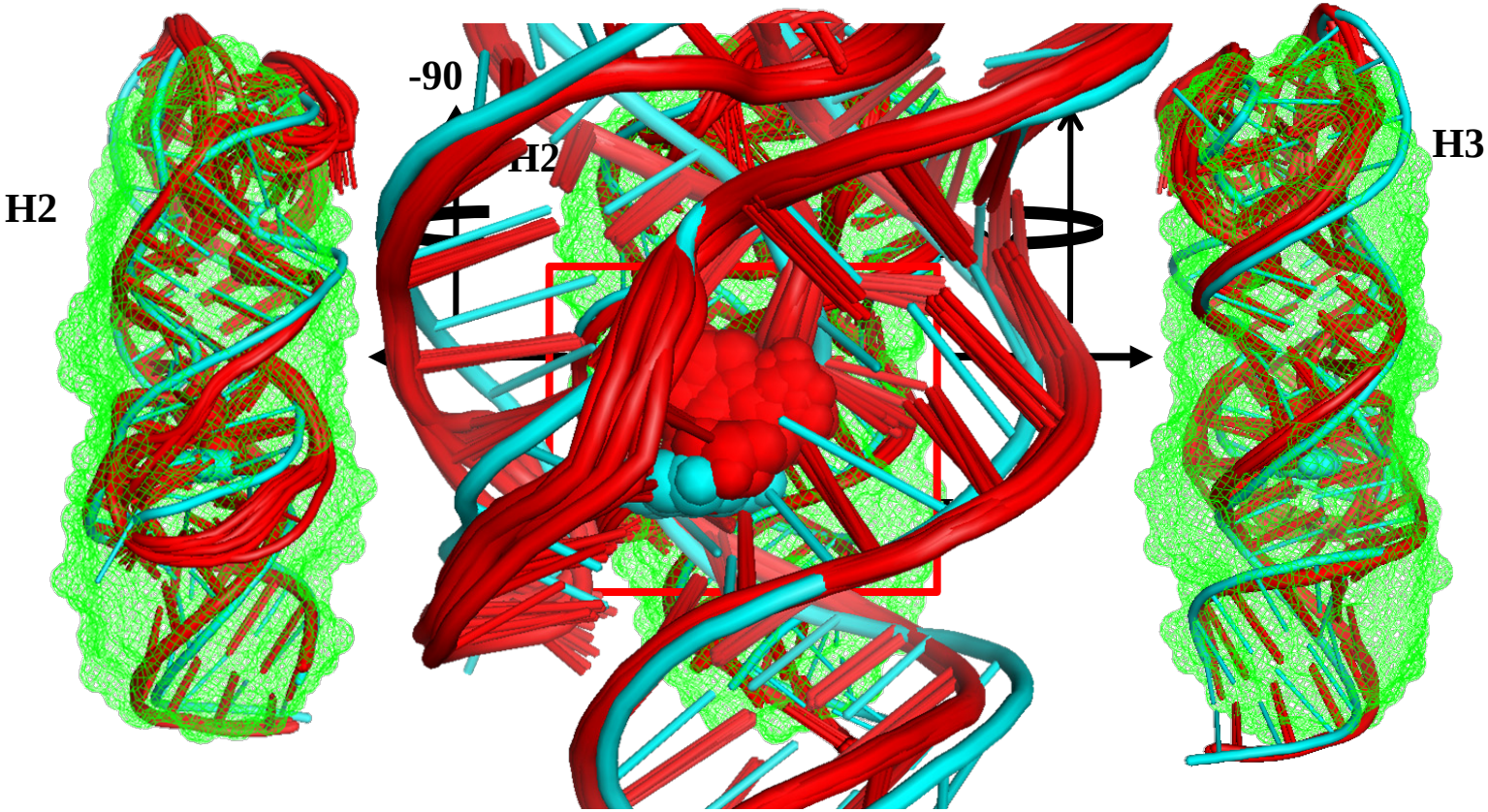



- Duplexes H1, H2, H3 are either parallel or anti-parallel fashion

# From 2D topology to 3D initial structure



# G2G can determine a structure to an accuracy that the approximate ligand binding pocket is well defined

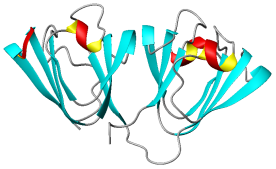


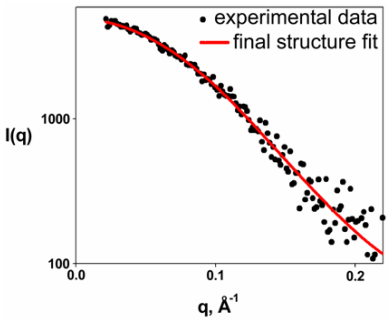

Overall backbone RMSD between the G2G (red) and X-ray crystal (cyan) structures: ~3.0  0.4 Å

Wang et al. (2009) J. Mol. Biol. 393, 717-734.

# High-resolution structure refinement with SAXS data: some observations

## γS crystallin (21 kDa, 2 domains)





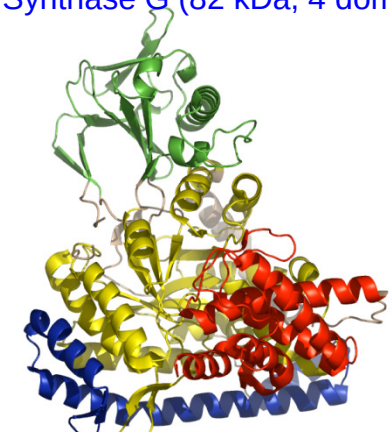

1.4 NOEs and 6 RDCs / residue

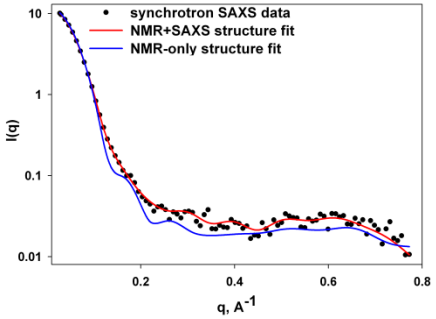
Accuracy improvements for both 2-domain and 1-domain geometries

Validated by later measurement of long-range $CH_3$-$CH_3$ NOEs

Rigid-body domain positioning against RDS and SAXS data unsuccessful due to high symmetry

## Malate Synthase G (82 kDa, 4 domains)

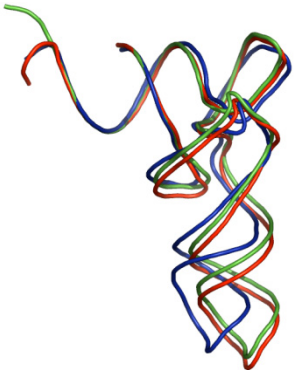


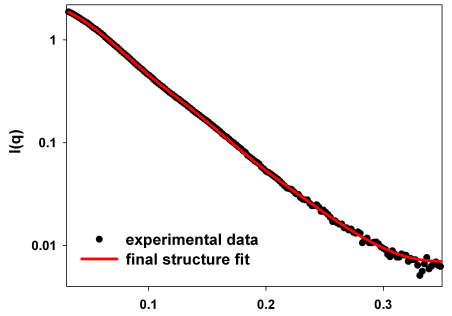


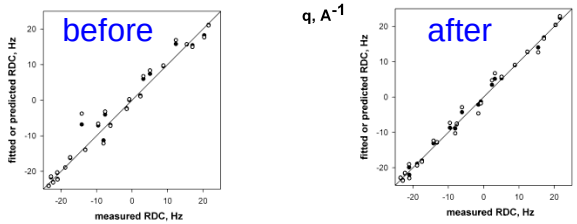

Accuracy improvement depends on the amount of NMR data:

from 4.6 Å (vs 1D8C) to:
3.2 Å with TALOS $\phi/\psi$
2.6 Å with TALOS+ $\phi/\psi$(100 extra)

## $tRNA^{Val}$ (24 kDa, 4 stem-loops)






Only 24 base N-H RDCs!

Locally rigid, globally flexible refinement

Accuracy improvement is confirmed by RDC cross-validation

# Part Three

## Neutron Scattering

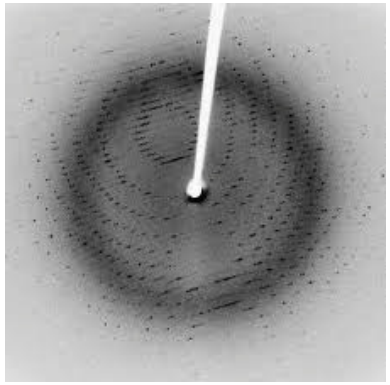
Andrew Whitten

University of Queensland, Australia

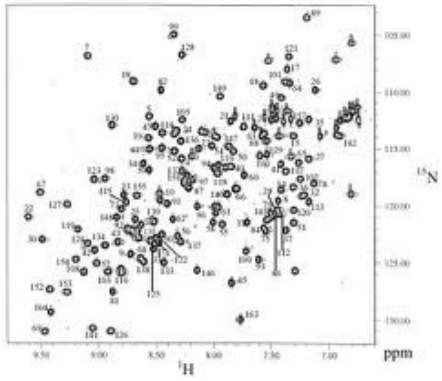
# Structure determination

## High-resolution

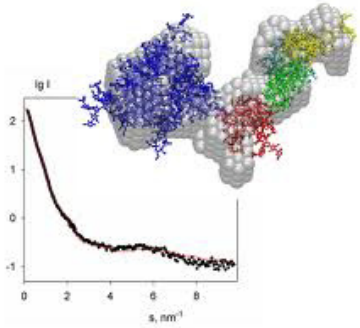
## Low-resolution



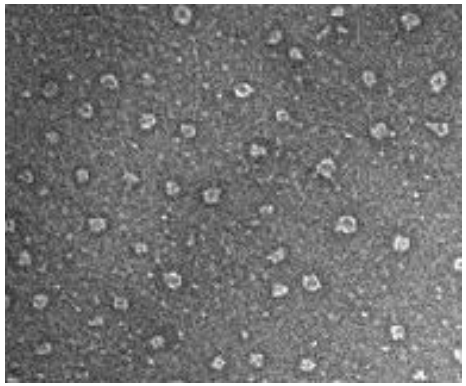
Crystallography




NMR




Scattering



Electron Microscopy

- Theory
- Experimental design
- Data collection
- Data analysis and modeling
- Examples

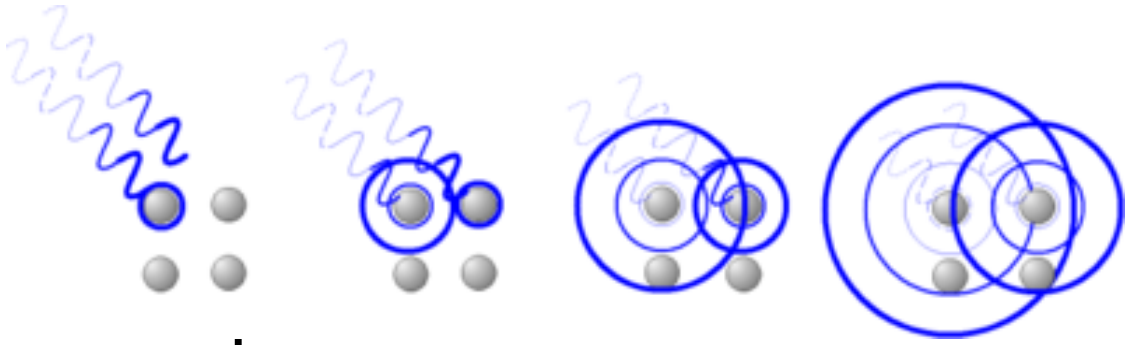
# Contributions to the scattering profile

$$I(q) = N\left(\Delta\bar{\rho}V\right)^2 P(q)\, S(q)$$

- Number of particles
- Composition of each particle and solvent
- Size of each particle
- Distribution of the contents
- Correlations between particles

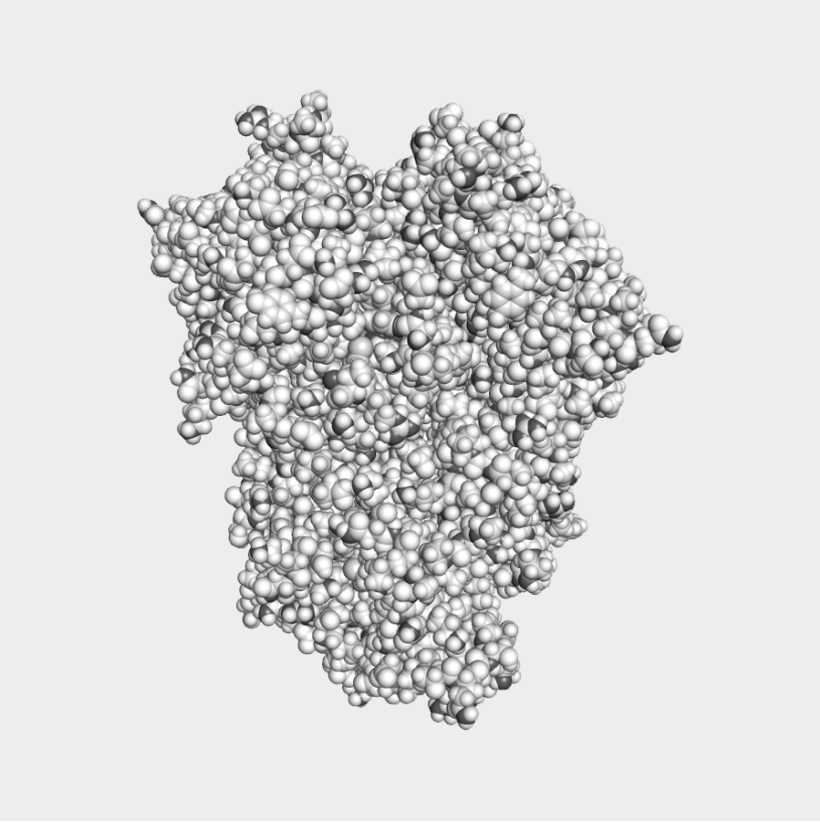
# Volume

- Scattering arises from interference



- This means that
  - While doubling the number of particles doubles the intensity
  - Doubling the volume of a particles quadruples the intensity (i.e. $I(q)$ is proportional to $V^2$

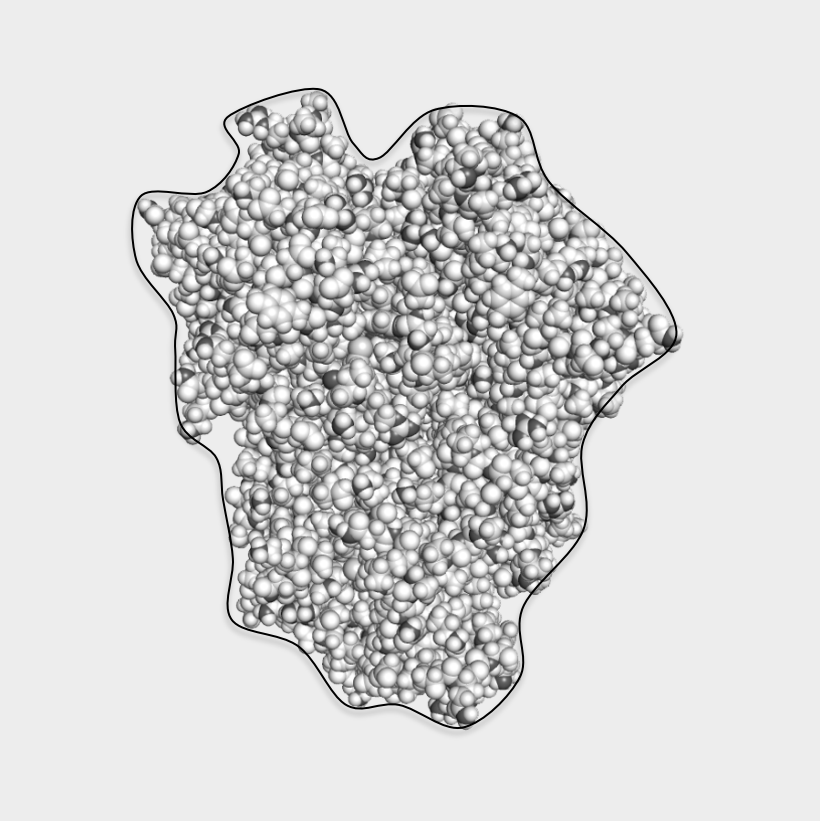
# Contrast



High-resolution image

Smeared image ~10Å/pixel

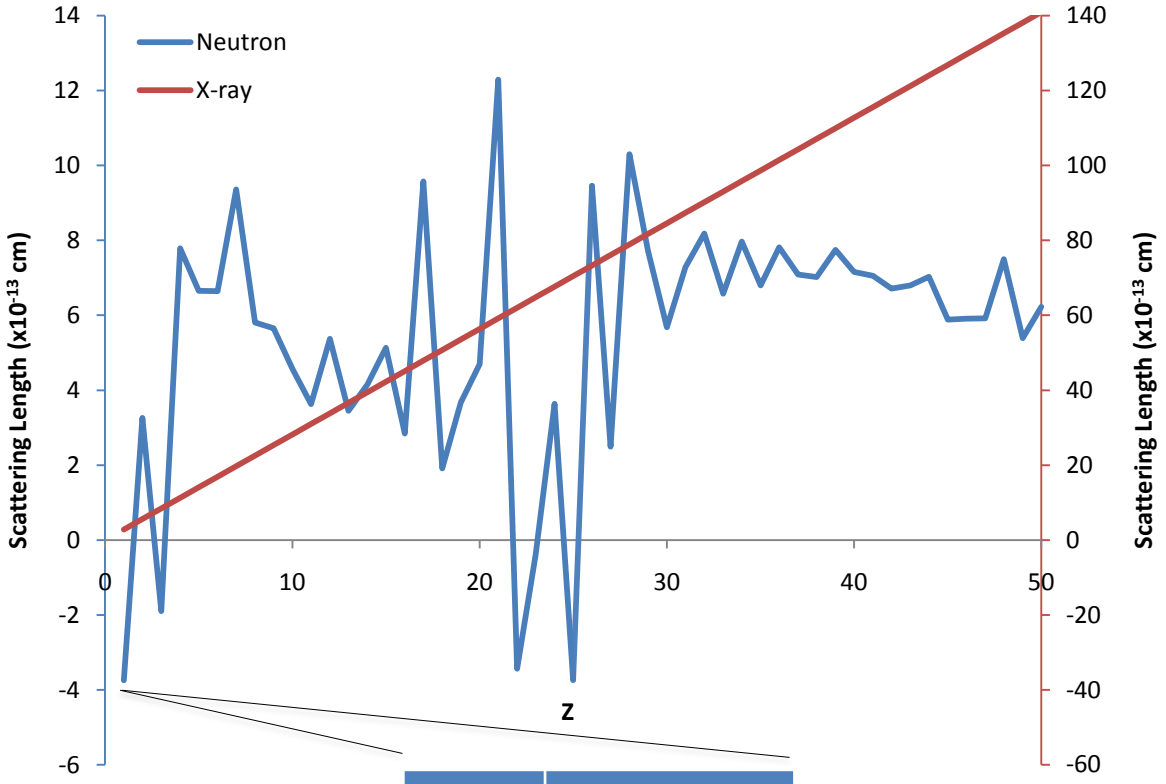
# Contrast



High-resolution image

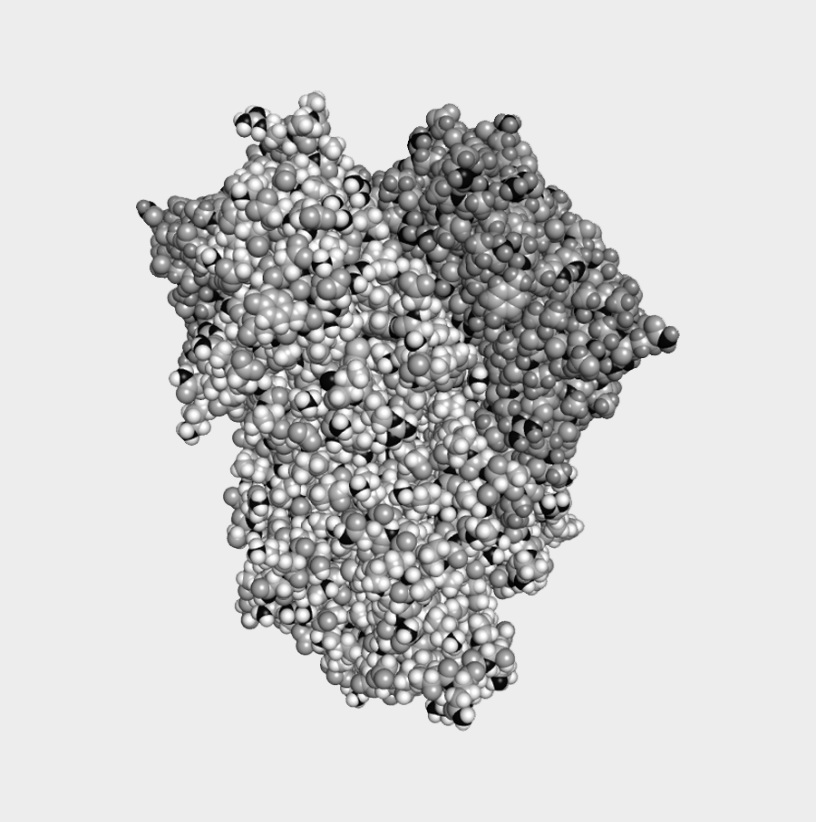
Constant density ~10Å/pixel

# Creating contrast differences

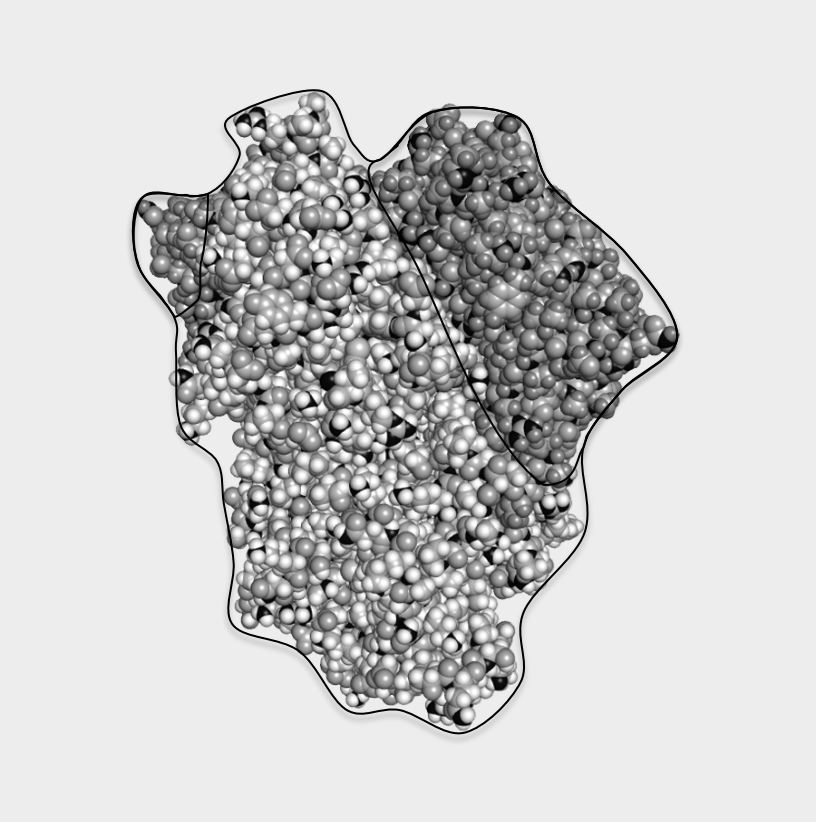


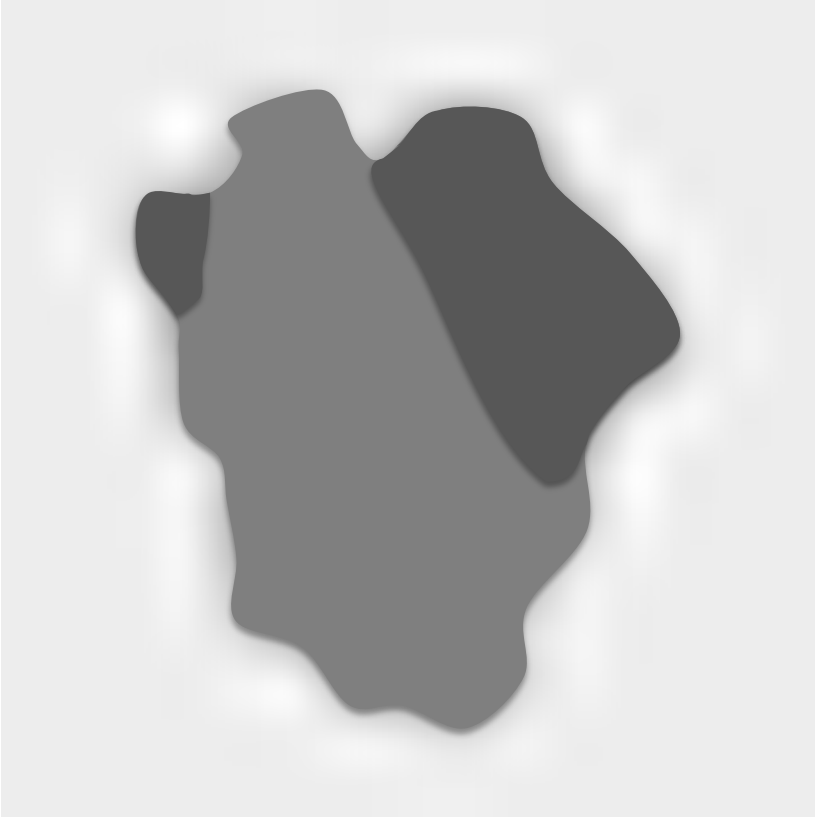

| Isotope | S.L. (x$10^{-13}$ cm) |
|---|---|
| $^{1}$H | -3.74 |
| $^{2}$H (D) | 6.67 |
| $^{3}$H | 4.79 |

# Contrast variation



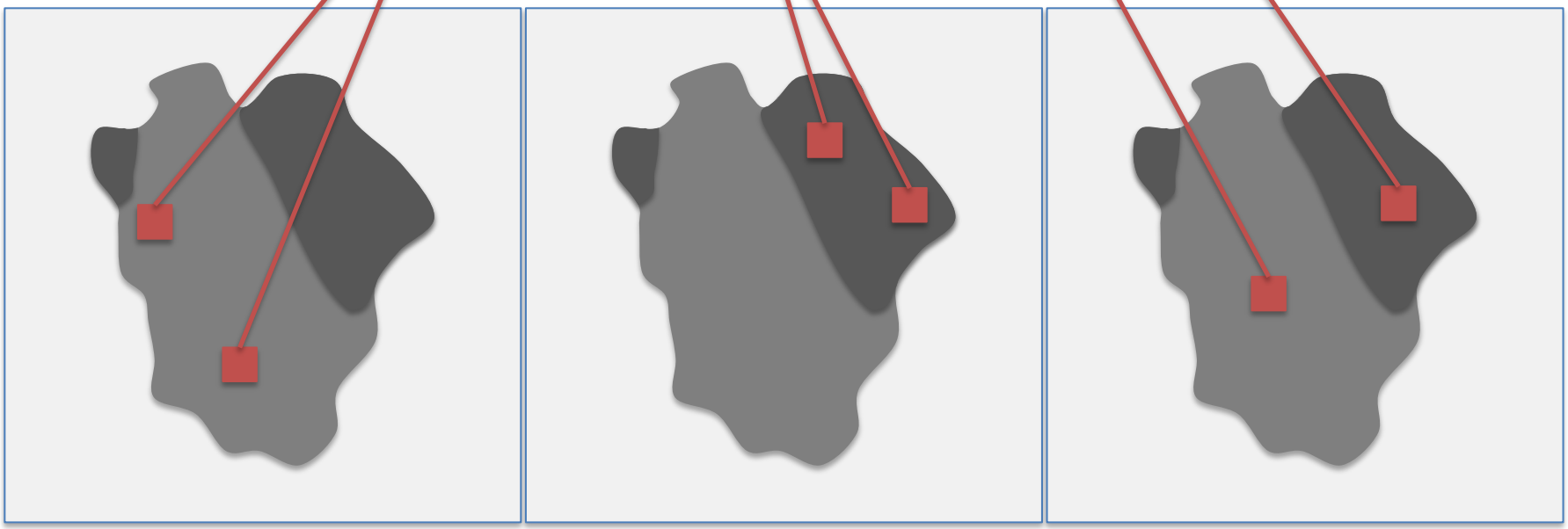
# Contrast variation





# Contrast variation

$$I(q) = \left(\Delta\overline{\rho}_A V_A\right)^2 P_A(q) + \left(\Delta\overline{\rho}_B V_B\right)^2 P_B(q) + \Delta\overline{\rho}_A V_A \Delta\overline{\rho}_B V_B P_{AB}(q)$$



# Contrast variation

$$I(q) = \left(\Delta\overline{\rho}_A V_A\right)^2 P_A(q) + \left(\Delta\overline{\rho}_B V_B\right)^2 P_B(q) + \Delta\overline{\rho}_A V_A \Delta\overline{\rho}_B V_B P_{AB}(q)$$

# Contrast variation

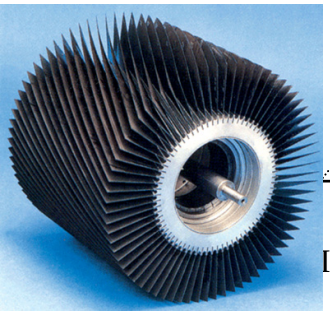
$$I(q) = \left(\Delta\overline{\rho}_A V_A\right)^2 P_A(q) + \left(\Delta\overline{\rho}_B V_B\right)^2 P_B(q) + \Delta\overline{\rho}_A V_A \Delta\overline{\rho}_B V_B P_{AB}(q)$$

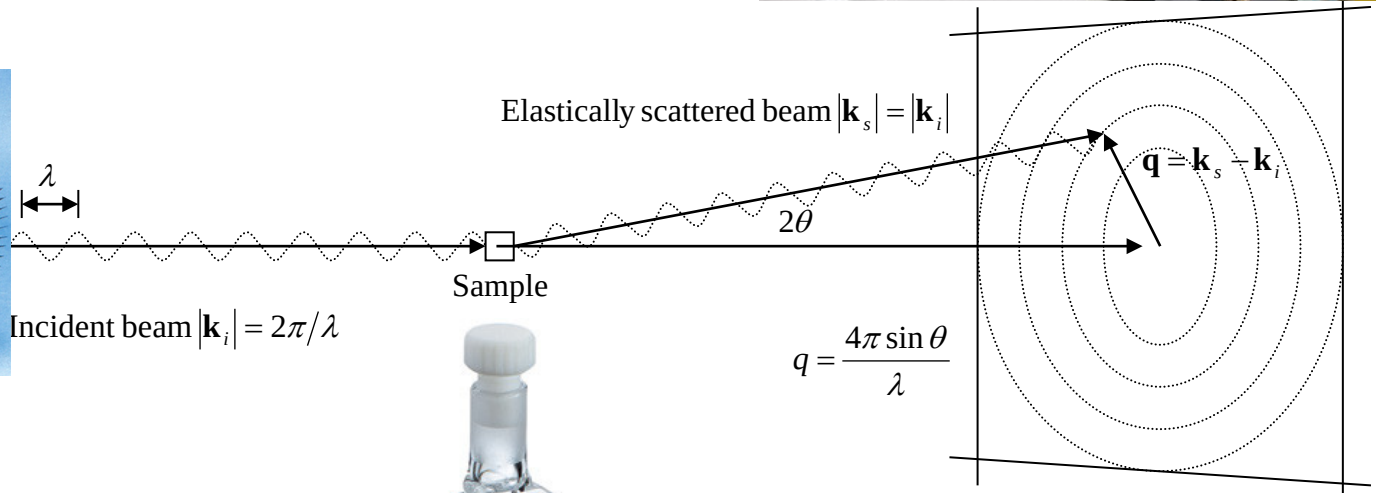
# Contrast variation

$$I(q) = \left(\Delta\overline{\rho}_A V_A\right)^2 P_A(q) + \left(\Delta\overline{\rho}_B V_B\right)^2 P_B(q) + \Delta\overline{\rho}_A V_A \Delta\overline{\rho}_B V_B P_{AB}(q)$$

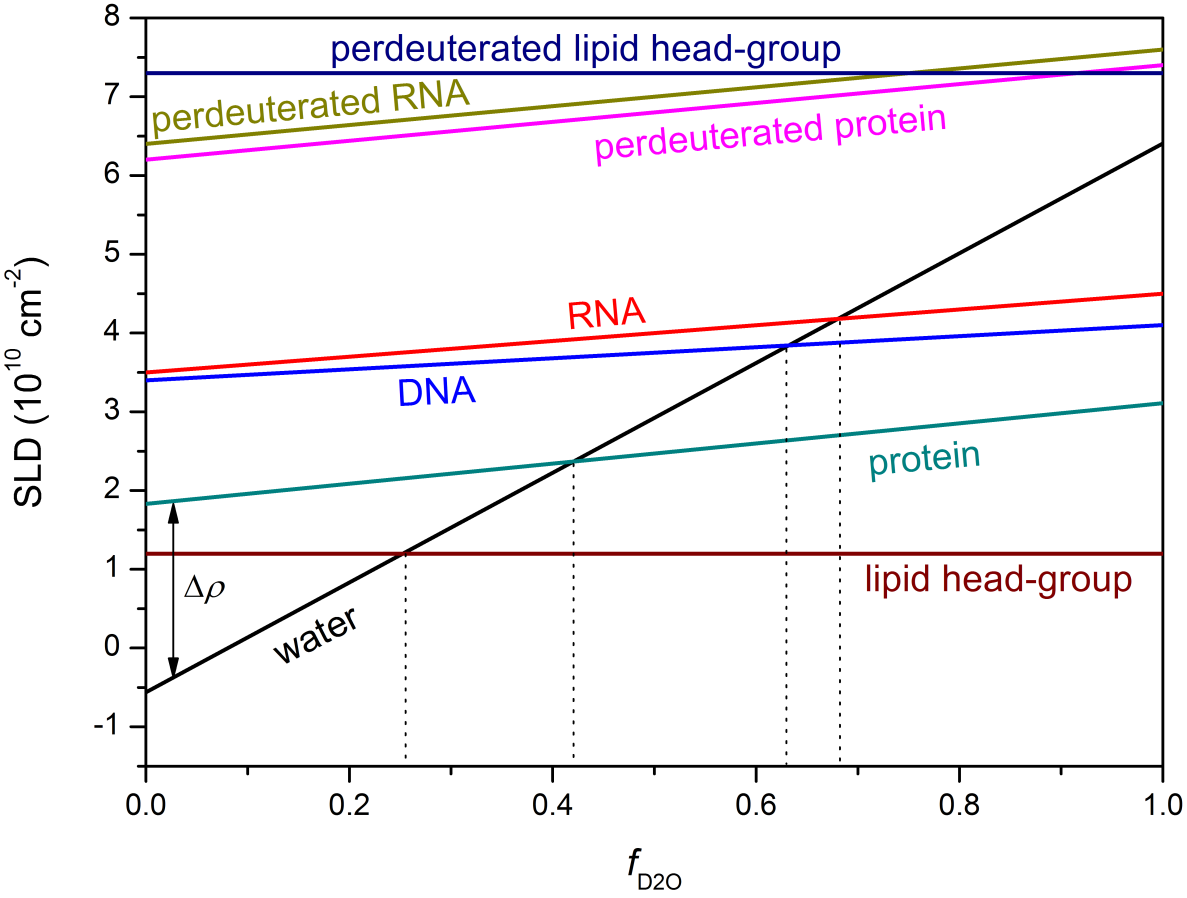
- Theory
- Experimental design
- Data collection
- Data analysis and modeling
- Examples

# The scattering experiment



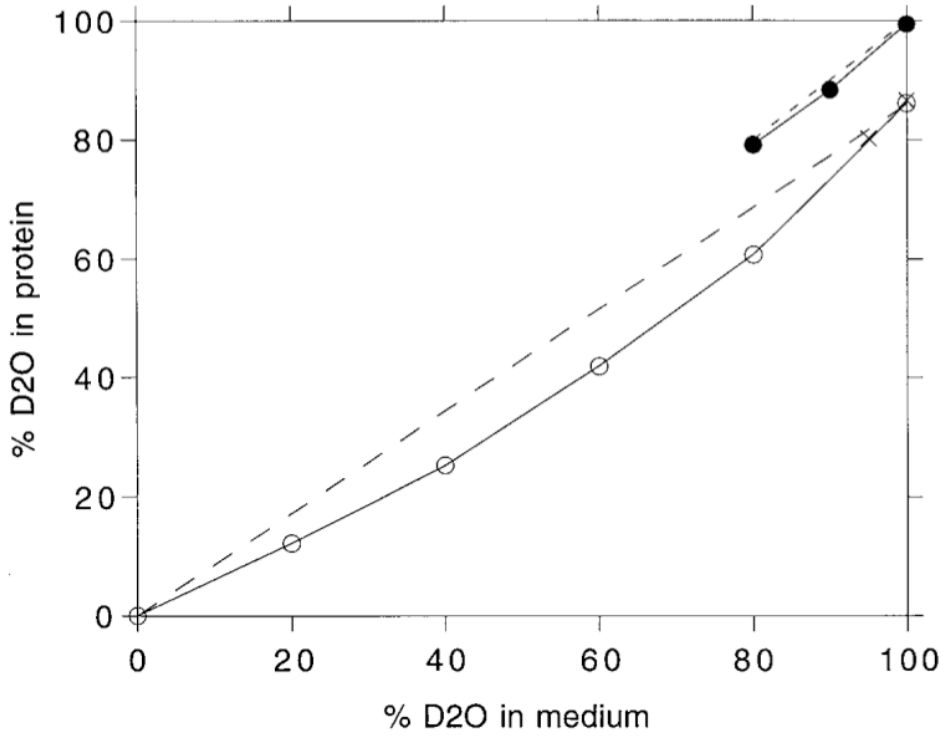



# Contrast of various particles



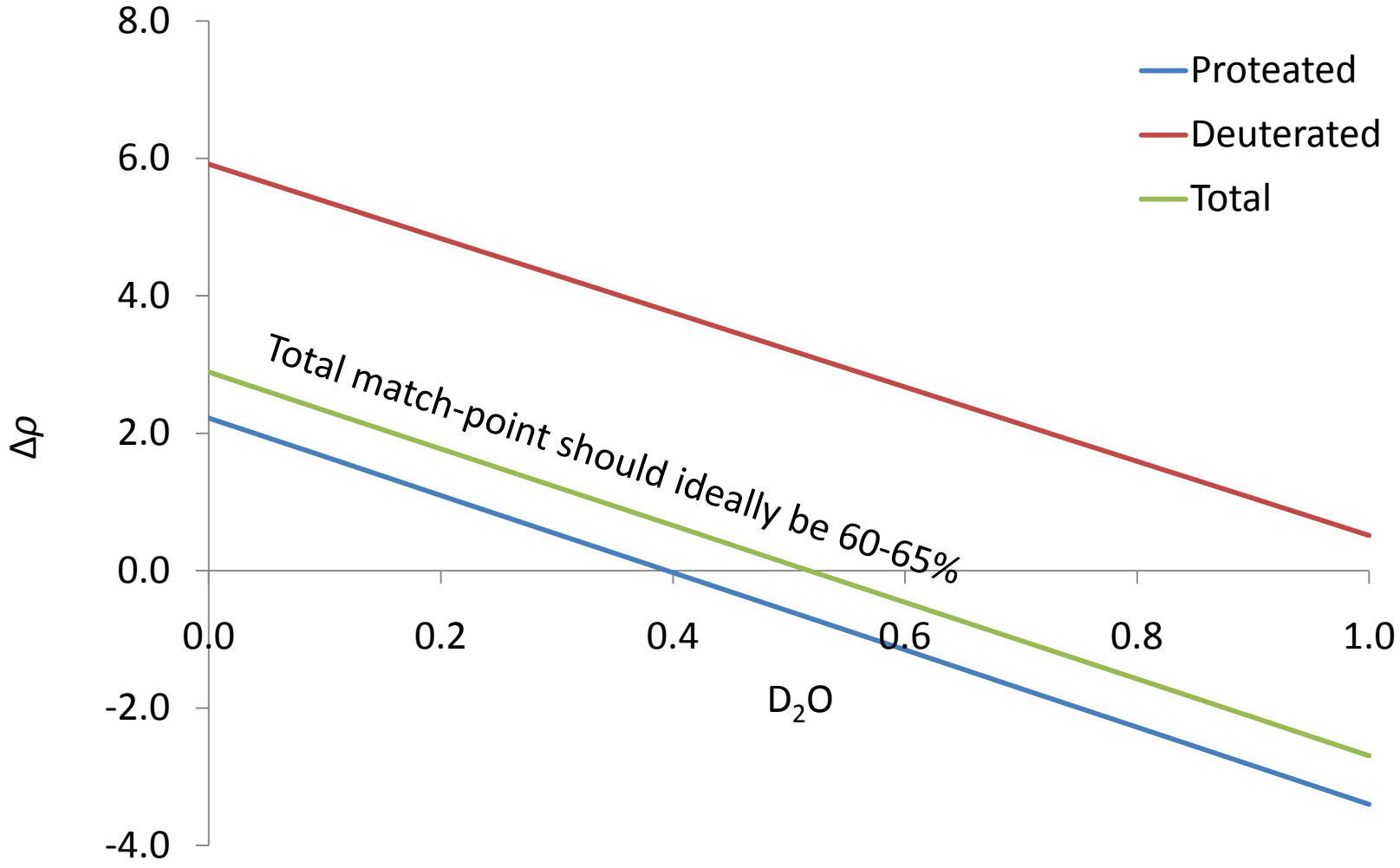
# Choosing the deuteration level

- Trade off between:
  - Cost
  - Practicality
  - Purpose
    - Contrast matching requires high levels of deuteration
    - Contrast variation requires a balance

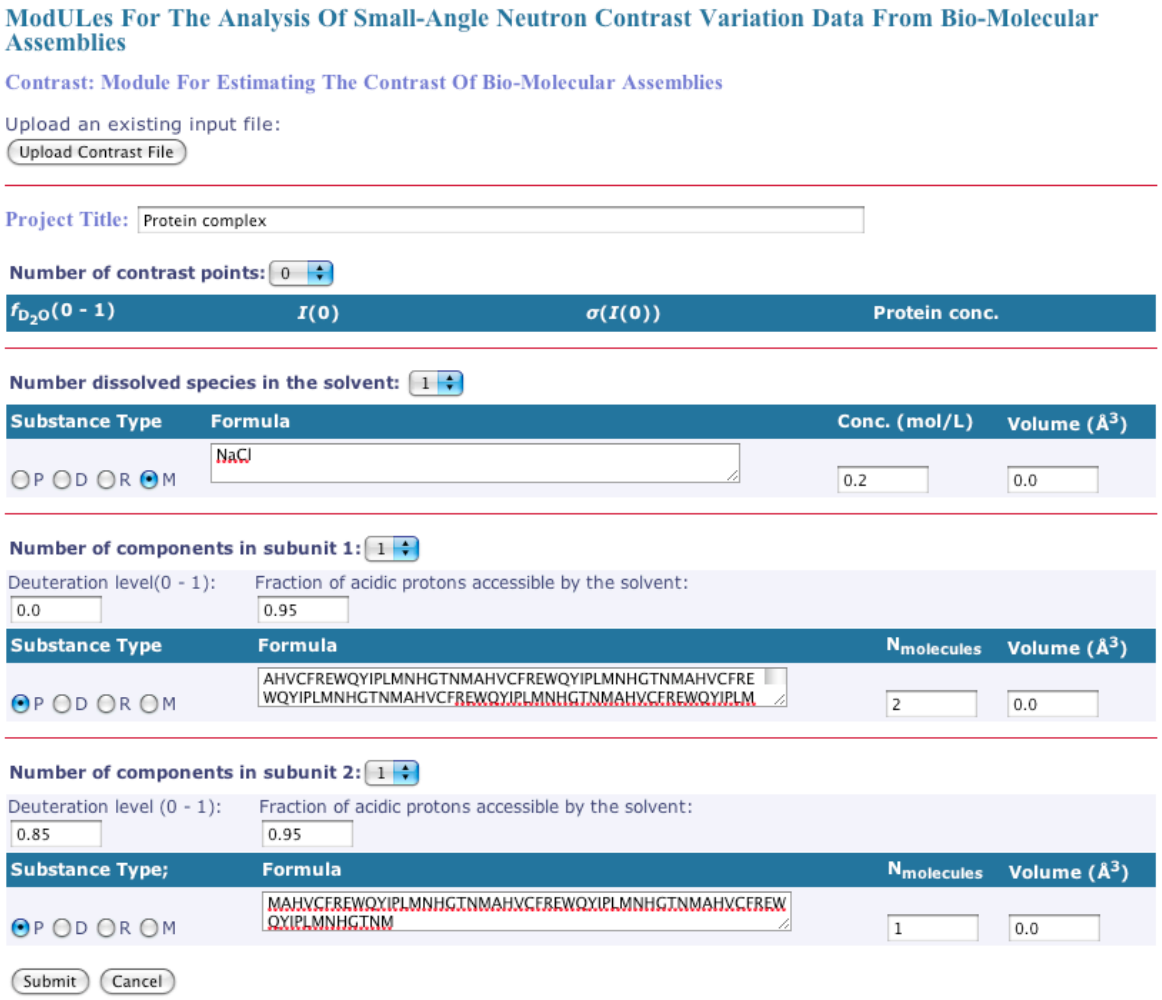



Leiting et al. Analytical Biochemistry, 265, 351–355 (1998)

# Choosing the deuteration level



# Contrast calculation




- (http://research.mmb.usyd.edu.au/NCVWeb/)

# Contrast calculation

*Tabulated scattering length densities and contrasts*

| | $\rho$ ($10^{10}$cm$^{-2}$) | | | $\Delta\rho$ ($10^{10}$cm$^{-2}$) | | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **Solvent** | **1** | **2** | **Total** |
| **X-RAY** | 12.343 | 12.337 | 9.471 | 2.871 | 2.866 | 2.870 |
| **NEUTRON** | | | | | | |
| **0.0** | 2.018 | 5.631 | -0.542 | 2.560 | 6.173 | 3.294 |
| **0.1** | 2.136 | 5.749 | 0.150 | 1.986 | 5.599 | 2.720 |
| **0.2** | 2.254 | 5.867 | 0.842 | 1.412 | 5.025 | 2.146 |
| **0.3** | 2.373 | 5.986 | 1.535 | 0.838 | 4.451 | 1.572 |
| **0.4** | 2.491 | 6.104 | 2.227 | 0.264 | 3.877 | 0.998 |
| **0.5** | 2.609 | 6.222 | 2.920 | -0.310 | 3.302 | 0.424 |
| **0.6** | 2.728 | 6.340 | 3.612 | -0.884 | 2.728 | -0.150 |
| **0.7** | 2.846 | 6.458 | 4.304 | -1.458 | 2.154 | -0.724 |
| **0.8** | 2.964 | 6.577 | 4.997 | -2.032 | 1.580 | -1.298 |
| **0.9** | 3.083 | 6.695 | 5.689 | -2.606 | 1.006 | -1.872 |
| **1.0** | 3.201 | 6.813 | 6.381 | -3.180 | 0.432 | -2.446 |
| **Calculated match-point ($f_{D_2O}$)** | | | | 0.446 | 1.075 | 0.574 |

# Choice of contrast points

- 0% $D_2O$ (maximize signal of particle)
- 40% $D_2O$ (signal dominated by deuterated component)
- 100% $D_2O$ (signal dominated by proteated component and incoherent scattering from $^1H$ minimized)

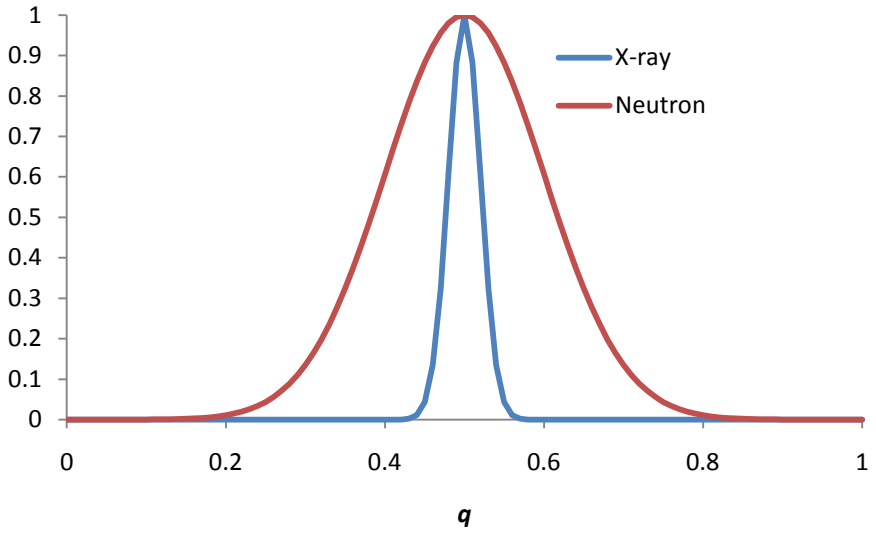
# Summary

- Contrast matching:
  - High  deuteration level (that is practical)
  - Measurement at 40% $D_2O$
- Contrast variation
  - Deuteration tuned so the total match-point of the complex is ~60-65% $D_2O$
  - Measurement at 0%, 20%, 40%, 80% and 100% $D_2O$

- Theory
- Experimental design
- Data collection
- Data analysis and modeling
- Examples

# Smearing

- Characteristics of the neutron beam
  - Large beam (~10 mm at sample)
  - High dispersion (10% wavelength spread)

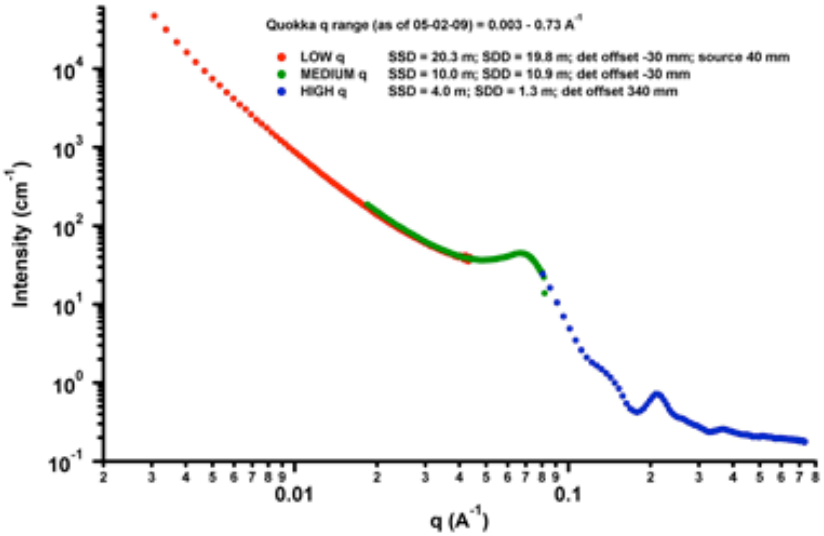


# Concentration

- Concentration requirements for a SANS experiment are high:
  - Total concentration > 5 mg/mL (for a 50 kDa protein)
  - Concentration of deuterated component > 2 mg/mL (for a 20 kDa protein)
  - Conditions optimized so that inter-particle interactions are low

# Merging data

- To cover an appropriate q-range data is collected in segments
  - Long detector position (low-q data, long exposure)
  - Short detector postion (high-q data, shorter exposure)



# Comparison

| Synchrotron | Neutron |
| --- | --- |
| Small sample (20-50 µL) | Large sample (>200 µL) |
| Concentration (< 5 mg/mL) | Concentration (> 5mg/mL) |
| Small beam size (200 x 100 µm) | Large beam size (10 x 10 mm) |
| High flux (1 x $10^{13}$ photons/sec) | Low flux (1 x $10^{8}$ photons/sec) |
| Ionizing | Non-ionizing |
| Fast experiment (seconds) | Slow experiment (hours) |
| | |

- Theory
- Experimental design
- Data collection
- Data analysis and modeling
- Examples

# Rg Analysis

- The Stuhrmann plot, is a method for analysing the dependence of the radius of gyration upon contrast due to scattering density fluctuations

$$R_g^2 = R_m^2 + \frac{\alpha}{\Delta\overline{\rho}} - \frac{\beta}{\Delta\overline{\rho}^2}$$

- $R_m$: Radius of gyration of the object with a homogenous distribution of scattering density

- The parallel axis theorem provides a straightforward path to $R_H$, $R_D$ and $D$

$$R_g^2 = \frac{\Delta\overline{\rho}_H V_H}{\Delta\overline{\rho}V} R_H^2 + \frac{\Delta\overline{\rho}_D V_D}{\Delta\overline{\rho}V} R_D^2 + \frac{\Delta\overline{\rho}_H V_H \Delta\overline{\rho}_D V_D}{\Delta\overline{\rho}^2 V^2} D^2$$

# Composite scattering functions

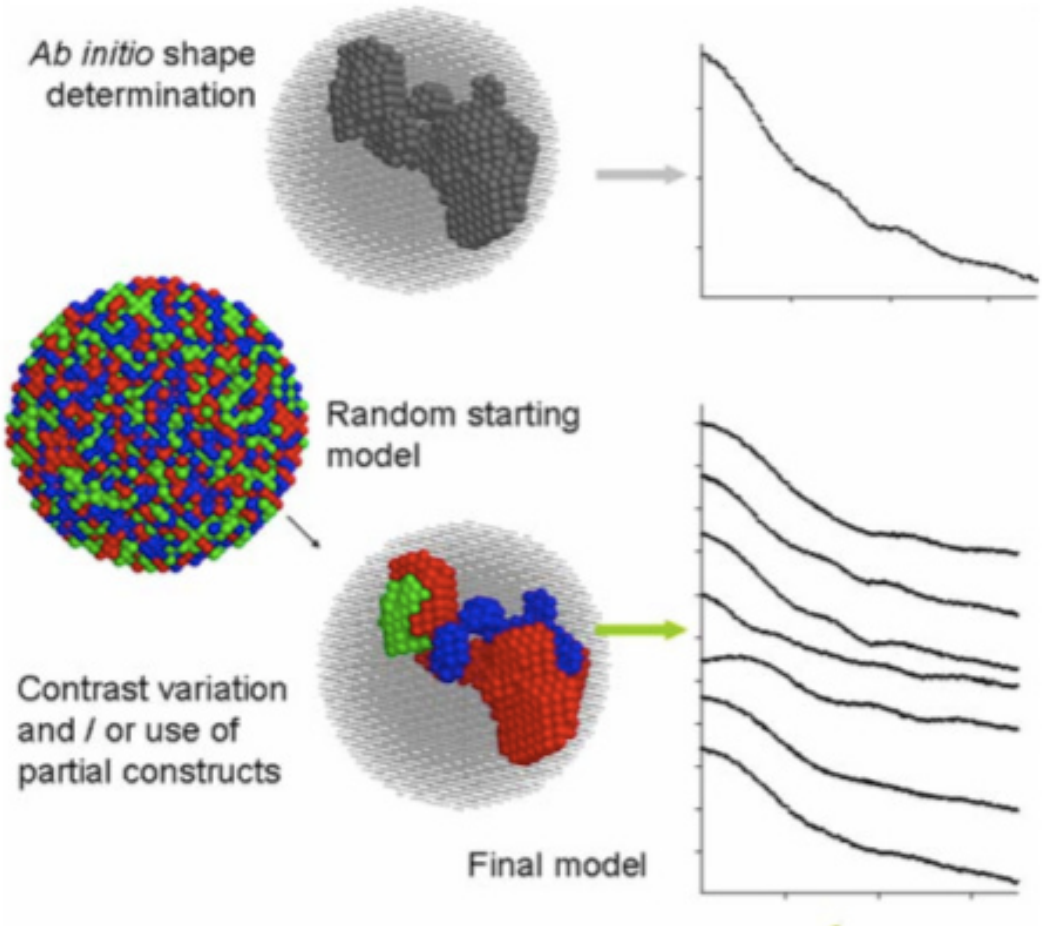
- Involves decomposing the contrast variation data into three profiles based on the equation

$$I(q) = \left(\Delta\overline{\rho}_A V_A\right)^2 P_A(q) + \left(\Delta\overline{\rho}_B V_B\right)^2 P_B(q) + \Delta\overline{\rho}_A V_A \Delta\overline{\rho}_B V_B P_{AB}(q)$$

- Allows the extraction of profiles akin to contrast matching experiments

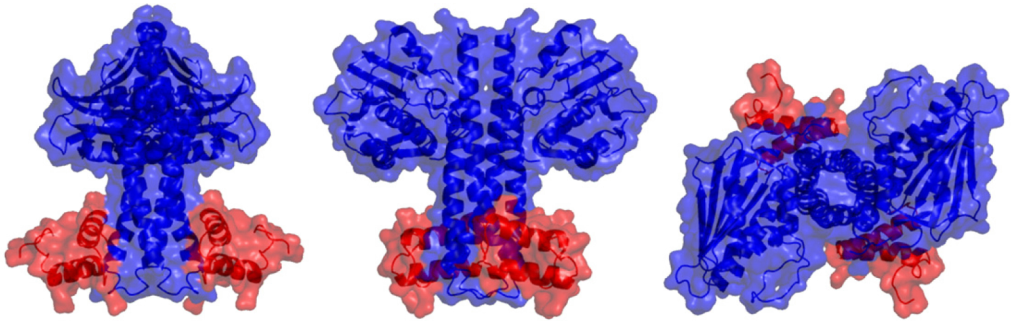
# Dummy-atom modeling

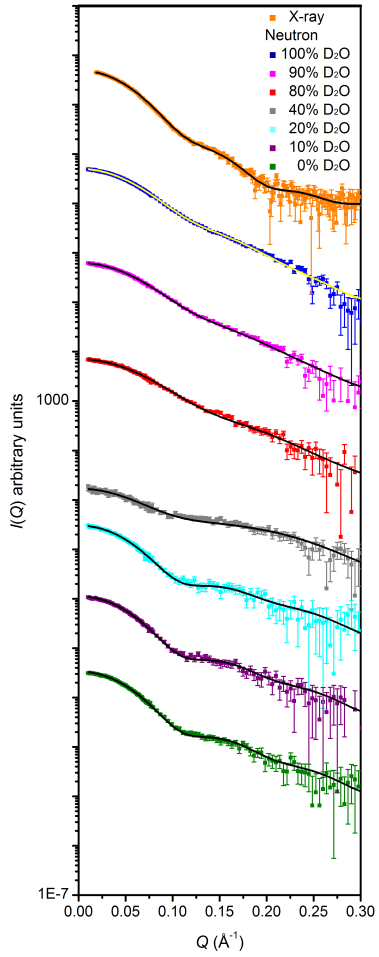
- MONSA



# Rigid body modeling

- SASREF7

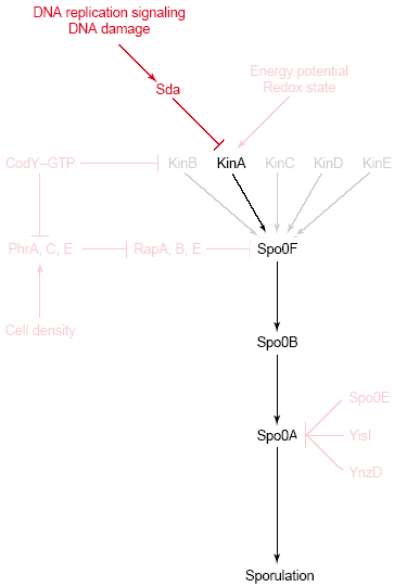




- Theory
- Experimental design
- Data collection
- Data analysis and modeling
- Examples

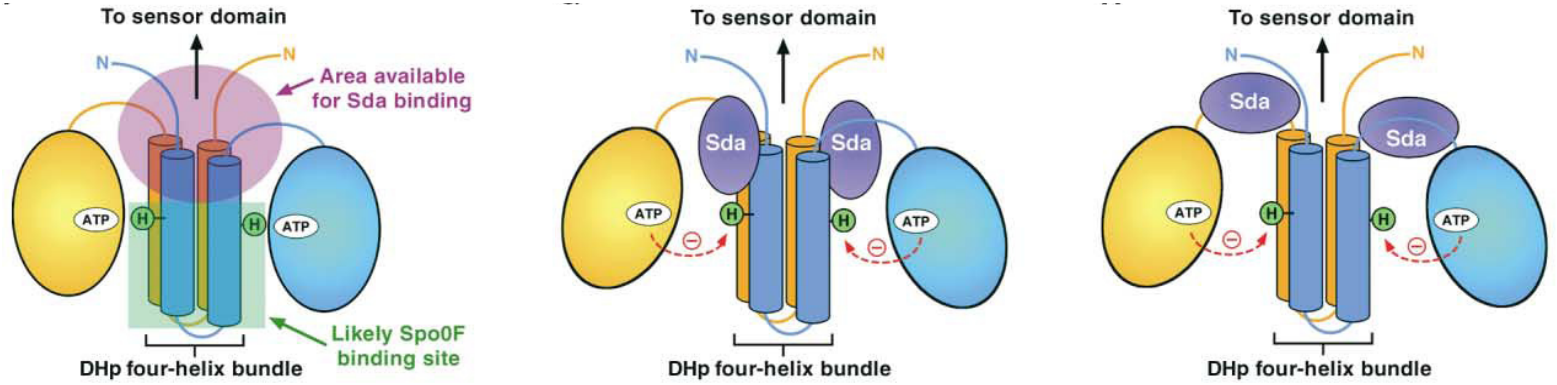
# Example: KinA:Sda

- Histidine kinases are involved in many signalling pathways in bacteria
  - Cell cycle
  - Metabolism
  - Chemotaxis
  - Antibiotic resistance

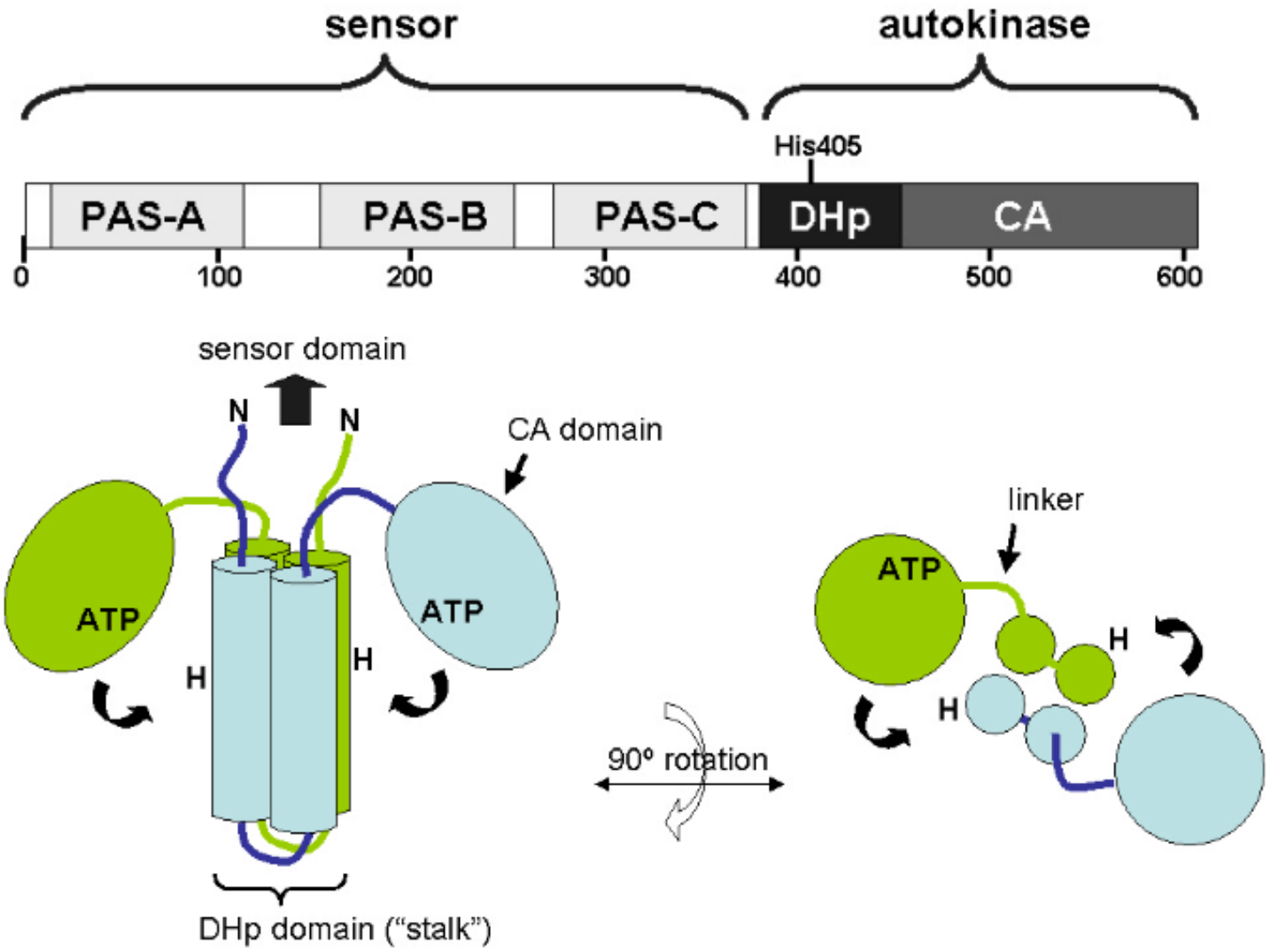


# Proposed mechanism

- The proposed binding region is a the top of the stalk

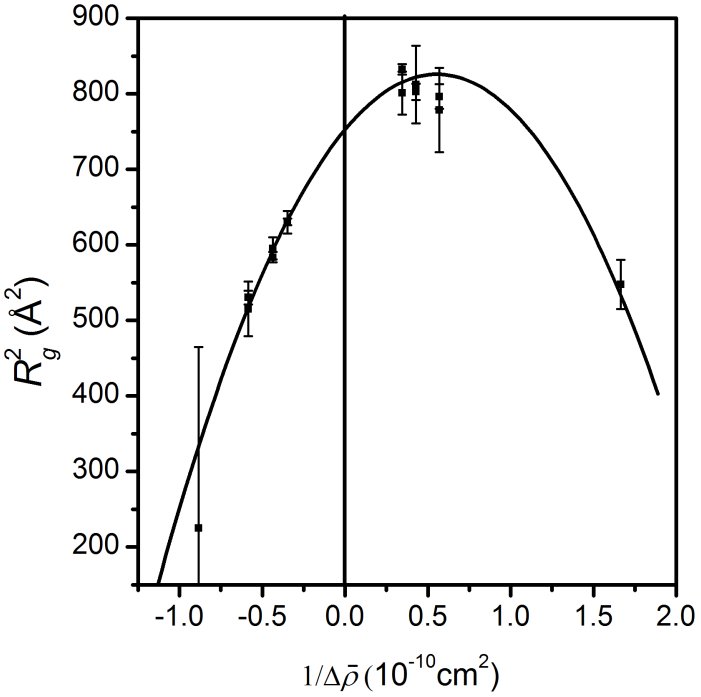


# Histidine Kinases

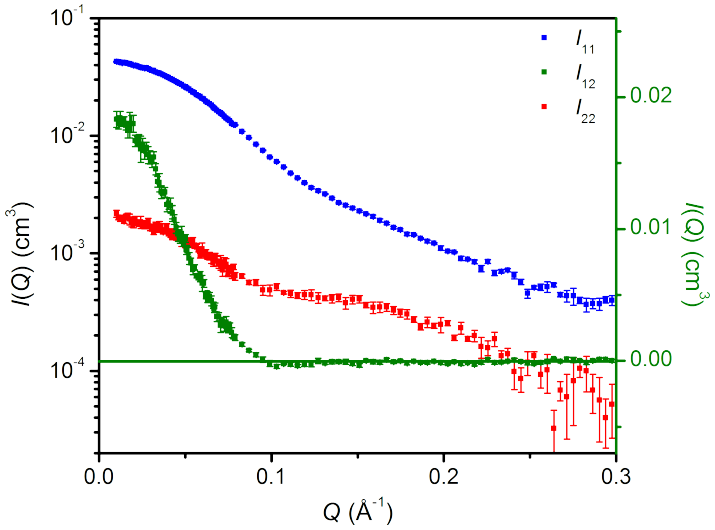


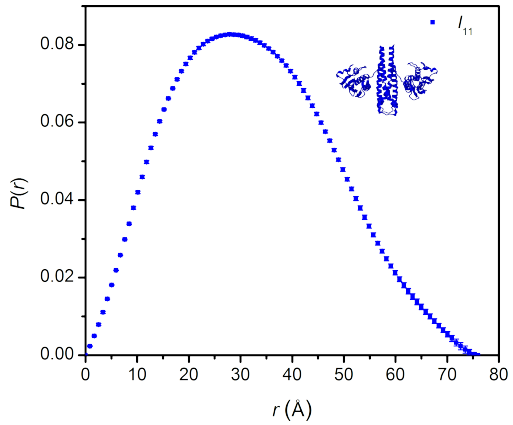
# Rg analyses

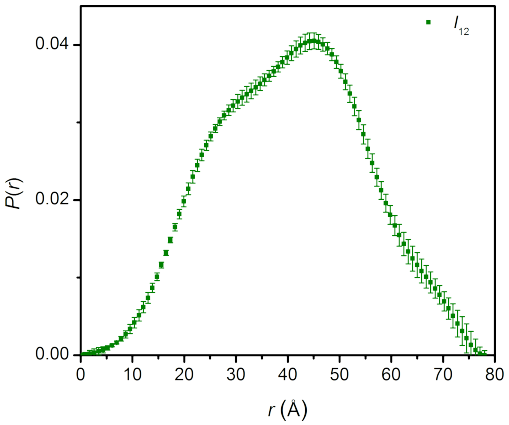
- Stuhrmann plot
  - Apex on RHS of y-axis means the deuterated component is on the outer of the molecule
- Parallel axis theorem
  - RH = 25.5 Å
  - RD = 25.4 Å
  - D = 26.6 Å

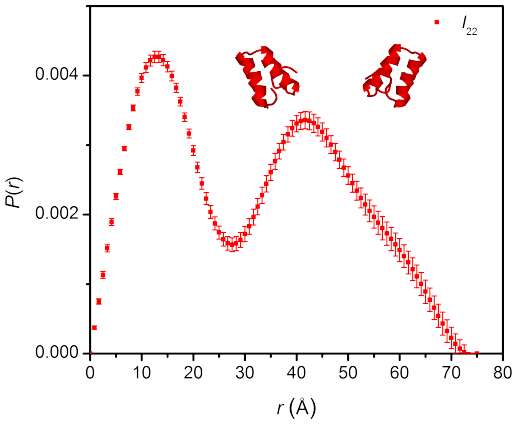


# Composite scattering functions

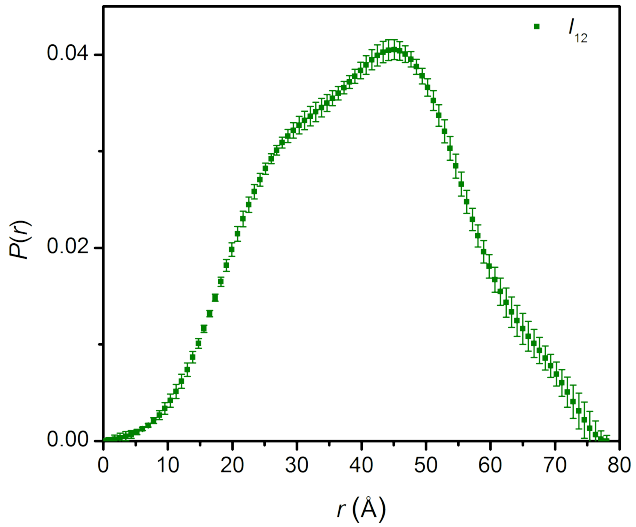


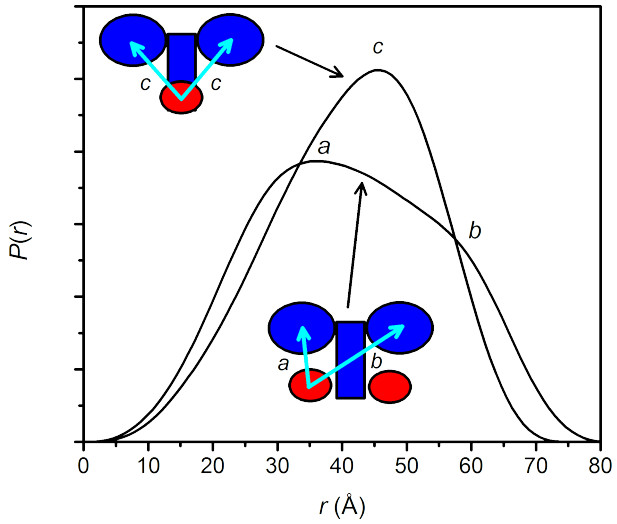



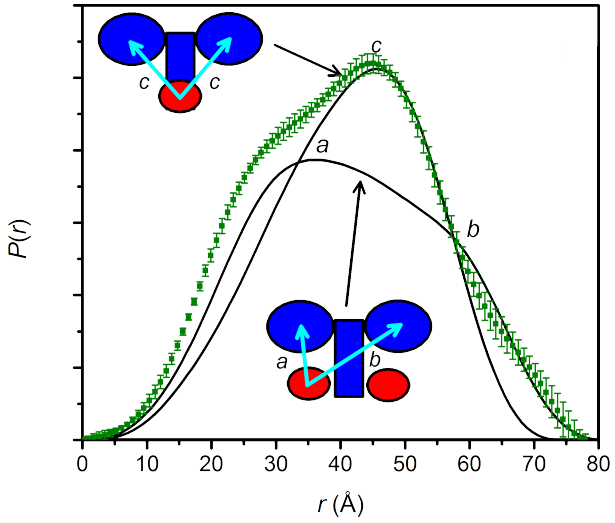



# The cross-term

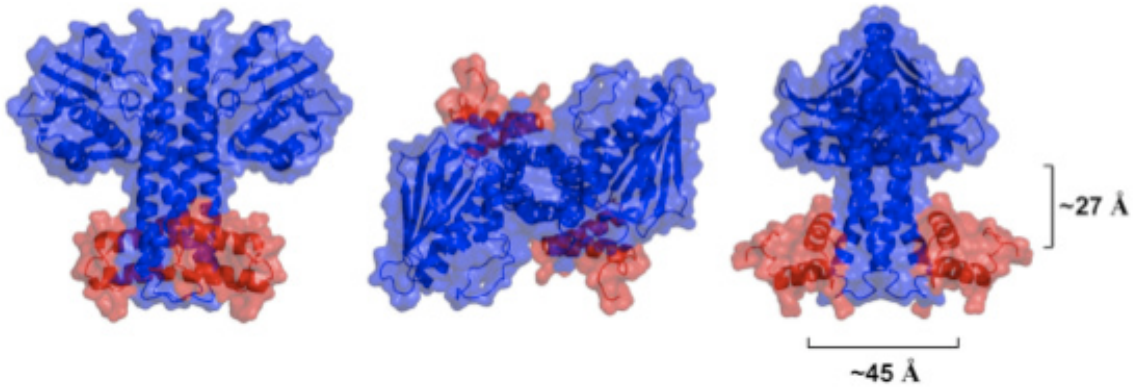


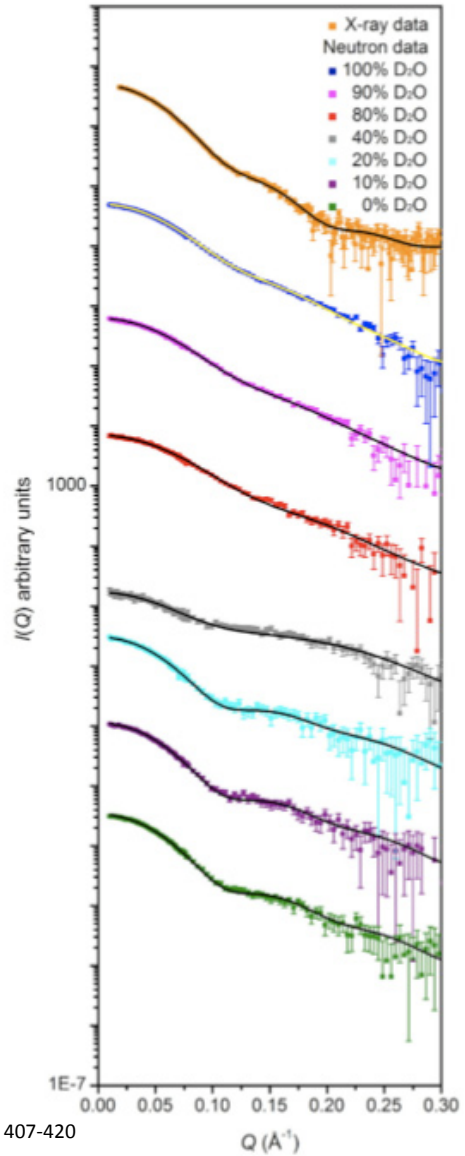



# Rigid Body modeling

- Modeling performed with SASF
- Excellent fit to the data




A.E. Whitten, D.A. Jacques, B. Hammouda, T. Hanley, G.F. King, J.M. Guss, J. Trewhella, D.B.Langley, *Journal of Molecular Biology* (2007), 368, 407-420

# Summary

- Neutron contrast variation experiments were able to determine that:
  - Two  molecules of Sda bound opposite to each other on the stalk of the the KinA dimer
- Neutron contrast variation experiments were unable to determine
  - Exactly how and where the Sda molecules bound to the KinA dimer

Part Four

# Xplor-NIH Structure Determination Including SAXS/SANS Data

Charles Schwieters

Center for Information Technology
National Institutes of Health

Bethesda, MD USA

Charles@Schwieters.org

August 21, 2010

# outline

1. Efficiently calculating solution scattering curves.
   - Calculating the bound-solvent scattering contribution.
2. Refinement against solution scattering data.
   - Overview of structure calculation
   - Overview of Xplor-NIH, and available refinement facilities
3. Overview of an Xplor-NIH Python script
4. Use of scattering data in an Xplor-NIH script

# Xplor-NIH Major Contributors

| | |
|---|---|
| Yaroslav Ryabov | Robin Thottungal |
| Marius Clore | John Kuszewski |
| Nico Tjandra | Guillermo Bermejo |
| Kirsten Frank | |

Support: Andy Byrd, Yun-Xing Wang, Ad Bax

developed in the Imaging Sciences Laboratory, DCB, CIT, NIH.

---

Those whose input contributed to SAXS/SANS capabilities:

Alex Grishaev, David Tiede, Yun-Xing Wang, Xiaobing Zuo

CIT
Center for Information Technology


# Solution Scattering Intensity

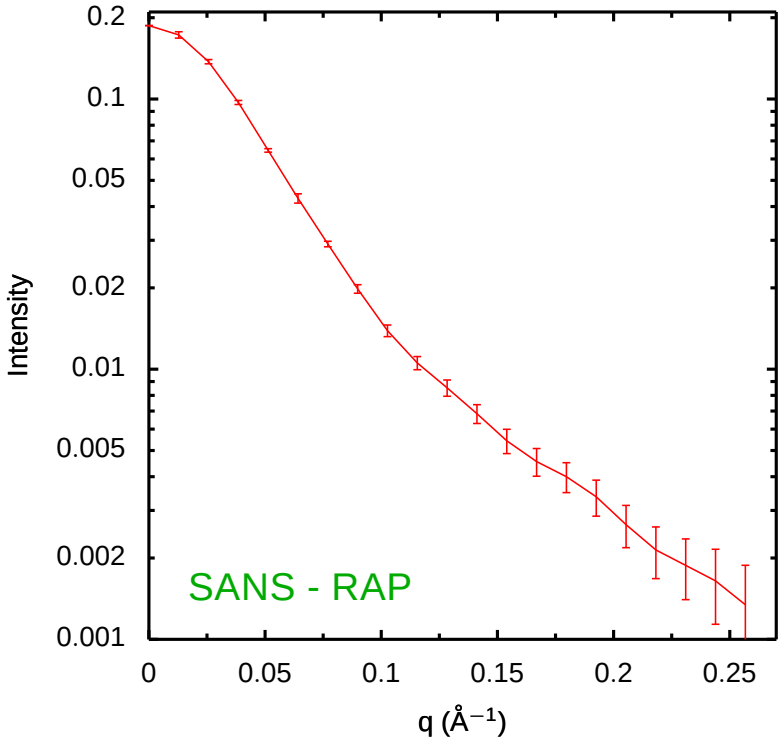
types of experiments:

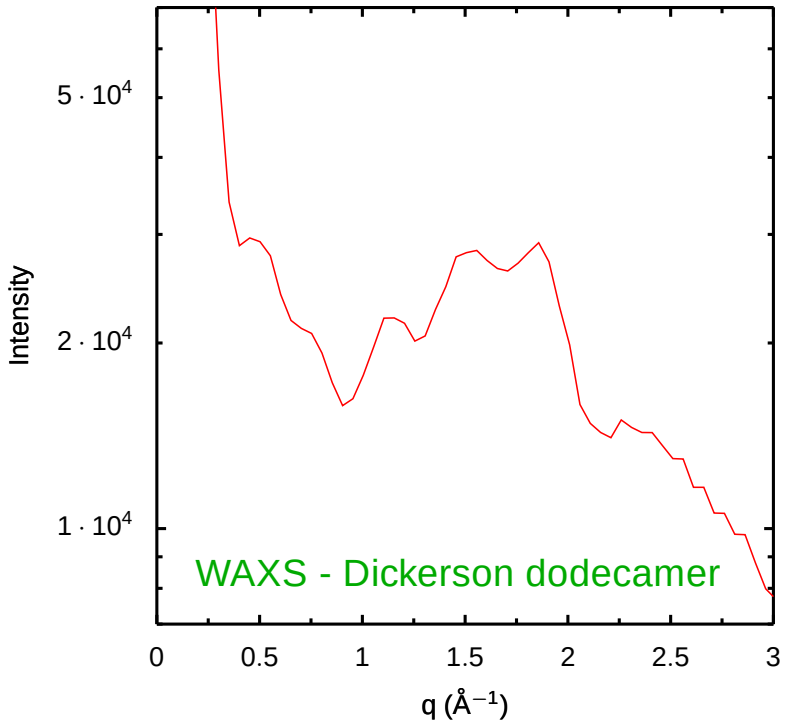
- small-angle X-ray scattering (SAXS)
- wide (or large) angle X-ray scattering (WAXS)
- Neutron scattering (SANS)

Provides information on overall molecular shape, size

$\rightarrow$ complementary to short-range information available from NMR

Example Spectra:



# Calculating Scattering Intensity

Sum over all atoms: point-source scatterers

$$A(\mathbf{q}) = \sum_j f_j^{\text{eff}}(q) e^{i\mathbf{q}\cdot\mathbf{r}_j},$$

scattering vector amplitude: $q = 4\pi \sin(\theta)/\lambda$

$\theta = 0$ is the forward scattering direction

effective atomic scattering amplitude: $f_j^{\text{eff}}(q) = f_j(q) - \rho_s g_j(q)$

$f_j(q)$: vacuum atomic scattering amplitude

$\rho_s g_j(q)$: contribution from excluded solvent

->boundary layer contribution can be optionally included

Difference between neutron and X-ray calculation: different $f_i^{\text{eff}}(q)$

Measured intensity

$$I(q) = \langle |A(\mathbf{q})|^2 \rangle_\Omega$$

$\langle \cdot \rangle_\Omega$: average over solid angle

Closed form solution: the Debye formula:

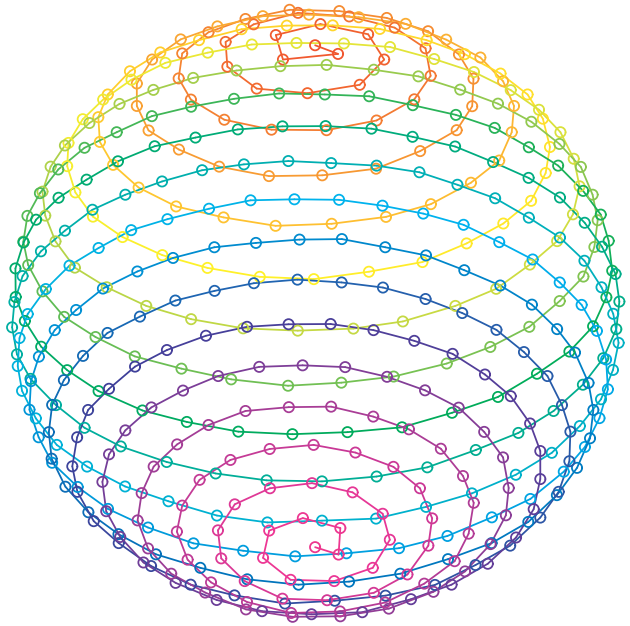
$$I(q) = \sum_{i,j} f_i^{\text{eff}}(q) f_j^{\text{eff}}(q) \text{sinc}(q r_{ij}),$$

sum is over all pairs of atoms. Expensive!

# Scattering Intensity Approximations

Instead, compute $A(\mathbf{q})$ on a sphere and integrate over solid angle numerically.

Points are selected quasi-uniformly on the sphere using the Spiral algorithm:



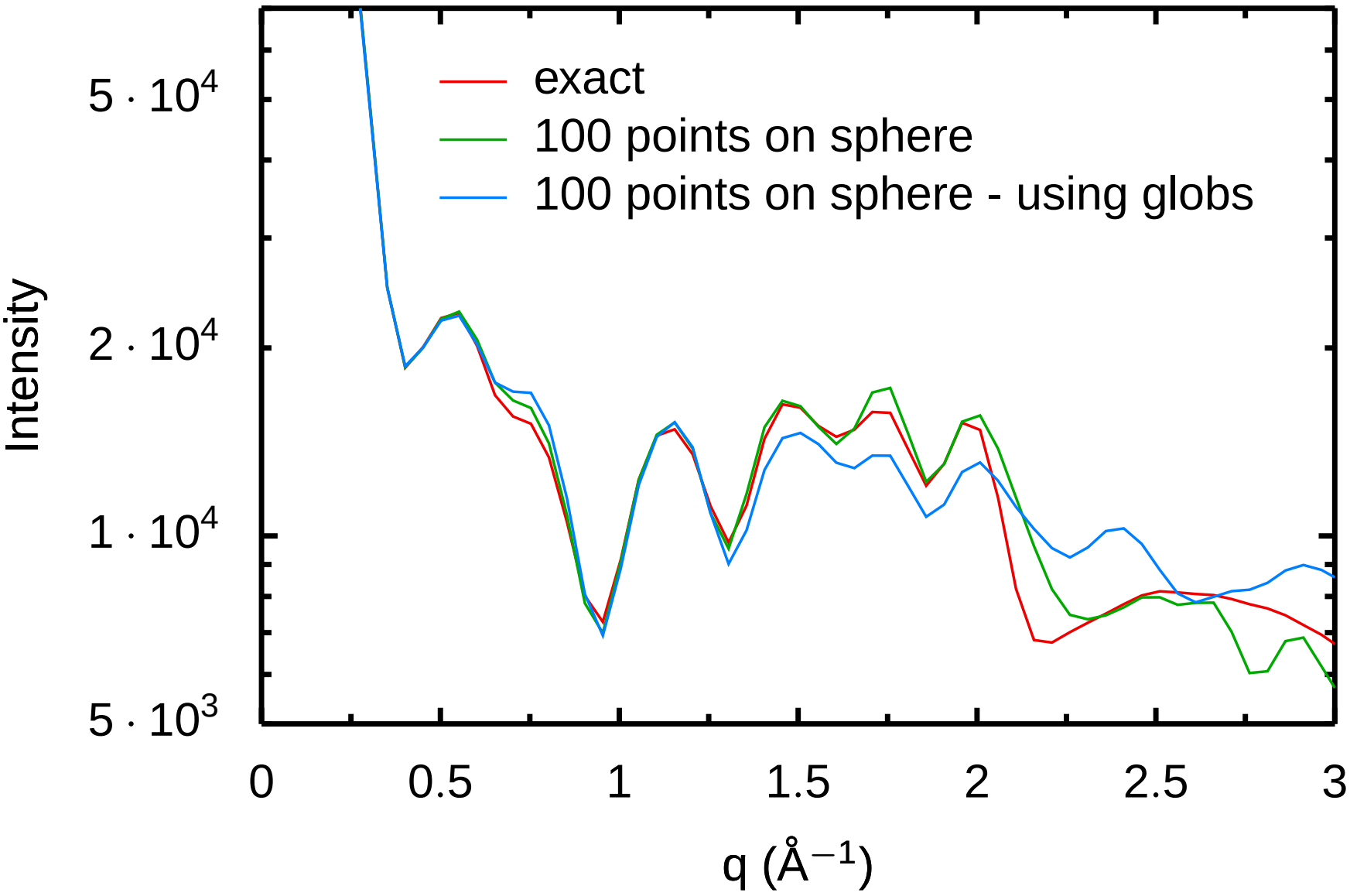
Additionally, combine atoms in "globs":

$$f_{\text{glob}}(q) = [\sum_{i,j} f_i^{\text{eff}}(q) f_j^{\text{eff}}(q) \text{sinc}(q r_{ij})]^{1/2},$$

Correct globbing, numerical integration errors with a multiplicative $q$-dependent correction factor $c_{\text{glob}}$:

$$I(q) = c_{\text{glob}}(q) I_{\text{glob}}(q),$$

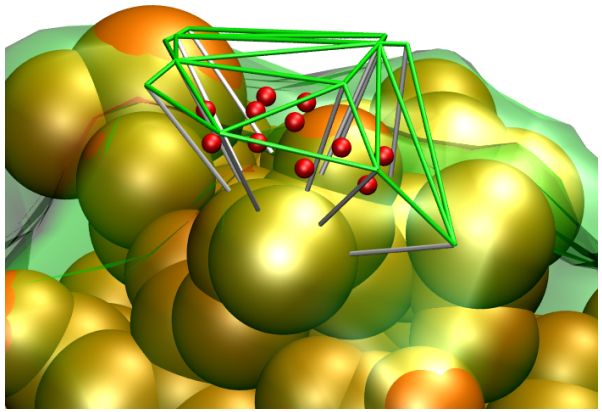
Calculated intensity for DNA scattering: numerical and globbing approximations:



# Boundary layer contribution

Bound water contributes to the scattering amplitude.

Model as a layer of uniform thickness around the molecular structure with density $\rho_b$.



- Use the Varshney[a] algorithm to efficiently generate an outer surface: roll solvent molecule over atoms whose radii are increased by $r_b$.
- Inner surface is generated using the points and surface normals.
- Each voxel defined by the tesselization procedure contributes to the scattering amplitude:

$$\sum_k f^{\mathrm{sph}}(q; r_k) e^{i\mathbf{q}\cdot\mathbf{y}_k},$$

with

$$f^{\mathrm{sph}}(q; r_k) = \rho_b 4\pi/q^2[\sin(qr_k)/q - r_k \cos(qr_k)]$$

[a] A. Varshney, F.P. Brooks, W.V. Wright, *IEEE Comp. Graphics App.* **14**, 19-25 (1994)

# Determining Solvent Scattering Parameters

as in Crysol[a] three parameters are fit

Effective atomic scattering amplitude:

$$f_j^{\text{eff}}(q) = f_j(q) - \rho_s g_j(q),$$

$f_j(q)$: vacuum atomic scattering amplitude

$\rho_s$: bulk solvent electron density

amplitude due to excluded solvent:

$$g_j(q) = s_V V_j \exp(-\pi q^2 V_j^{2/3}) \times \exp[-\pi (q r_m)^2 (4\pi/3)^{2/3} ({s_r}^2 - 1)]$$

$V_j$: atomic volume

$r_m$: is the radius corresponding to the average atomic volume

$s_V, s_r$: scale factors to be fit.

Bound solvent scattering amplitude

$$f^{\text{sph}}(q; r_k) = \rho_b 4\pi/q^2 [\sin(q r_k)/q - r_k \cos(q r_k)]$$

$\rho_b$: boundary layer electron density

$r_k$: radius corresponding to voxel volume.

three parameters are fit using a grid search.

For SANS: one additional parameter: isotropic background added to calculated $I(q)$.

[a] D. Svergun, C. Barberato and M.H.J. Koch, *J. Appl. Cryst.* **28**, 768-773 (1995).

# Refinement against solution scattering data

Refinement target function

$$E_{\text{scat}} = w_{\text{scat}} \sum_j \omega_j (I(q_j) - I^{\text{obs}}(q_j))^2,$$

$w_{\text{scat}}$, $\omega_j$: weight factors

Typically set $\omega_j = 1/\Delta I^{\text{obs}}(q_j)^2$ - inverse square of error

Normalization of $I(q)$: minimize $E_{\text{scat}}$.

There is no need to extrapolate to $q = 0$.

Efficient computation of $I(q)$ requires uniform spacing in $q$: linear interpolation is used.

Periodically (not every timestep) solvent fit parameters are recalculated.

Choice of $w_{\text{scat}}$: system/size dependent. Need for tuning.

When docking: use $R_g$ potential term to avoid local minima.

# Example Xplor-NIH SAXS setup

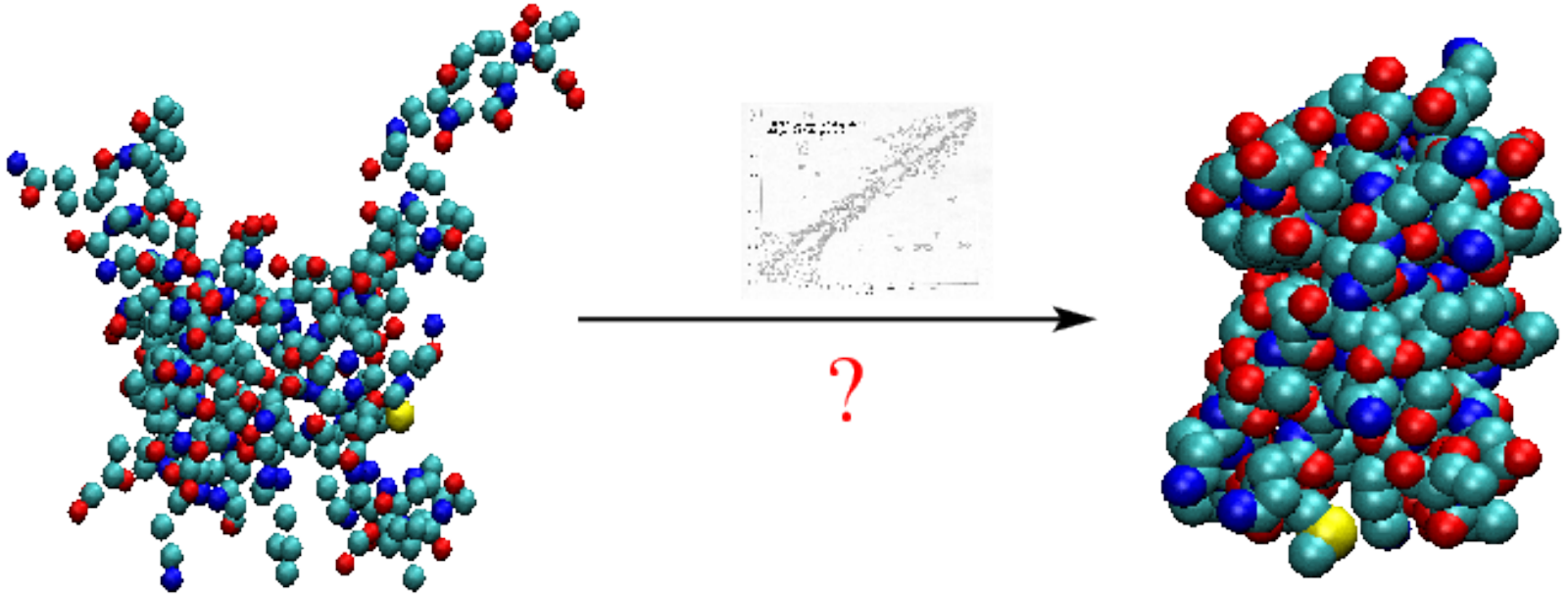
```
from solnXRayPotTools import create_solnXRayPot
import solnXRayPotTools
xray=create_solnXRayPot('xray',
                        experiment='saxs.dat',
                        numPoints=26,
                        normalizeIndex=-3,preweighted=False)

xrayCorrect=create_solnXRayPot('xray-c',
                               experiment=saxs.dat',
                               numPoints=26,
                               normalizeIndex=-3,preweighted=False)

solnXRayPotTools.useGlobs(xray)
xray.setNumAngles(50)
xrayCorrect.setNumAngles(500)
potList.append(xray)
crossTerms.append(xrayCorrect)

#corrects I(q) for globbing, small angular grid and
#  includes solvent contribution corrections
from solnScatPotTools import fitParams
rampedParams.append( StaticRamp("fitParams(xrayCorrect)") )
rampedParams.append(StaticRamp("xray.calcGlobCorrect(xrayCorrect.calcd())"))
```

# Example Xplor-NIH SANS setup

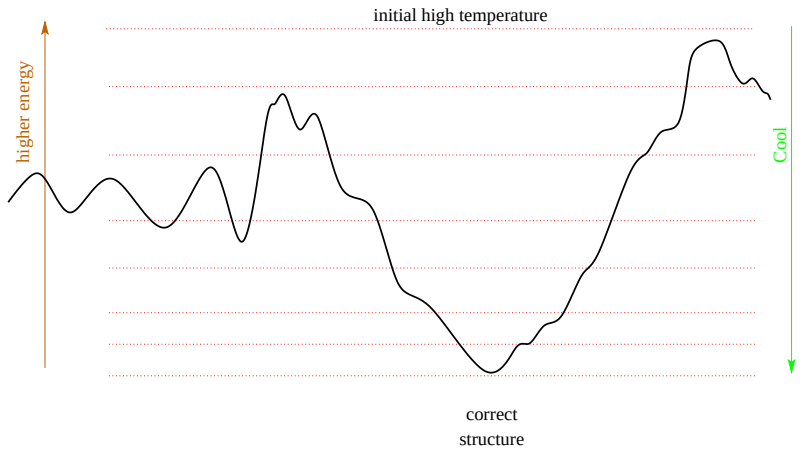
bound-solvent contribution frequently much less important

```
from sansPotTools import create_SANSPot
import sansPotTools
sans=create_SANSPot('sans',
                    experiment='sans.dat',
                    numPoints=20,
                    fractionD2O=0.41,
                    fractionDeuterated=1.,
                    altDeuteratedSels=[("resid 601:685",0.)],
                    cmpType="plain",
                    normalizeIndex=-3,preweighted=False)

sansPotTools.useGlobs(sans)

sans.setNumAngles(80)
sans.setScale(40)
potList.append(sans)

#correct using the Debye equation
rampedParams.append( StaticRamp("sans.calcGlobCorrect('n2')") )
```

# Overview of structure determination



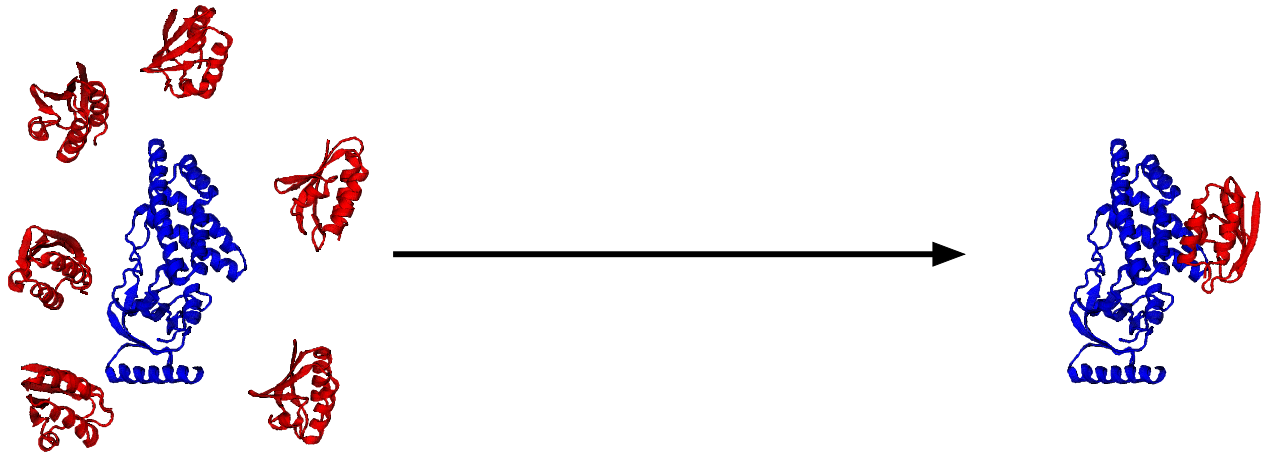

Minimize energy: $V_{\text{tot}} = V_{\text{noe}} + V_{\text{bond}} + V_{\text{SAXS}} + \ldots$




- molecular dynamics to explore the energy surface.
- slowly decrease the temperature to find the global minimum.
- surface smoothed at high temperature

# Types of calculations

- refinement of close initial structure
- structure determination from random coordinates (torsion angles)
- docking of subunits with known structures



# Structure Calculation Overview: Script Skeleton

```
import protocol
protocol.loadPDB("model.pdb")     #initialize coordinates

coolParams=[] # a list which specifies potential smoothing
# set up potential terms from NMR experiments, covalent geometry,
# and knowledge-based terms

# initialize coolParams for annealing protocol for each energy term

from ivm import IVM   #configure which degrees of freedom to optimize
dyn = IVM()

from simulationTools import AnnealIVM
coolLoop=AnnealIVM(dyn,...)      #create simulated annealing object, specify temperature schedule

def calcOneStructure( structData ):
    """ a function to calculate a single structure """
    # [ randomize velocities ]
    # [ perform high temp dynamics ]
    dyn.run()
    # [ cooling loop ]
    coolLoop.run()
    # [ final minimization ]
    dyn.run()
    structData.writeStructure(potList)  #write out pdb record to file
                                        # with energies, rmsd's in headers
                                        # a separate .viols file also written

from simulationTools import StructureLoop
StructureLoop(numStructures=100,                        #calculate all structures
              pdbTemplate='SCRIPT_STRUCTURE.sa',        #  in parallel, if desired
              structLoopAction=calcOneStructure).run() #also report stats at end
```

# Loading and Generating Coordinates

PSF file - contains atomic connectivity, mass and covalent geometry information.

This information must be present before coordinates can be loaded.

generate via external helper scripts

1. seq2psf - generate a psf file from primary sequence

   ```
   % seq2psf file.seq
   ```

2. pdb2psf - generate a psf file from a pdb file

   ```
   % pdb2psf file.pdb
   ```

within the Python scripting interface (in the `protocol` module)

- `protocol.initStruct` - load pregenerated .psf file
- `protocol.initCoords` - read pdb file
- `protocol.loadPDB` - read pdb and generate psf info on the fly

# Using potential terms in Xplor-NIH

Available potential terms in the following modules:

- noePot - NOE distance restraints
- rdcPot - dipolar coupling
- csaPot - Chemical Shift Anisotropy
- cstMagPot - refine against chemical shift tensor magnitudes
- jCoupPot - $^3$J-coupling
- prePot - Paramagnetic relaxation enhancement
- gyrPot - pseudopotential enforcing correct protein density
- diffPot - refine against rotational diffusion tensor
- relaxRatioPot - refine directly against ratios of relaxation times
- residueAffPot - contact potential for hydrophobic attraction/repulsion
- planeDistPot - distance between atoms and plane
- xplorPot - use XPLOR potential terms
  - refinement against X-ray crystall fiber diffraction data
  - database terms for hydrogen bonds, torsion angles
- solnScatPot - potential for solution X-ray and neutron scattering
- posSymmPot - restrain atomic positions relative to those in a similar structure
- potList - a collection of potential terms in a list-like object.

# The IVM (internal variable module) for dynamics and minimization

Dynamics/Minimization in Cartesian, torsion-angle and other coordinates.

- automatic choice of MD stepsize.
- don't have to worry about messing up known coordinates.
- topological loops must be treated carefully.
- facility to constrain bonds which cause loops.

Xplor-NIH implementation: C.D. Schwieters and G.M. Clore; J. Magn. Reson. 152, 288-302 (2001).

Dynamics with variable timestep

```
import protocol
bathTemp=2000
protocol.initDynamics(ivm=integrator,      #note: keyword arguments
                      bathTemp=bathTemp,
                      finalTime=1,         # use variable timestep
                      printInterval=10,    # print info every ten steps
                      potList=pots)

integrator.run()          #perform dynamics
```

# Topology Setup

torsion angle dynamics with fixed region:

```
from ivm import IVM
integrator = IVM()                                  #create an IVM object
integrator.fix( AtomSel("resid 100:120") )          # these atoms are fixed in space
integrator.group( AtomSel("resid 130:140") )        # fix relative to each other,
                                                    # but translate, rotate in space

from protocol import torsionTopology
torsionTopology(integrator)                         # group rigid side chain regions
                                                    # break proline rings
                                                    # group and setup all remaining
                                                    #  degrees of freedom for
                                                    #  torsion angle dynamics
                                                    #
                                                    #  topology setup of pseudoatoms
                                                    #  e.g. alignment tensor atoms:
                                                    #  - tensor axis should rotate
                                                    #    only - not translate.
                                                    #  - only single dof of Da and Rh
                                                    #    parameter atoms is significant.
```

# High-Level Helper Classes

AnnealIVM: perform simulated annealing

```
from simulationTools import AnnealIVM
anneal= AnnealIVM(initTemp =3000,      #high initial temperature
                  finalTemp=25,        #final temperature
                  tempStep =25,        # temperature increment
                  ivm=integrator,      # ivm object used for molecular dynamics
                  rampedParams = coolParams) #list of energy parameters to scale

anneal.run()  # actually perform simulated annealing
```

force constants of some terms are geometrically scaled during refinement:

$$k_{\mathrm{NOE}} = \gamma^n k^{(0)}_{\mathrm{NOE}}$$

$\gamma^N = k^{(f)}_{\mathrm{NOE}}/k^{(0)}_{\mathrm{NOE}}$

```
from simulationTools import MultRamp    #multiplicatively ramped parameter
coolParams=[]
coolParams.append( MultRamp(2,30,       #change NOE scale factor
                            "noe.setScale( VALUE )") )
```

StructureLoop: calculate multiple structures

```
from simulationTools import StructureLoop
StructureLoop(structureNums=range(10),         # calculate 10 structures
              structLoopAction=calcStructure, # calcStructure is function
              pdbTemplate=pdbTemplate)         # template for output structures
pdbTemplate = 'SCRIPT_STRUCTURE.sa'
#SCRIPT -> replaced with the name of the input script (e.g. 'anneal.py')
#STRUCTURE -> replaced with the number of the current structure
```

`StructureLoop` also helps with analysis:

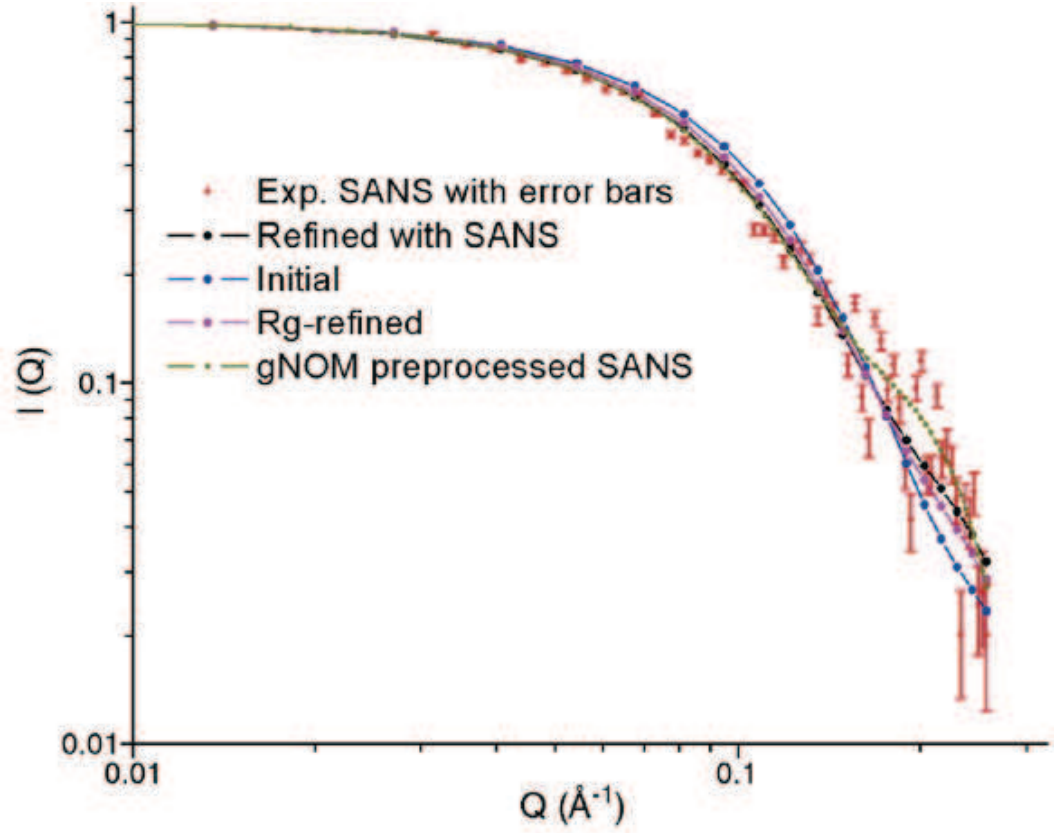
```
from simulationTools import StructureLoop, FinalParams
StructureLoop(structureNums=range(10),
              structLoopAction=calcStructure,
              pdbTemplate=outFilename,
              averageTopFraction=0.5,    # fraction of structures to use
              averageFitSel="not hydro", #atoms used for fitting structures
              averagePotList=potList,    #terms to use to compute of ave. struct
              averageContext=FinalParams(rampedParams),  #force constants used
              averageFilename="ave.pdb", #output filename
              genViolationStats=True,    # generate a .stats file with
                                         #  energy/violation/structure stats
             ).run()
```

`StructureLoop` transparently takes care of parallel structure calculation.

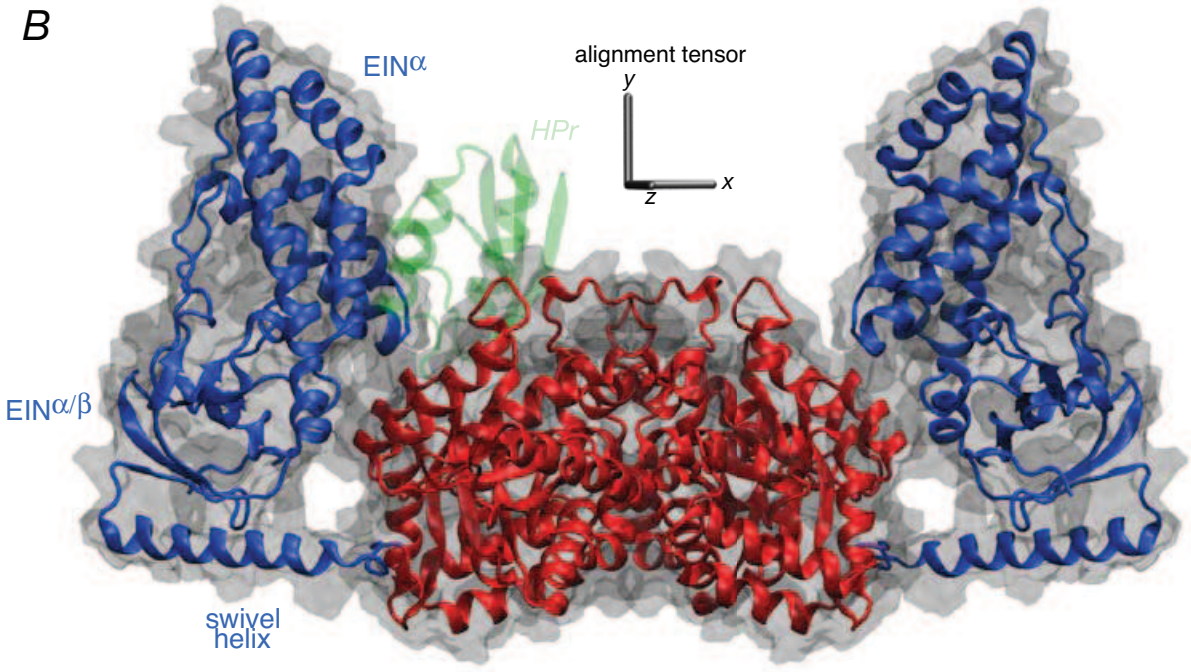
# Examples of SAXS/SANS structure determination

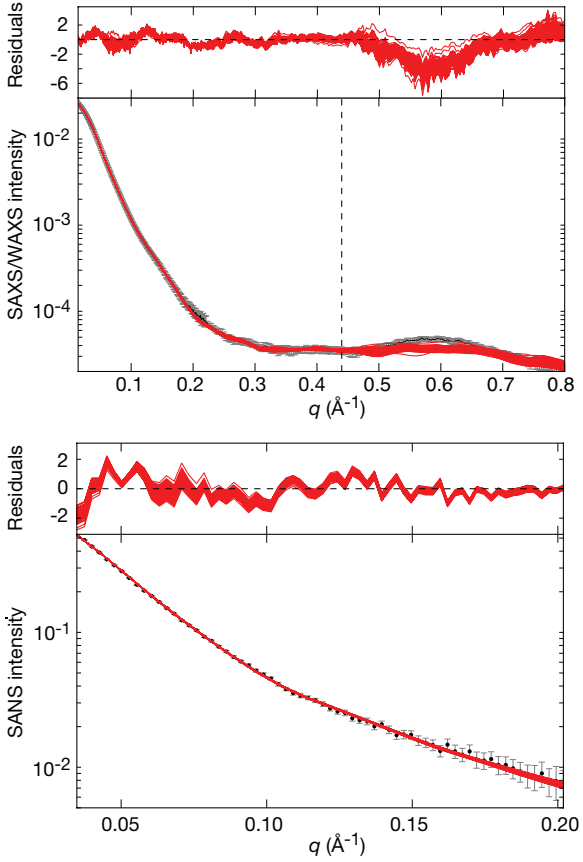
L11 Protein

refinement from NMR structure using NOE, RDC and SANS data.



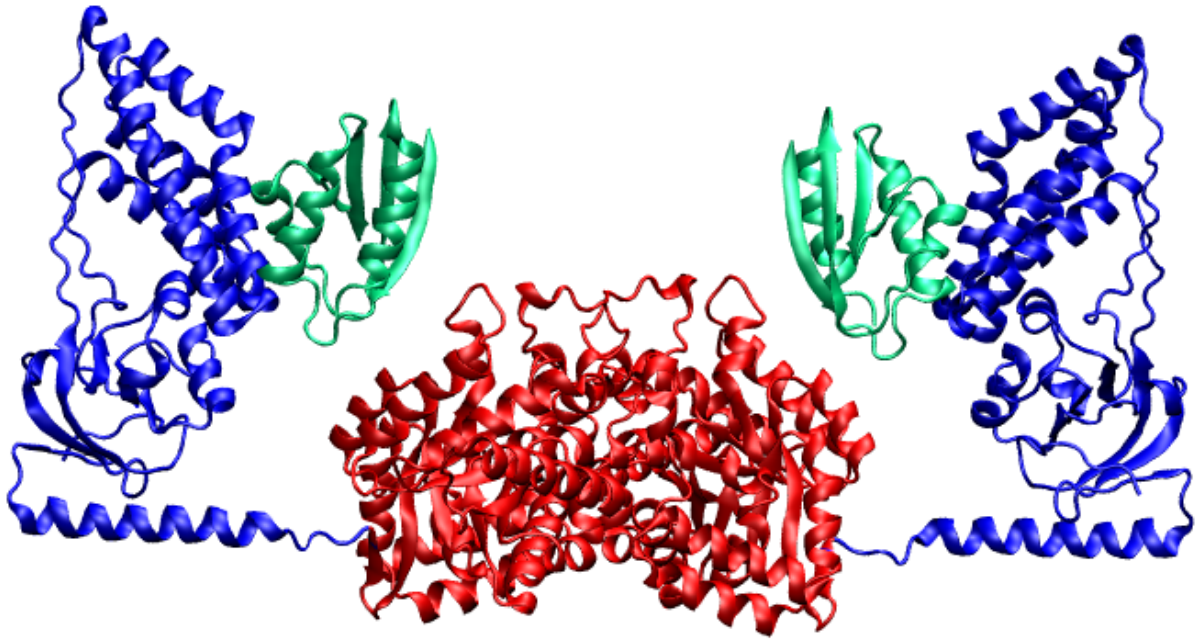
Rigid-body positioning of the N- and C-Termini based on SAXS and RDC data.

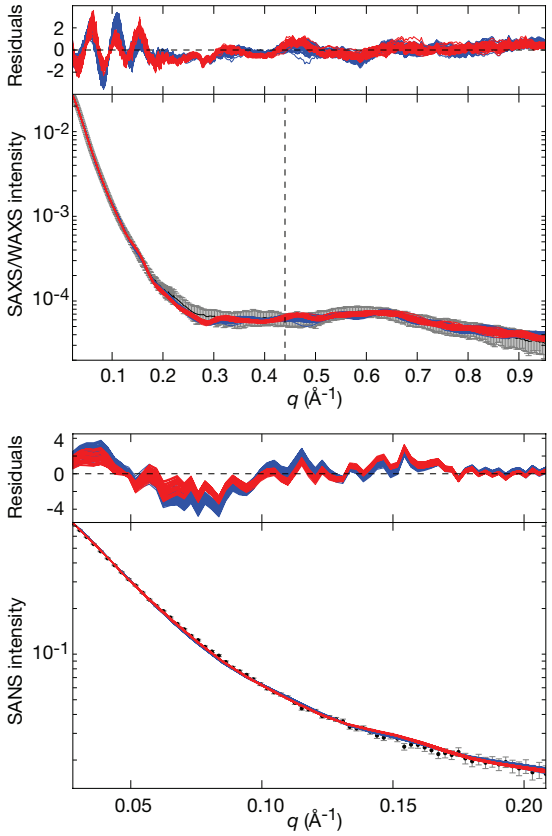


# Complex of Enzyme I / HPr (histidine phosphocarrier protein)

expansion to allow HPr

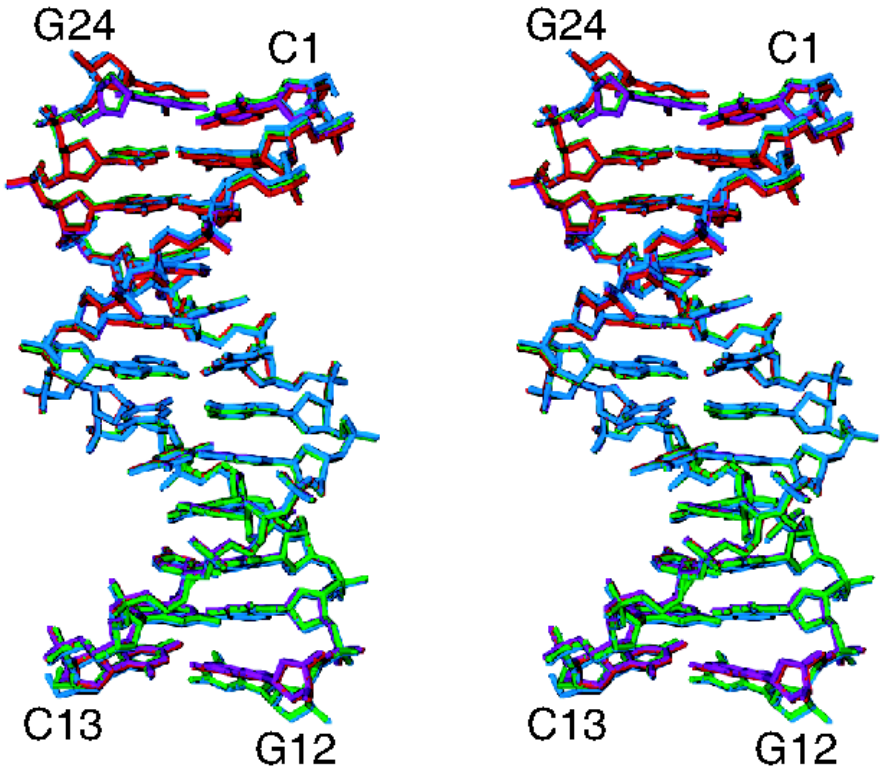




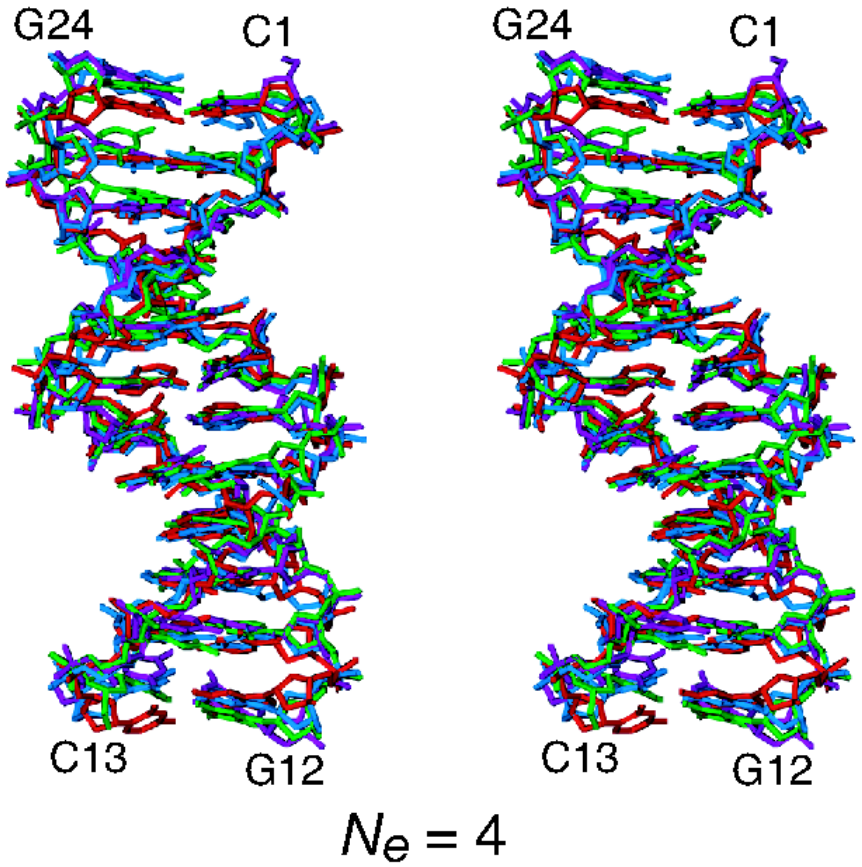
# DNA: the Dickerson Dodecamer

SAXS + RDCs, CSA, distance restraints

required the use of Xplor-NIH Ensemble-fitting facilities

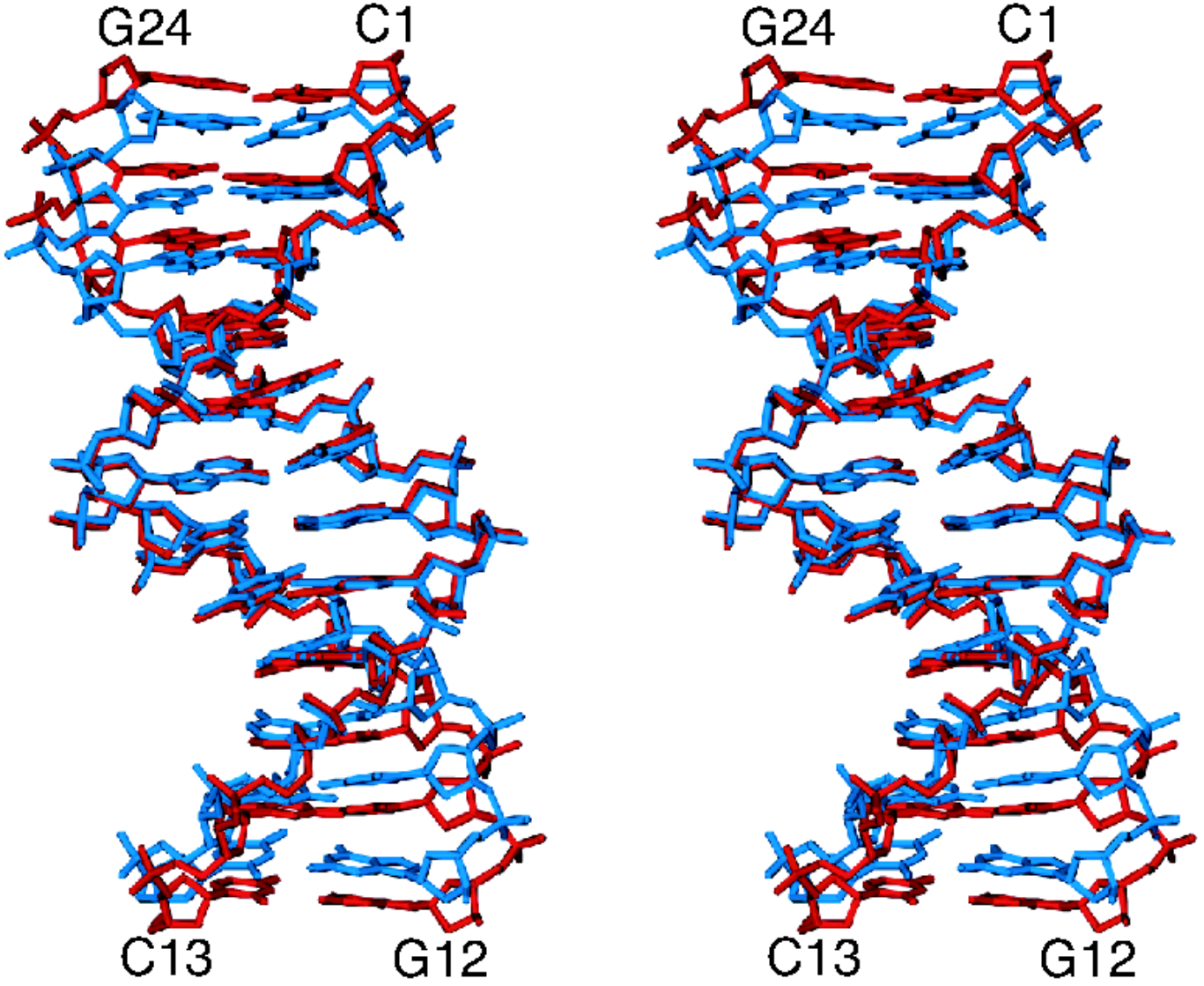



ensemble of four independent structures




4-membered ensemble refined simultaneously

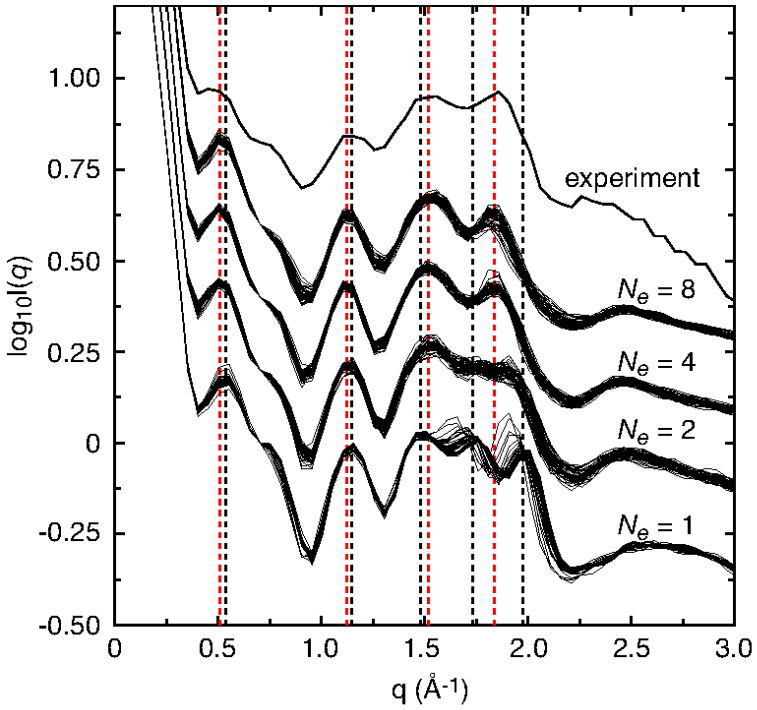
# DNA - average structure

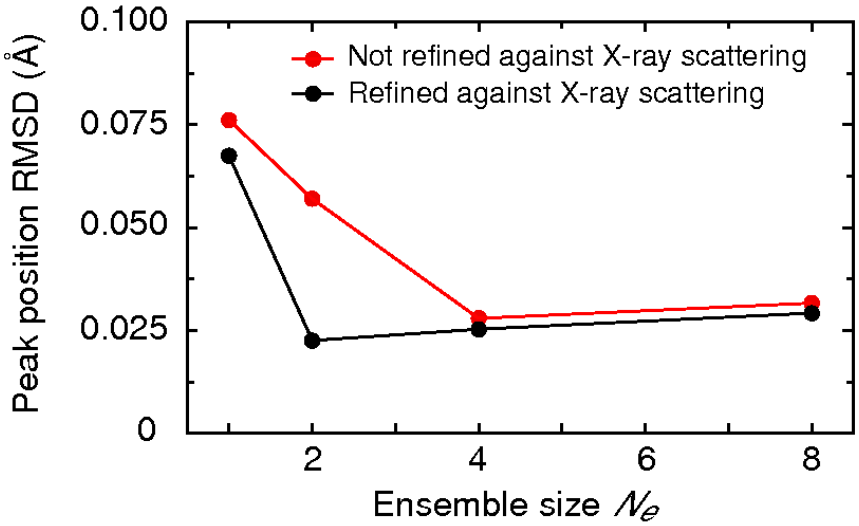



$N_e = 1$ average (blue) ; $N_e = 4$ average (red) – 1.2Å RMSD

# DNA - WAXS peak positions

peak positions incorrect if single structure representation used.



# Convenience Scripts

pdb2psf - generate a psf from a PDB file.

```
% pdb2psf 1gb1.pdb
```

creates 1gb1.psf. Please send us pdb files which fail.

seq2psf - generate a psf file from primary sequence.

```
% seq2psf -segname PROT -startresid 300 -protein protG.seq
```

creates protG.psf with segid PROT starting with residue id 300.

torsionReport - collect and average protein torsion angle values.

```
% torsionReport -psf=[psf file] [pdb files] >average.info
```

also:

aveStruct - average final structures and report per-atom RMSD to the mean

targetRMSD - report RMSD to a reference structure

pairRMSD.py - report pairwise RMSD

calcTensor.py - calculate an alignment tensor and report back-calculated RDC values given one or more structures

calcDaRh - calculate estimates of $D_a$ and rhombicity given only RDC values (no structures) - using a maximum likelihood approach.

mleFit - fit an ensemble of structures based on similarity using a maximum likelihood algorithm

domainDecompose - given an ensemble of structures, find regions of structural similarity, using maximum-likelihood fitting.

calcSAXS - given a structure, calculate a SAXS or SANS curve, optionally comparing with experiment. Can also compute optimal excluded solvent parameters (including boundary layer contribution).

findClusters - find clusters of similar structures within an ensemble.

# `calcSaxs` helper to calculate SAXS/SANS curve

calculate SAXS curve from a structure - include hydration layer.

```
calcSAXS -fit -numQ 100 -expt saxs.dat ei.pdb
```

calculate SANS curve for selectively deuterated EI-HPr

```
calcSAXS -sans  -fit \
         -fractionD2O .41 -fractionDeuterated 1 \
         -altDeuteratedSel "resid 601:685" 0 \
         -numQ 20 -expt sans.dat \
         ei-hpr.pdb
```

# Putting it together: a full script

Full script for refining protein Enzyme I with NOE, dihedral and SAXS restraints will be presented.

# Where to go for help

online:

| | | |
|---|---|---|
| `http://nmr.cit.nih.gov/xplor-nih/` | - | home page |
| xplor-nih@nmr.cit.nih.gov | - | mailing list |
| `http://nmr.cit.nih.gov/xplor-nih/faq.html` | - | FAQ |
| `http://nmr.cit.nih.gov/xplor-nih/doc/current/` | - | current Documentation, including the XPLOR manual |

subdirectories within the xplor distribution:

| | | |
|---|---|---|
| eginputs | - | newer complete example scripts |
| tutorial | - | repository of older XPLOR scripts |
| helplib | - | help files |
| helplib/faq | - | frequently asked questions |

SAXS/SANS refinement example:

eginputs/saxs

Python:

M. Lutz and D. Ascher, "Learning Python, 2nd Edition" (O'Reilly, 2004);

`http://python.org`

Please complain! and suggest!