Introduction to Biological Small Angle Scattering

Tom Grant, Ph.D.
Staff Scientist
BioXFEL Science and Technology Center
Hauptman-Woodward Institute
Buffalo, New York, USA
tgrant@hwi.buffalo.edu

SAXS Literature and Software

Reviews:

- Putnam et al, Q Rev Biophys. Aug 2007; 40(3): 191-285.
- Jacques and Trewhella, Protein Science 2010 Apr; 19(4): 642–657.
- Svergun et al, Oxford University Press 2013, *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*

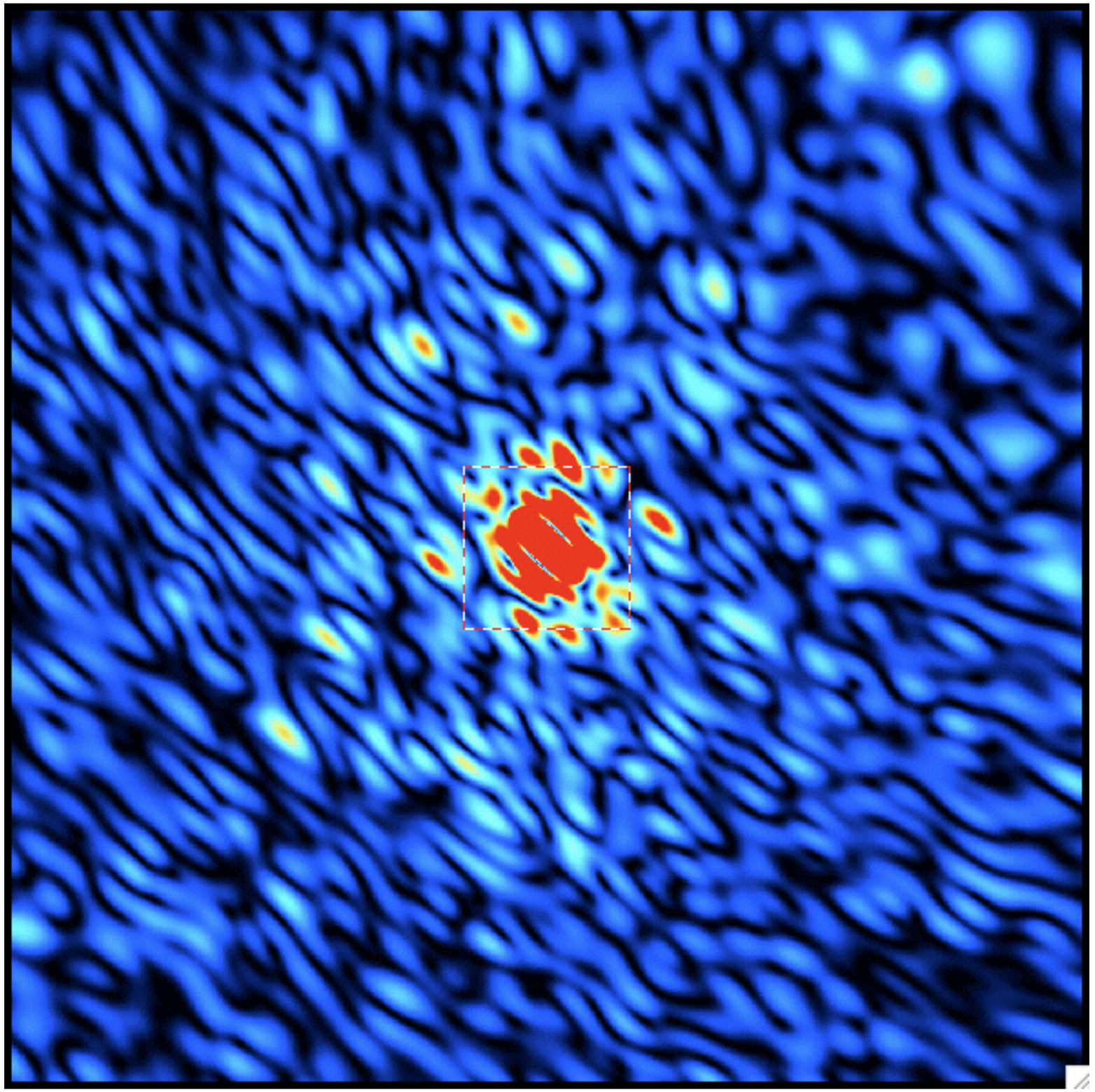
- Long list of software for SAS data analysis for biological and non-biological applications available at:

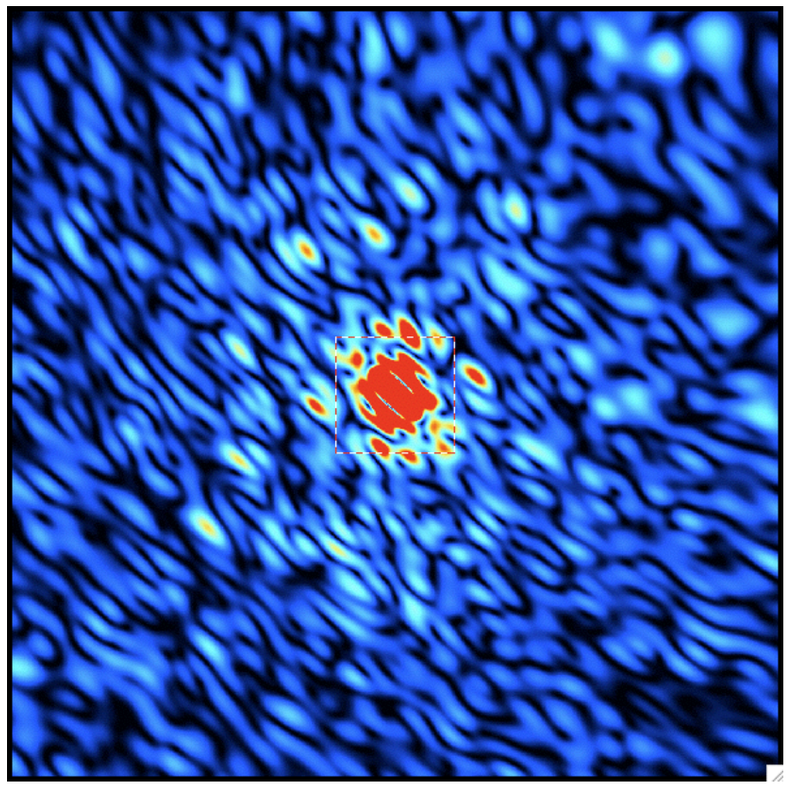
http://smallangle.org/content/software

- Most common package for analysis and modeling of biological SAS data is ATSAS, however many other excellent software packages exist

What is small angle scattering?

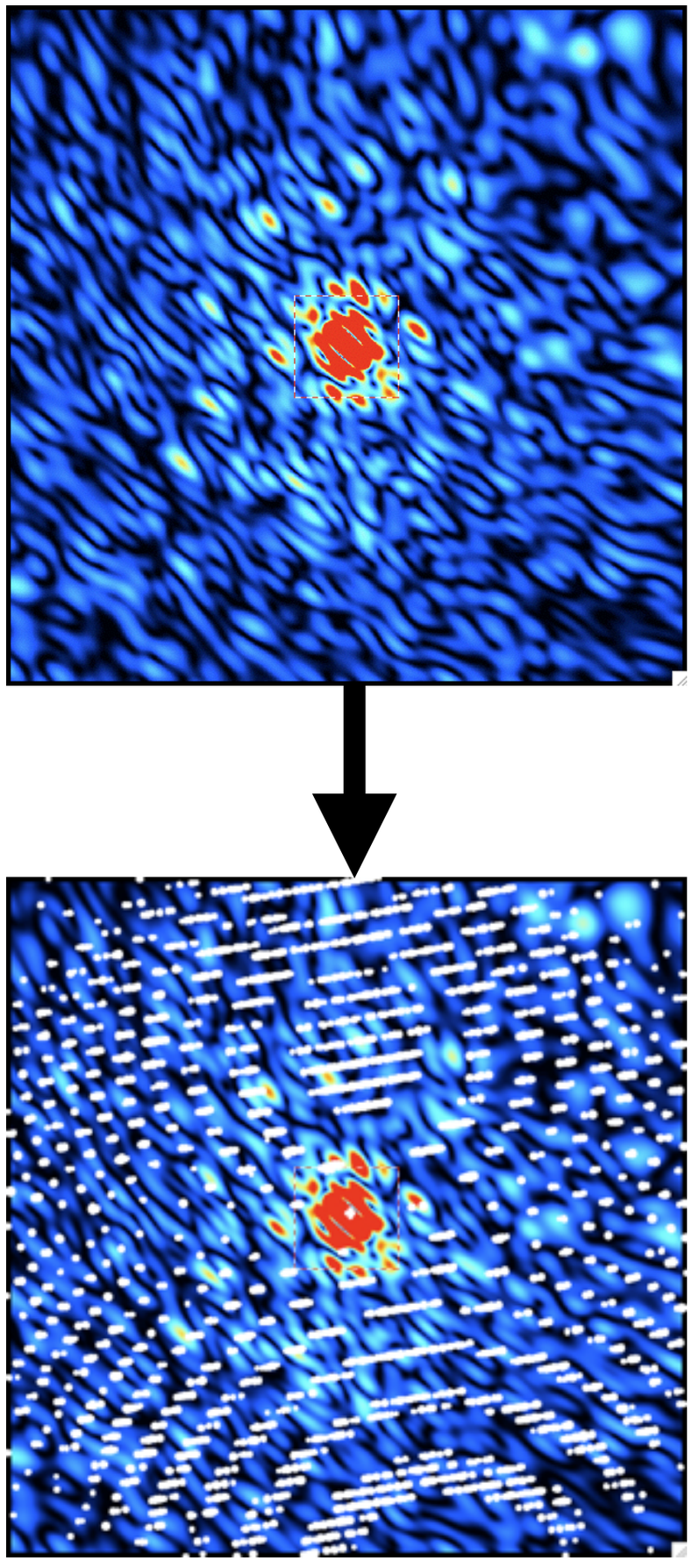
Molecular Transform





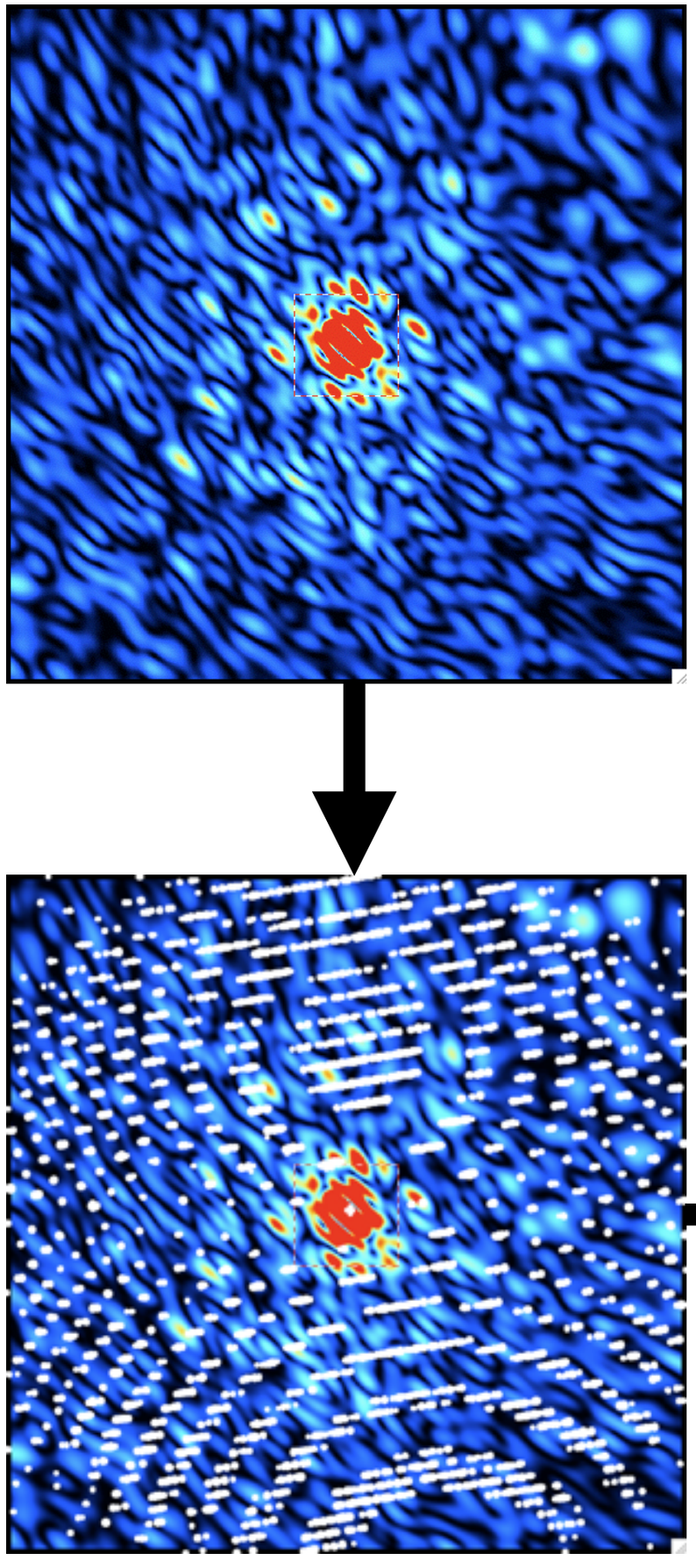
Molecular Transform

Bragg Sampling from X-ray Crystallography



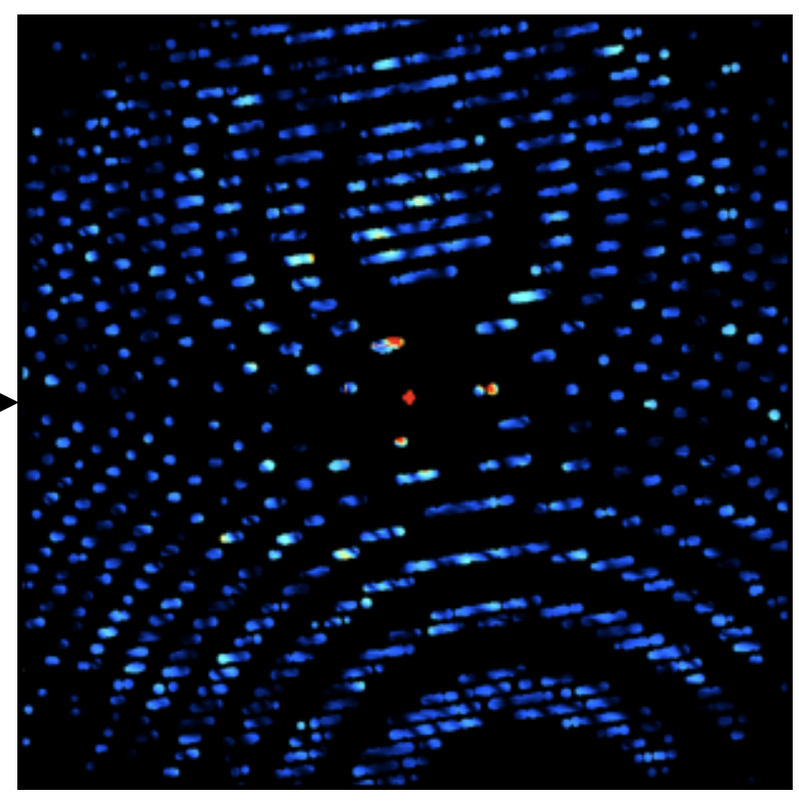
Molecular Transform

Bragg Sampling from
X-ray Crystallography

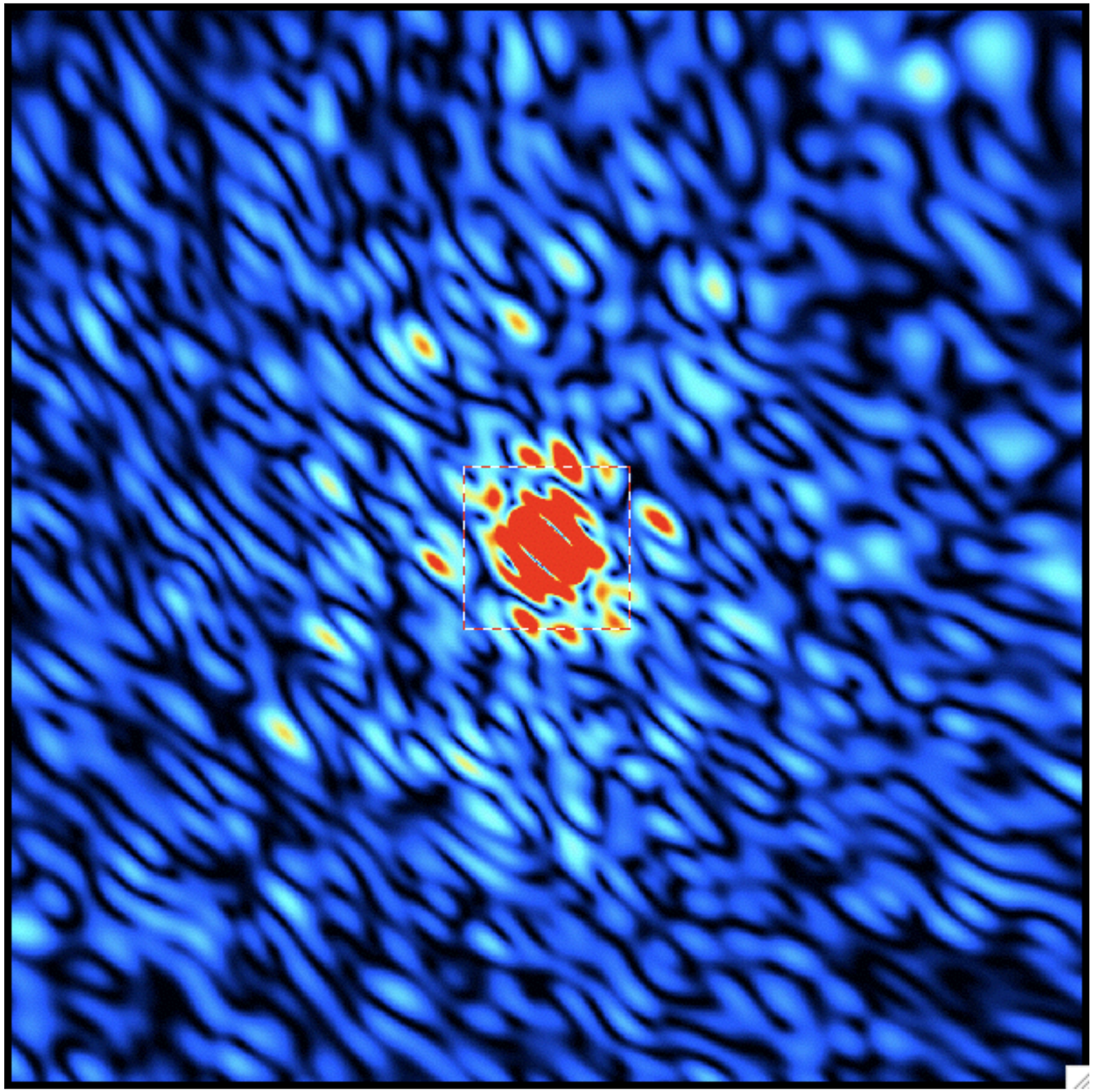


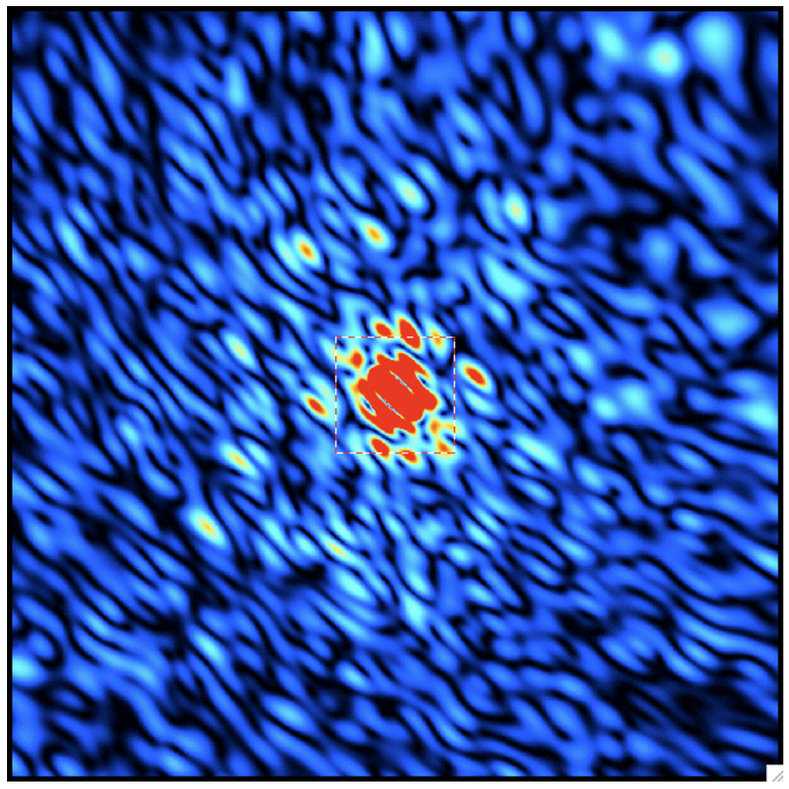
Molecular Transform

Bragg Sampling from
X-ray Crystallography

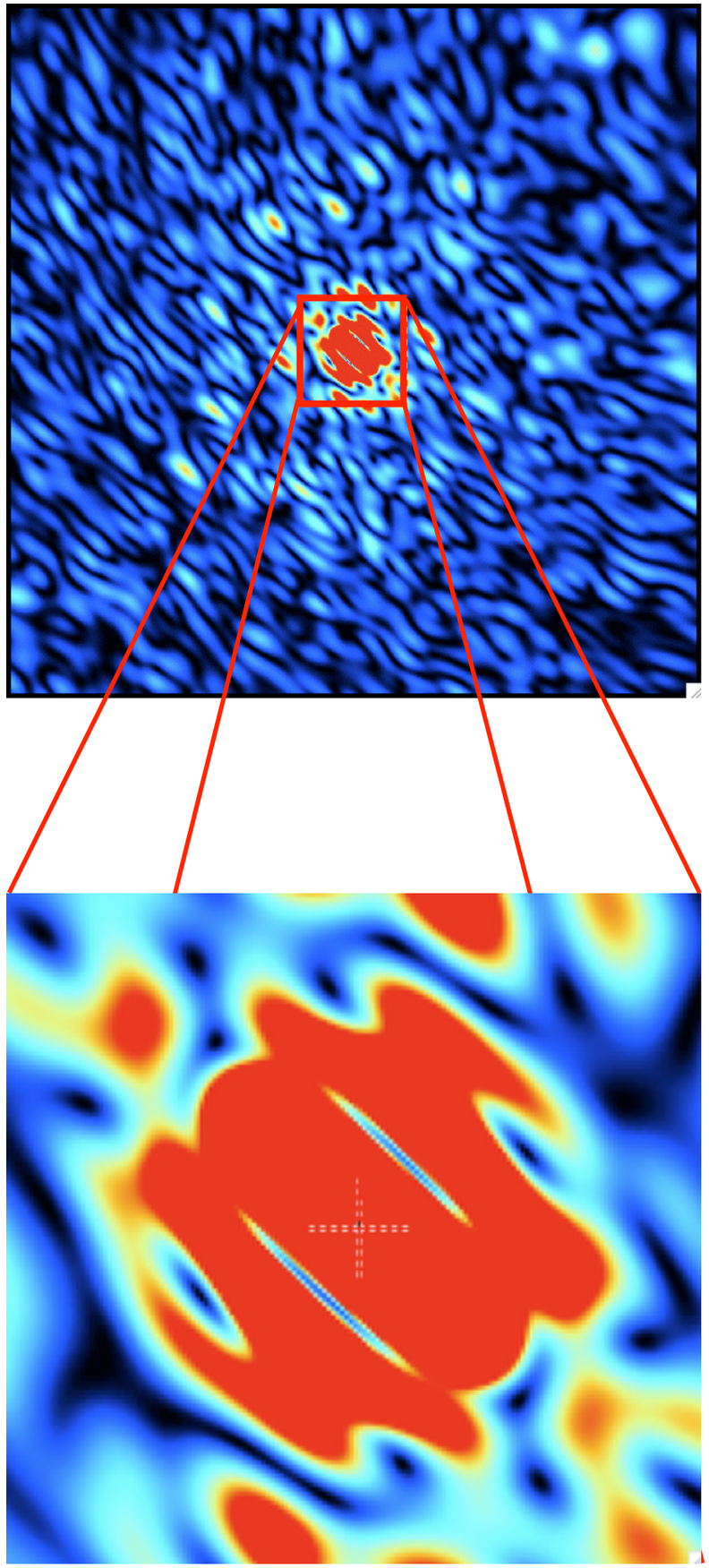


Molecular Transform



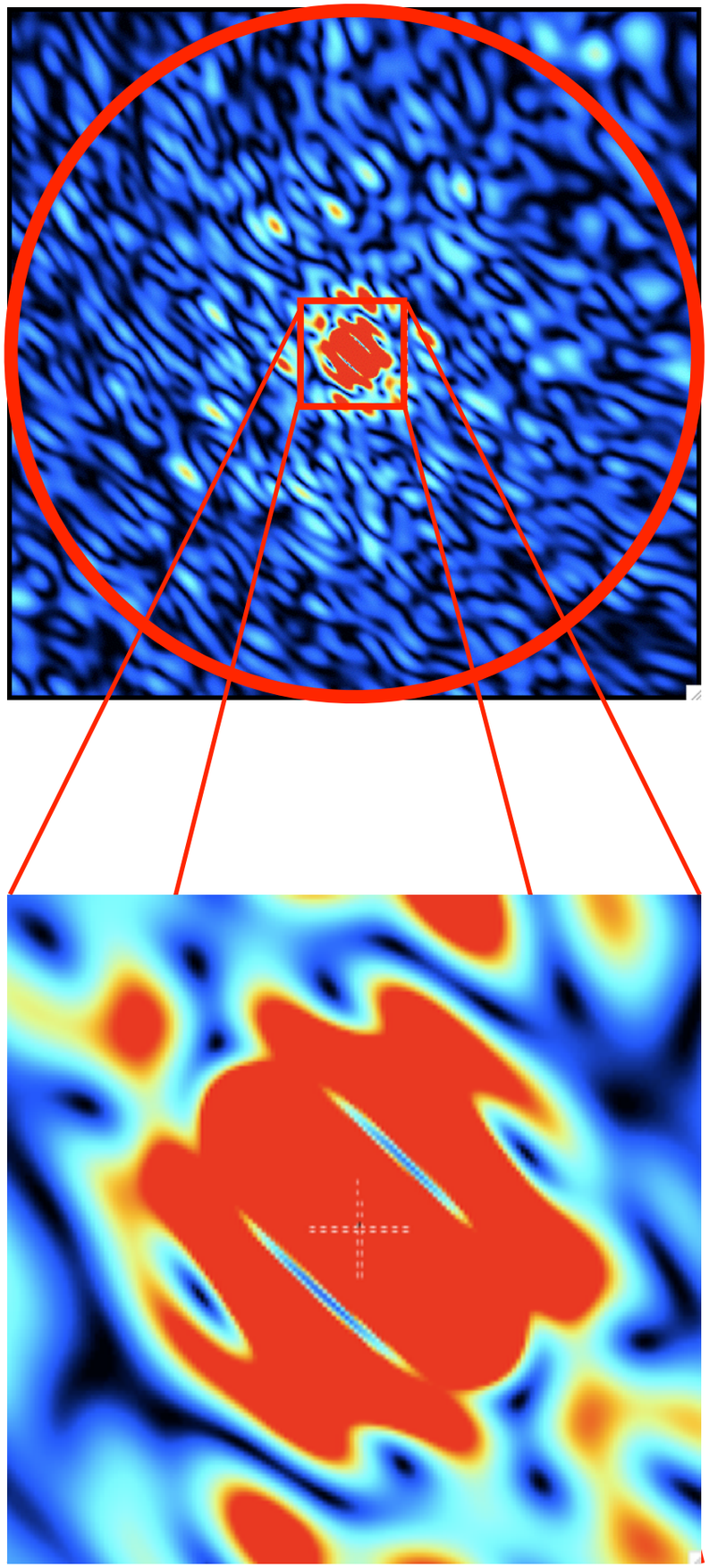


Molecular Transform



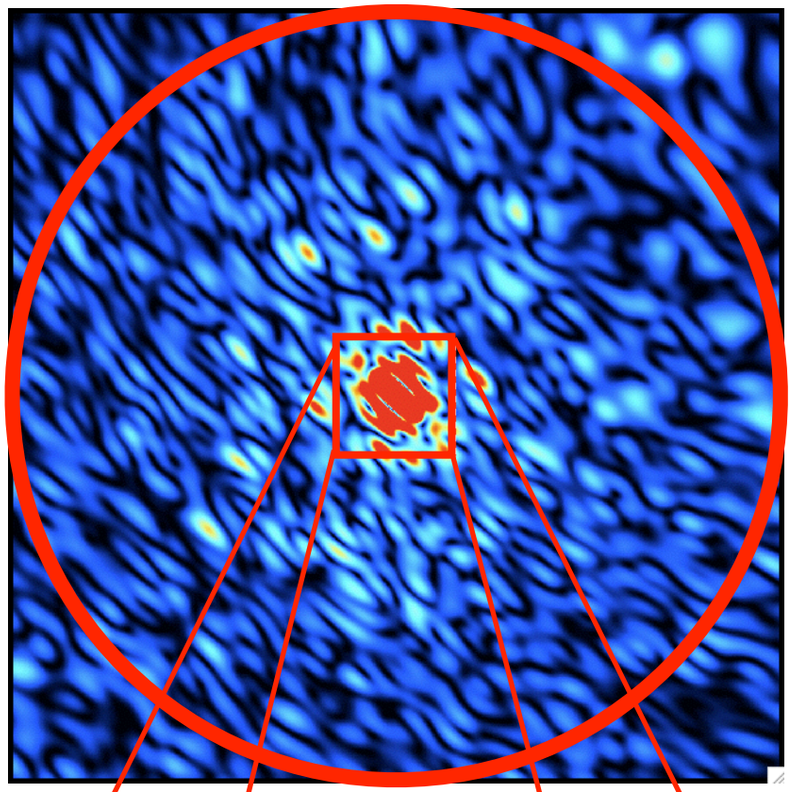
Molecular Transform

2Å

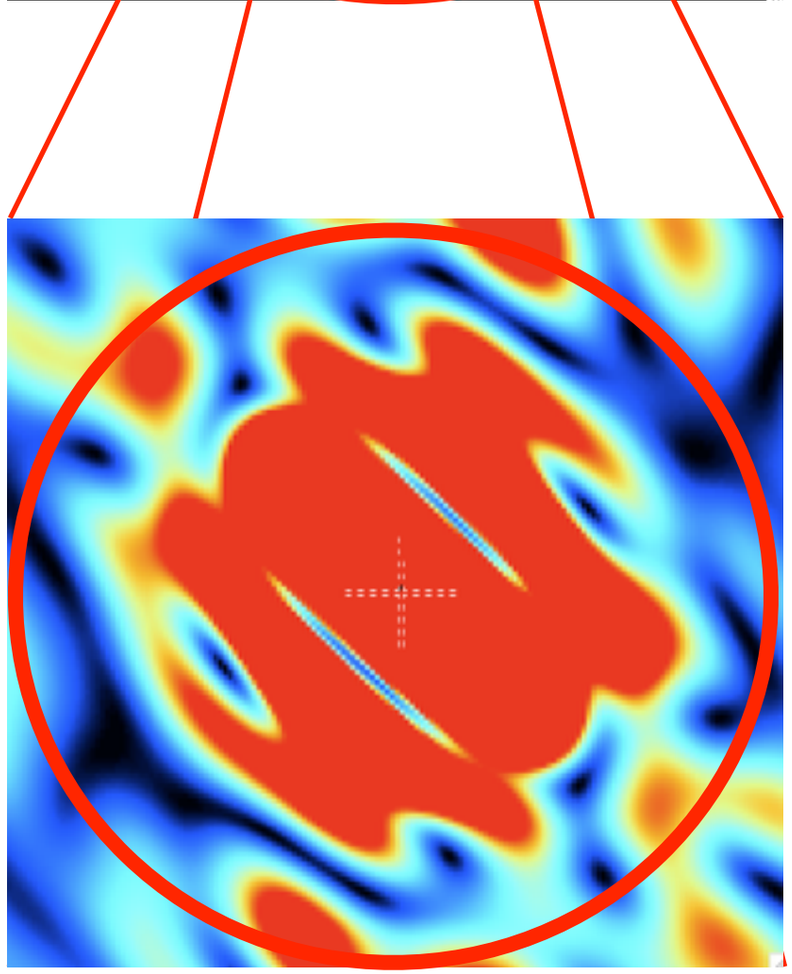


Molecular Transform

2Å

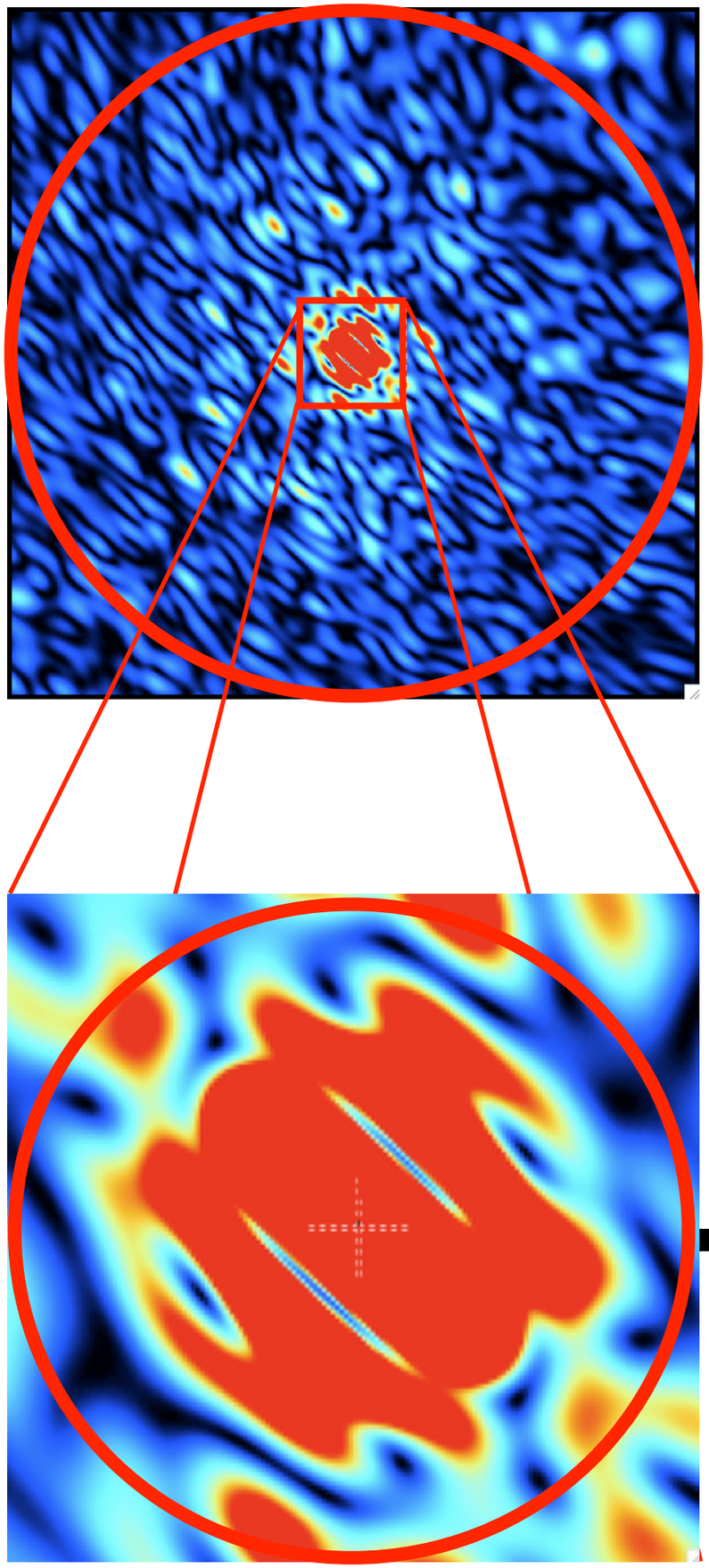


20Å



Molecular Transform

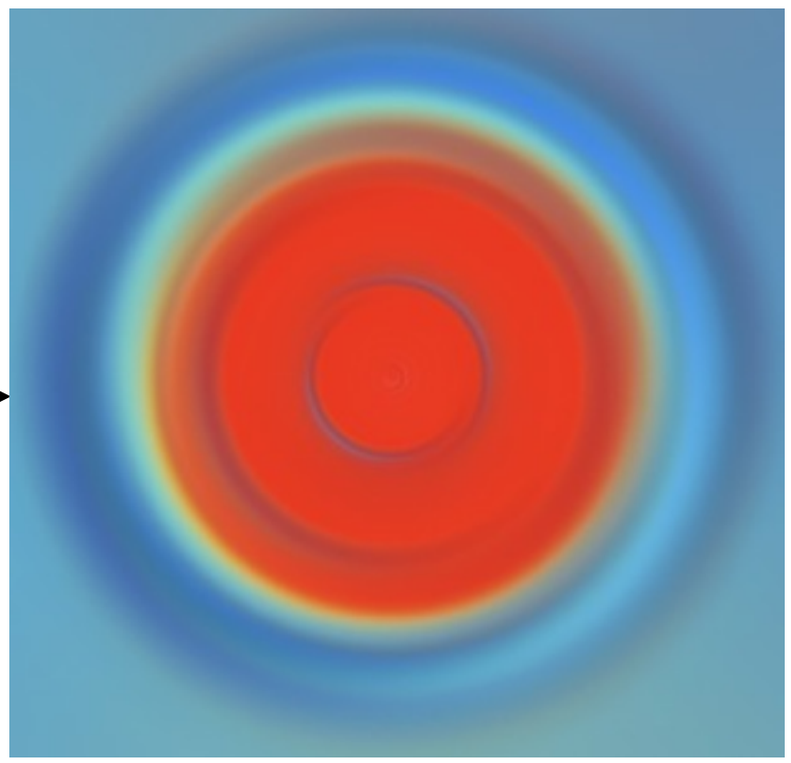
2Å



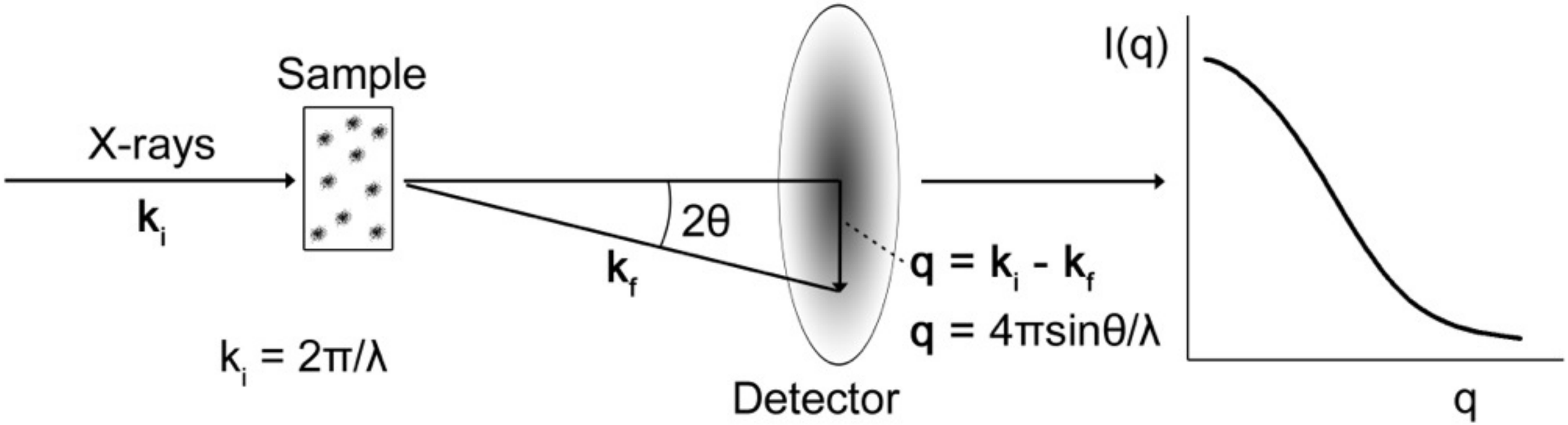
20Å

Molecular Transform

Spherical averaging
from solution of
tumbling molecules



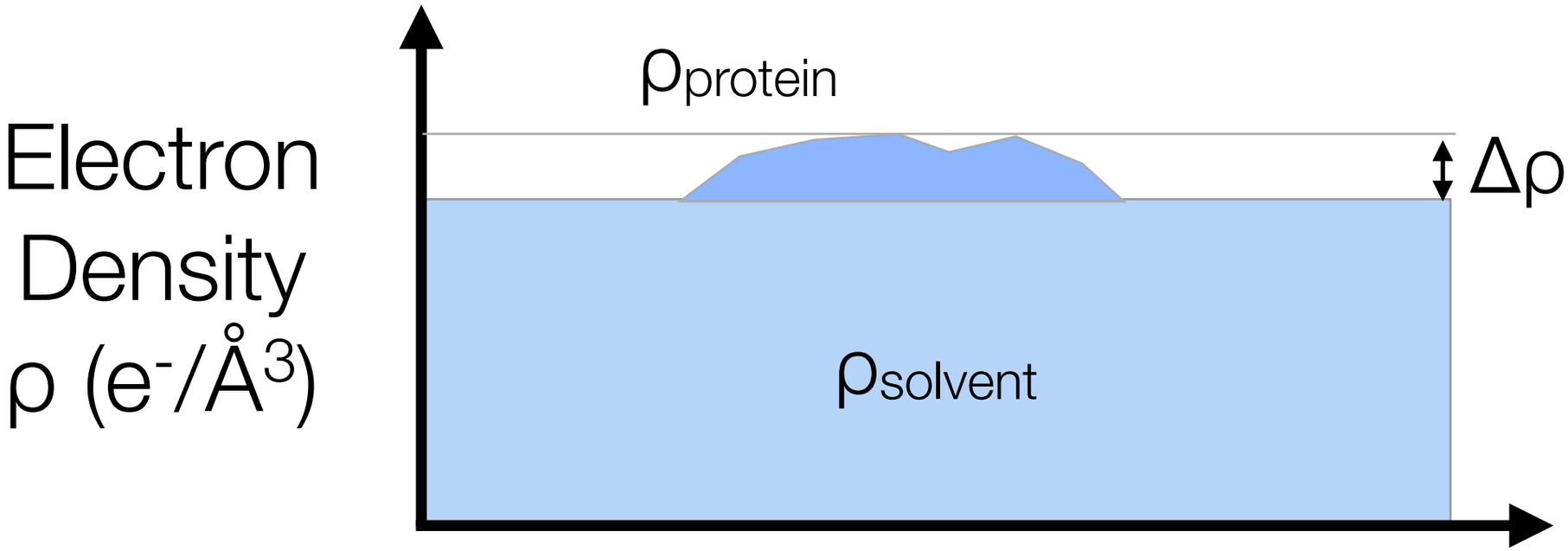
Basic SAXS Set-up



- particles in solution tumble - spherically averaged intensity is recorded
- radial integration results in one dimensional SAXS profile
- larger particles scatter at smaller angles → reciprocal space
- analysis of 1D profile yields info about size and shape of particles in solution

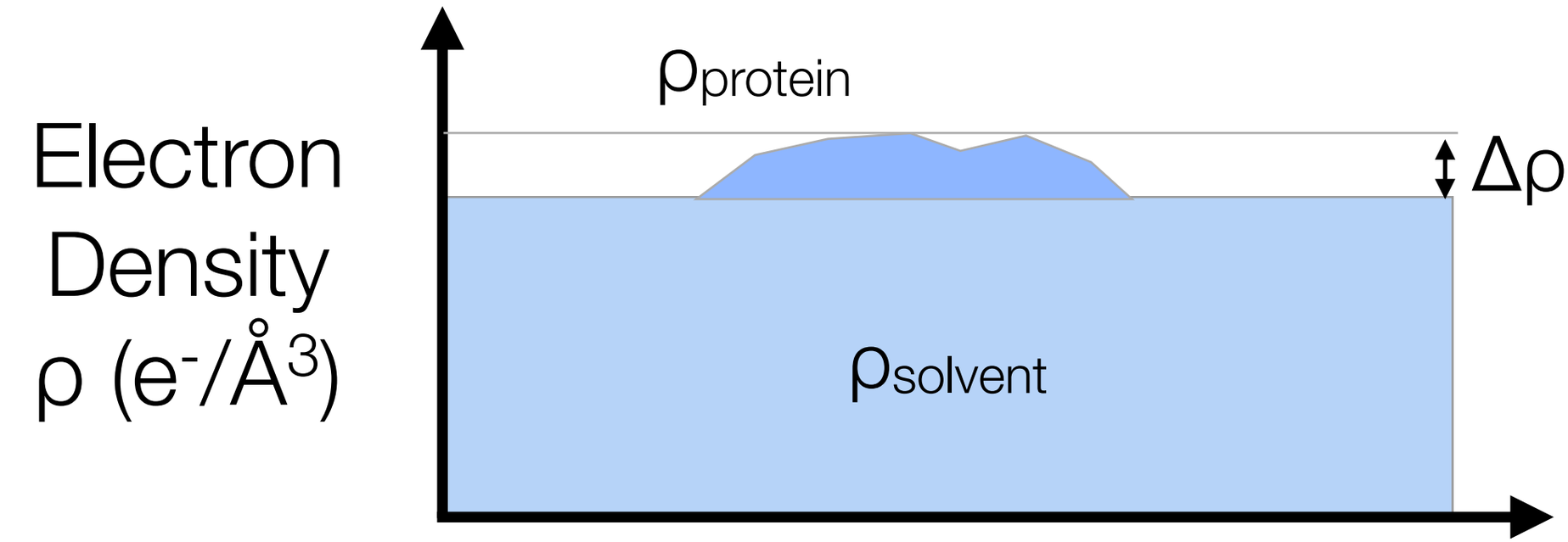
Contrast

- SAXS is a contrast method, i.e. it depends on the square of the difference in the electron density between the molecule and the solvent



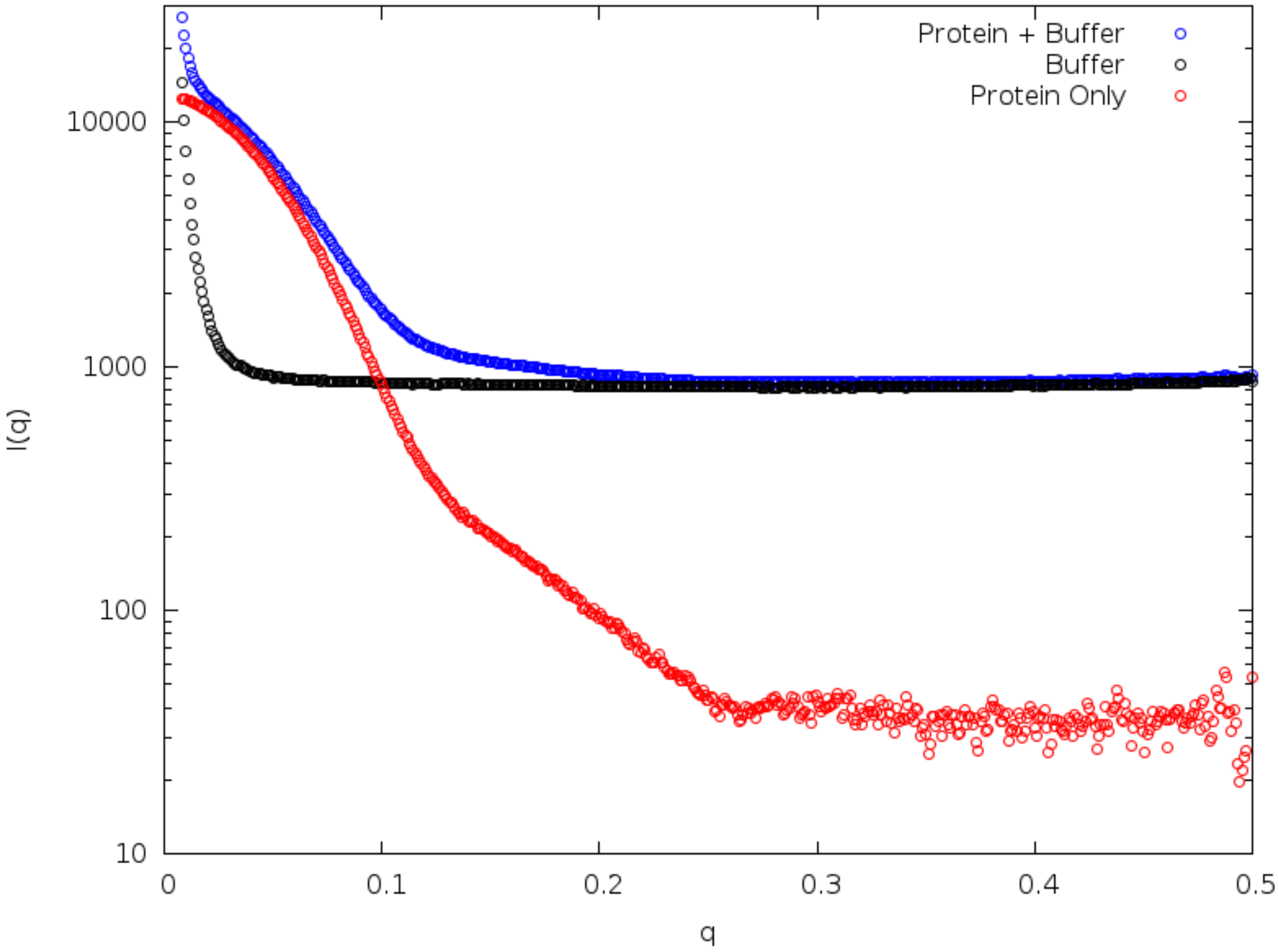
Contrast

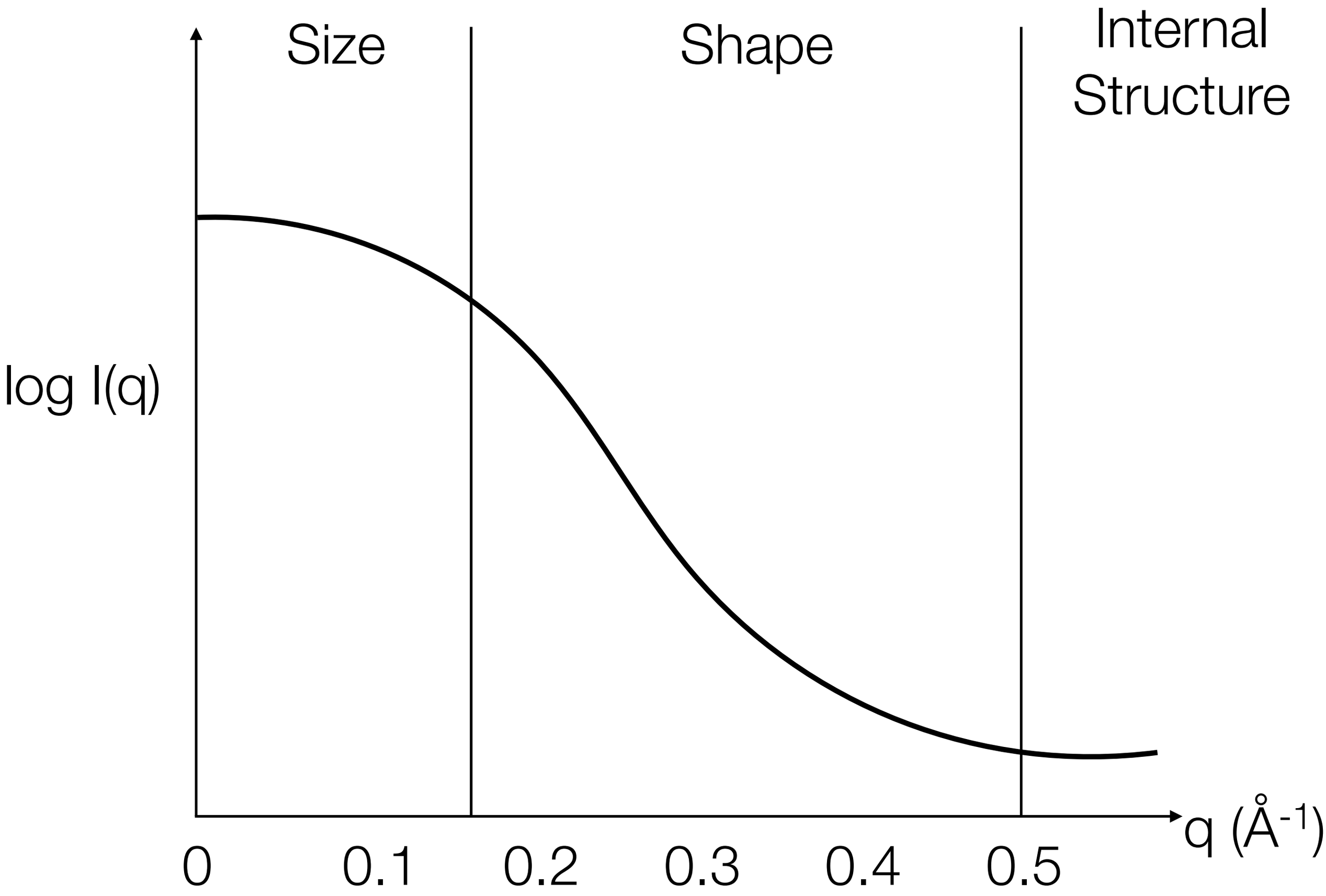
- SAXS is a contrast method, i.e. it depends on the square of the difference in the electron density between the molecule and the solvent



$$(\Delta\rho)^2 = (\rho_{protein} - \rho_{water})^2 = (0.44\text{-}0.33)^2 \approx$$ 10% above background

Buffer Subtraction




Size
Shape
Internal Structure
log I(q)
q (Å-1)
0
0.1
0.2
0.3
0.4
0.5

What can SAXS provide?

- Radius of gyration
- Maximum particle dimension
- Molecular weight
- Oligomeric state and organization in solution
- Amount of native flexibility or unfoldedness
- Visualization of disordered regions not seen in X-ray crystallography
- Low resolution molecular envelope

Interparticle Interactions

- Equation for scattering intensity:

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

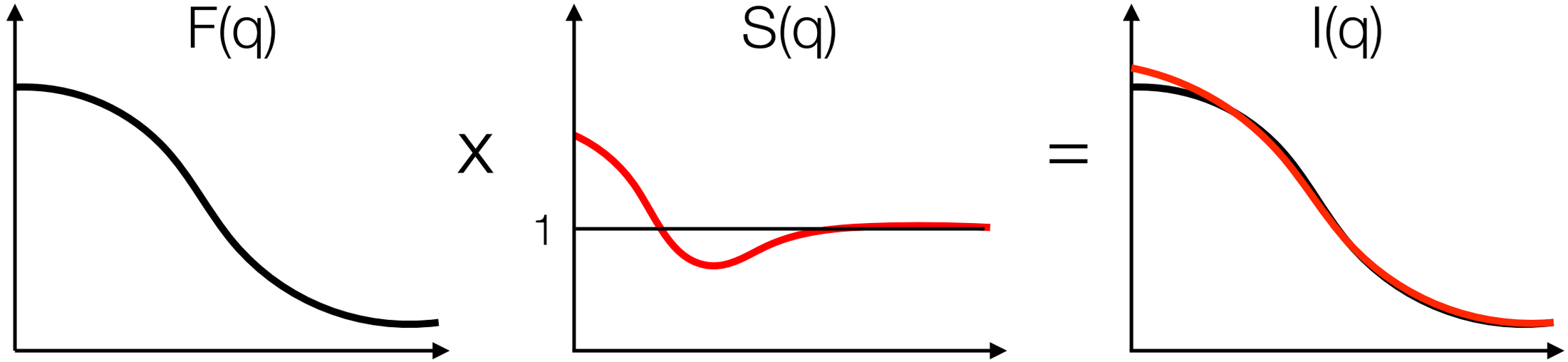
Experimental Intensity	Form factor of particle	Structure factor of solution

- Form factor describes *intra*particle interactions, i.e. size and shape
- Structure factor describes *inter*particle interactions, i.e. repulsion/attraction
- Ideally a monodisperse solution for SAXS should have no interparticle interactions, i.e. S(q) = 1

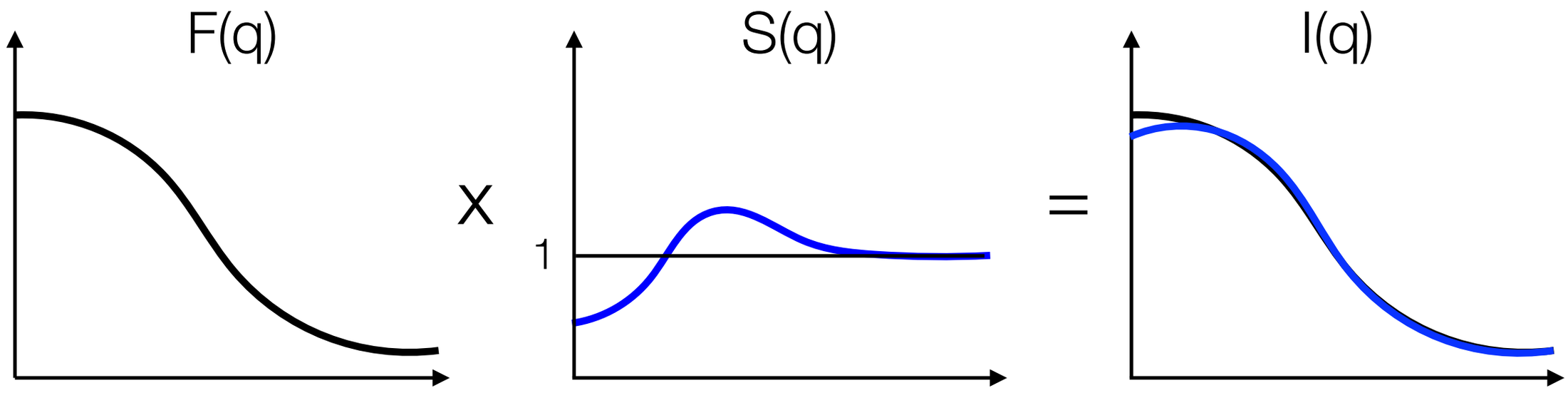
Interparticle Interactions

S(q) ≠ 1 affects low q data most

Attraction



Repulsion



Guinier Method

- Developed by André Guinier in 1939.
- As q → 0, intensity can be approximated by:

$$I(q) = I_0 e^{-q^2 R_g^2/3}$$

$$\ln I(q) = \ln I_0 - \frac{R_g^2}{3} q^2$$

Guinier Method

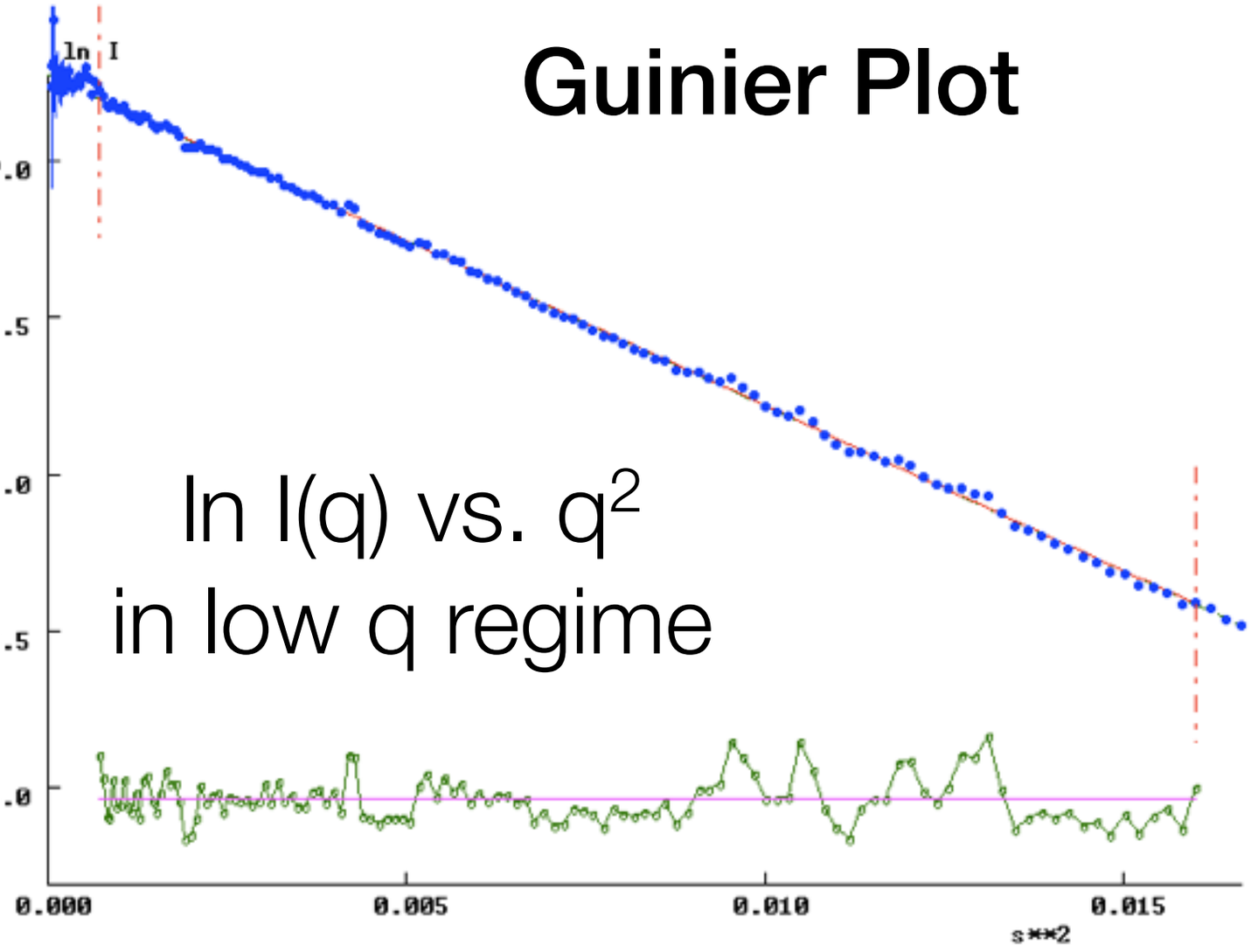
- Developed by André Guinier in 1939.
- As q → 0, intensity can be approximated by:

$$I(q) = I_0 e^{-q^2 R_g^2/3}$$

$$\ln I(q) = \ln I_0 - \frac{R_g^2}{3} q^2$$

y = b + m * x

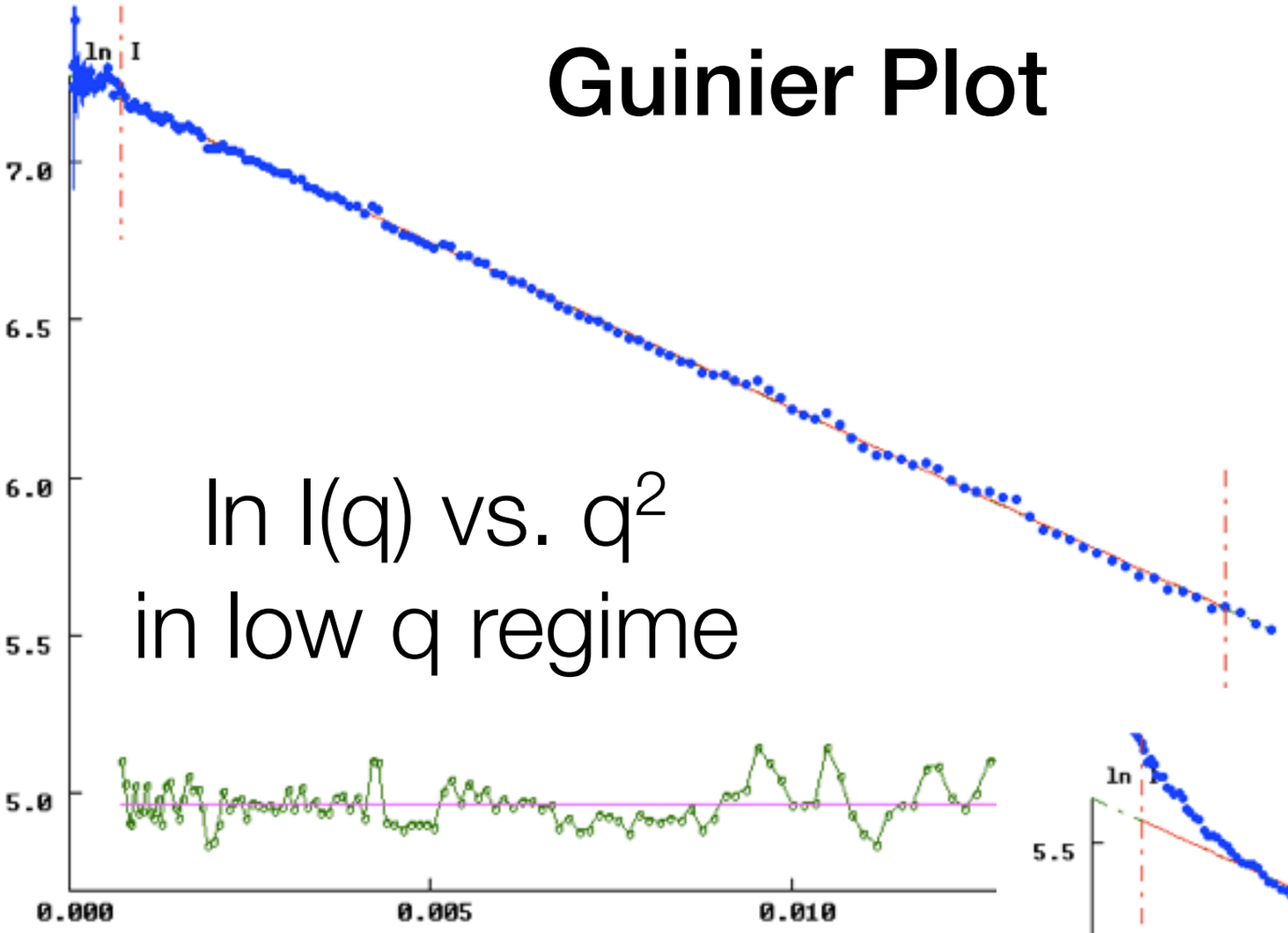
Guinier Plot



ln I(q) vs. q^2
in low q regime

Straight line shows no aggregation: Can determine R_g from slope of line, I(0) from intercept

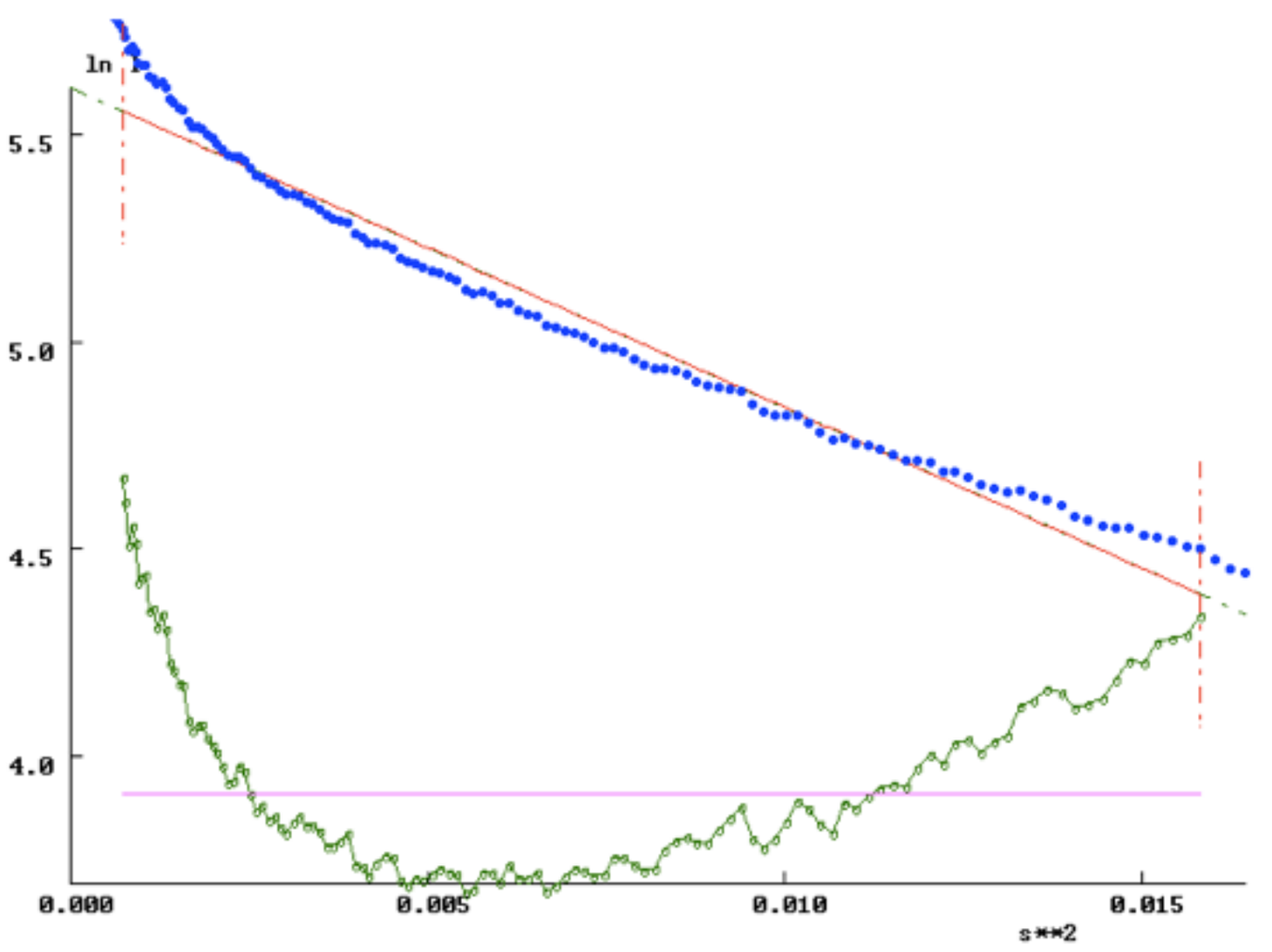
Guinier Plot



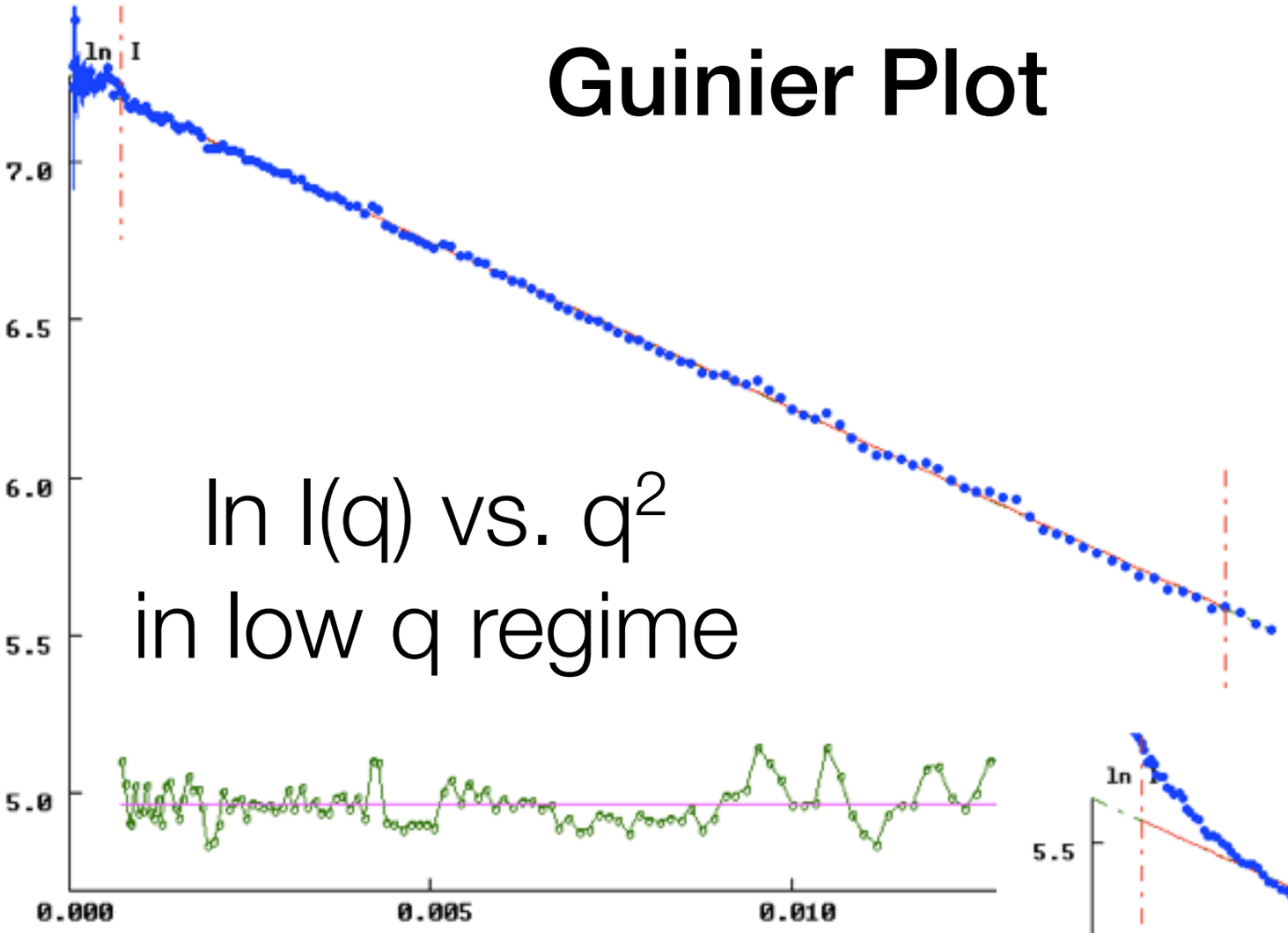
ln I(q) vs. q^2
in low q regime

Straight line shows no aggregation: Can determine R_g from slope of line, I(0) from intercept

Curved line shows attraction or aggregates present:
No SAXS processing should be done



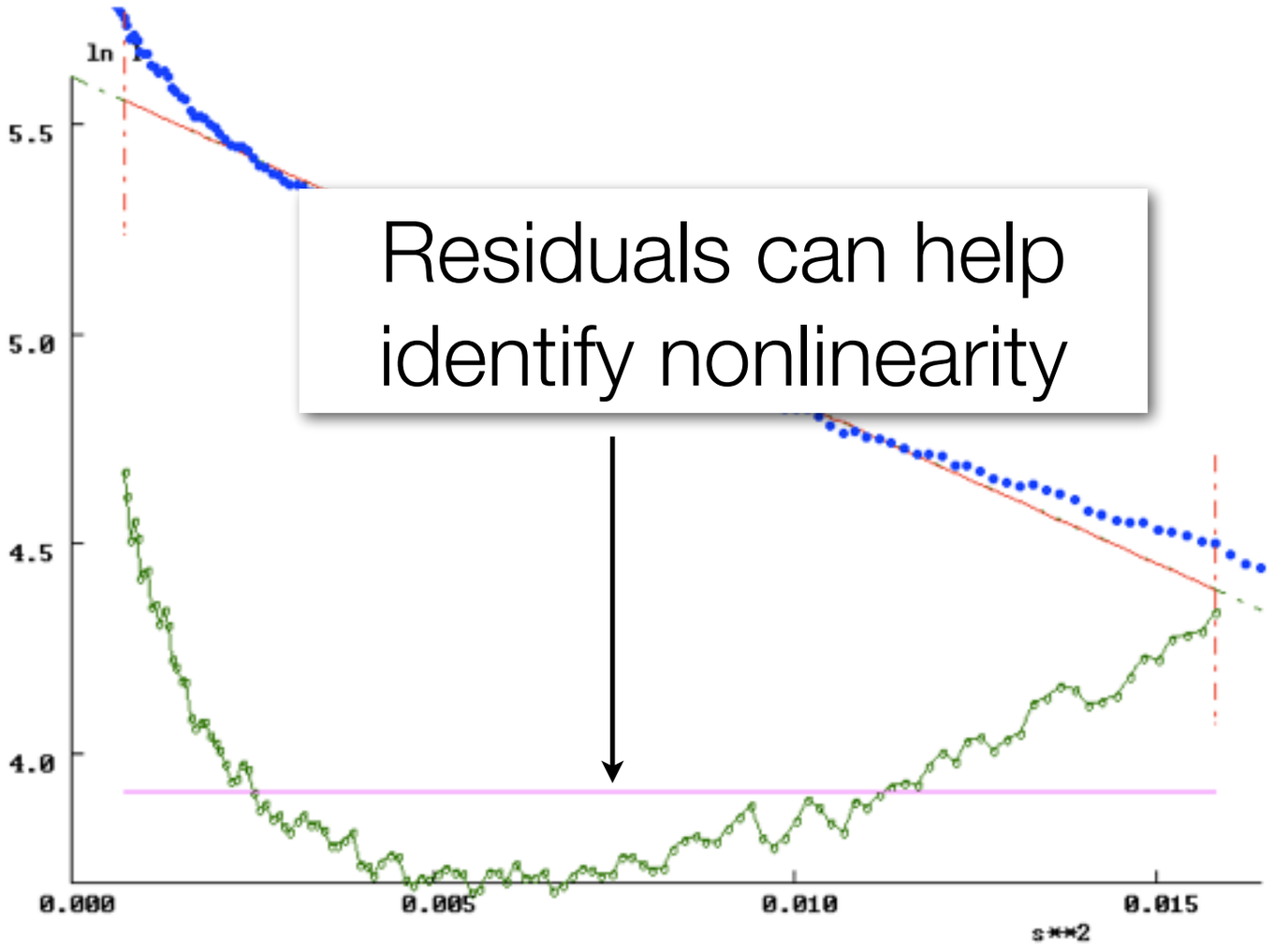
Guinier Plot



ln I(q) vs. q^2
in low q regime

Straight line shows no aggregation: Can determine R_g from slope of line, I(0) from intercept

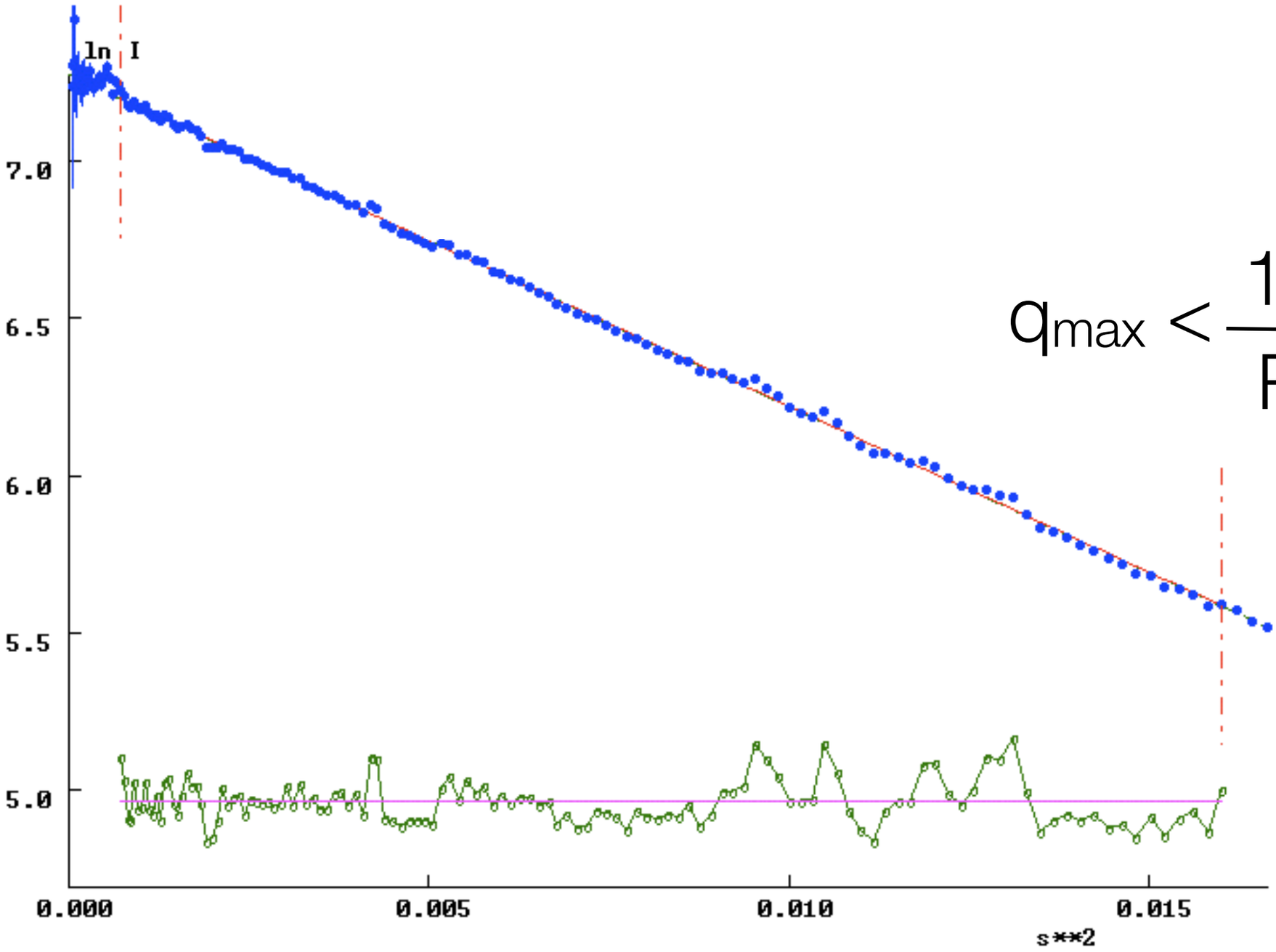
Curved line shows attraction or aggregates present:
No SAXS processing should be done



Residuals can help identify nonlinearity

Guinier Region

$$q_{min} < \frac{\pi}{D_{max}}$$

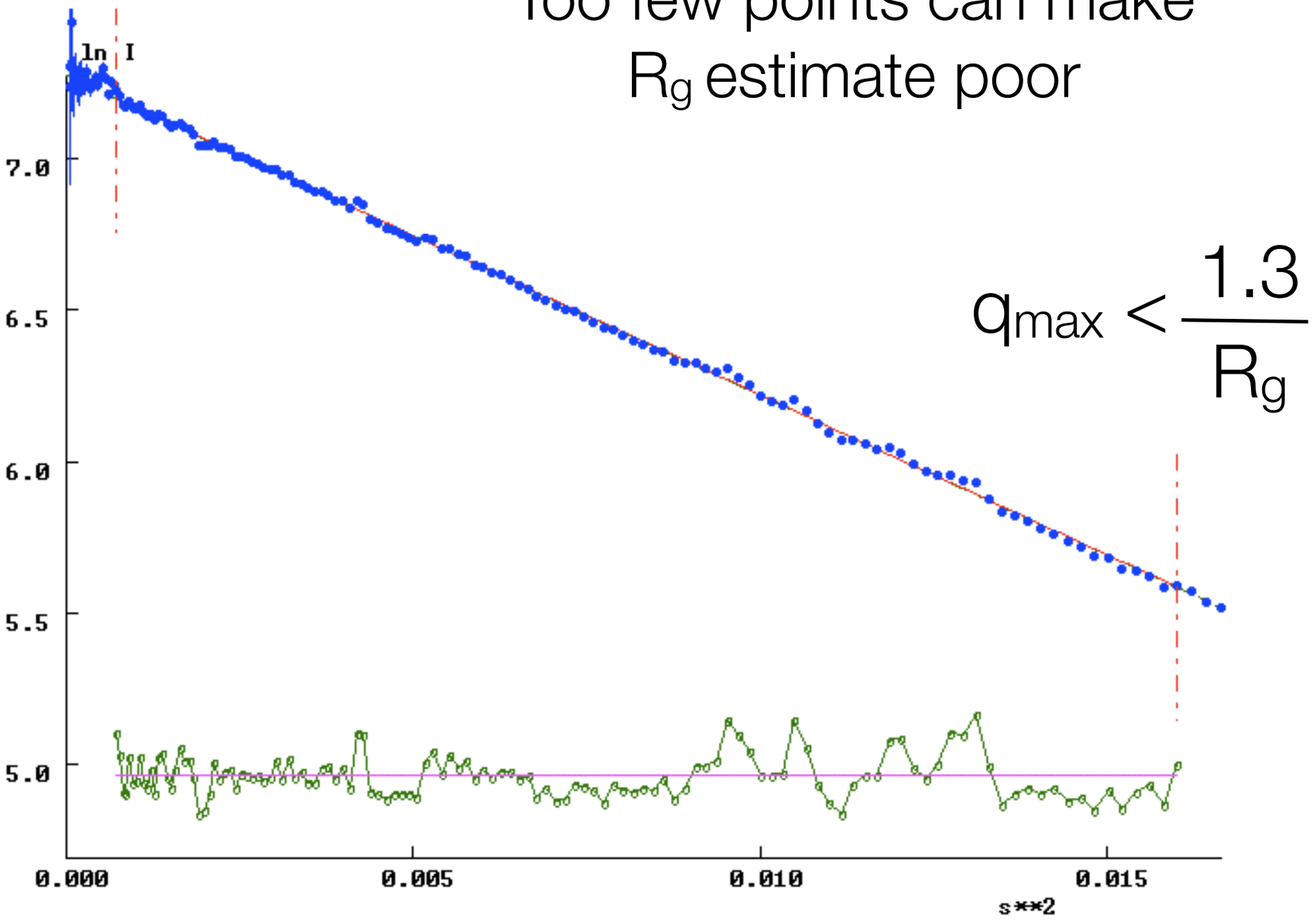


$$q_{max} < \frac{1.3}{R_g}$$

Guinier Region

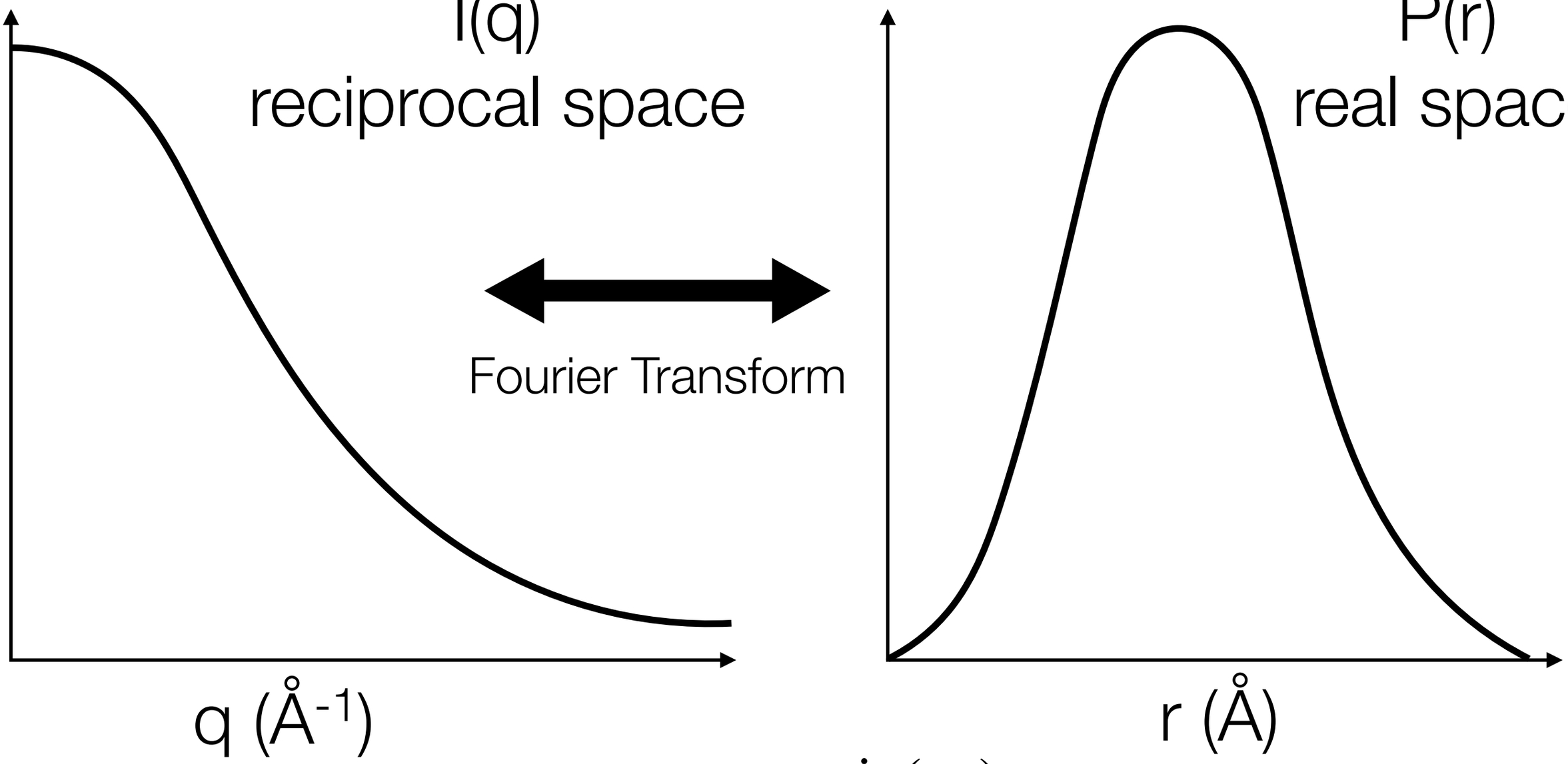
$$q_{min} < \frac{\pi}{D_{max}}$$

Very large particles -->
Narrow Guinier region
Too few points can make
R_g estimate poor



$$q_{max} < \frac{1.3}{R_g}$$

Pair Distance Distribution Function



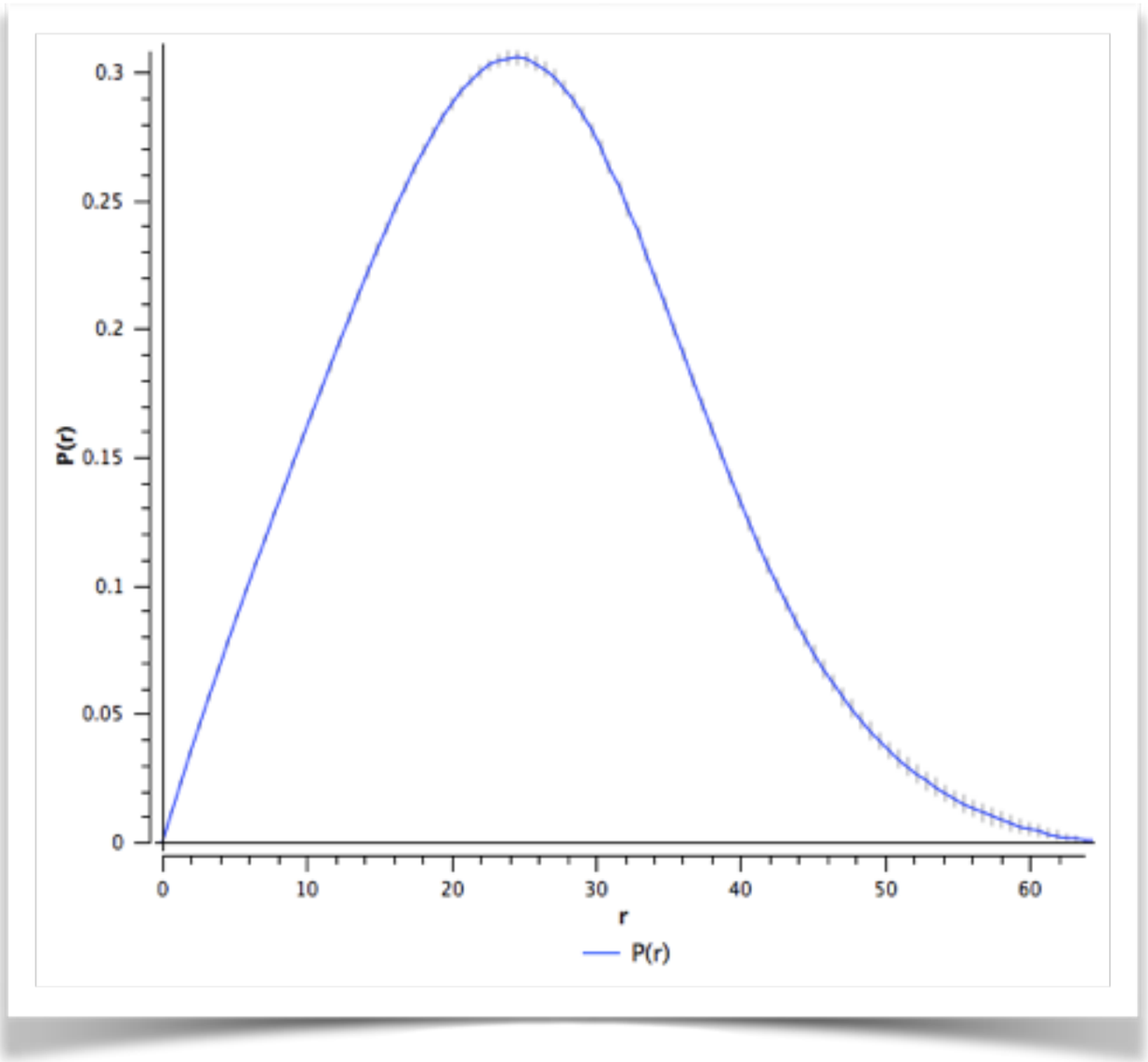
$$I(q) = \int p(r) \frac{\sin(qr)}{qr} dr$$

Pair Distance Distribution Function

- Rg can be calculated from

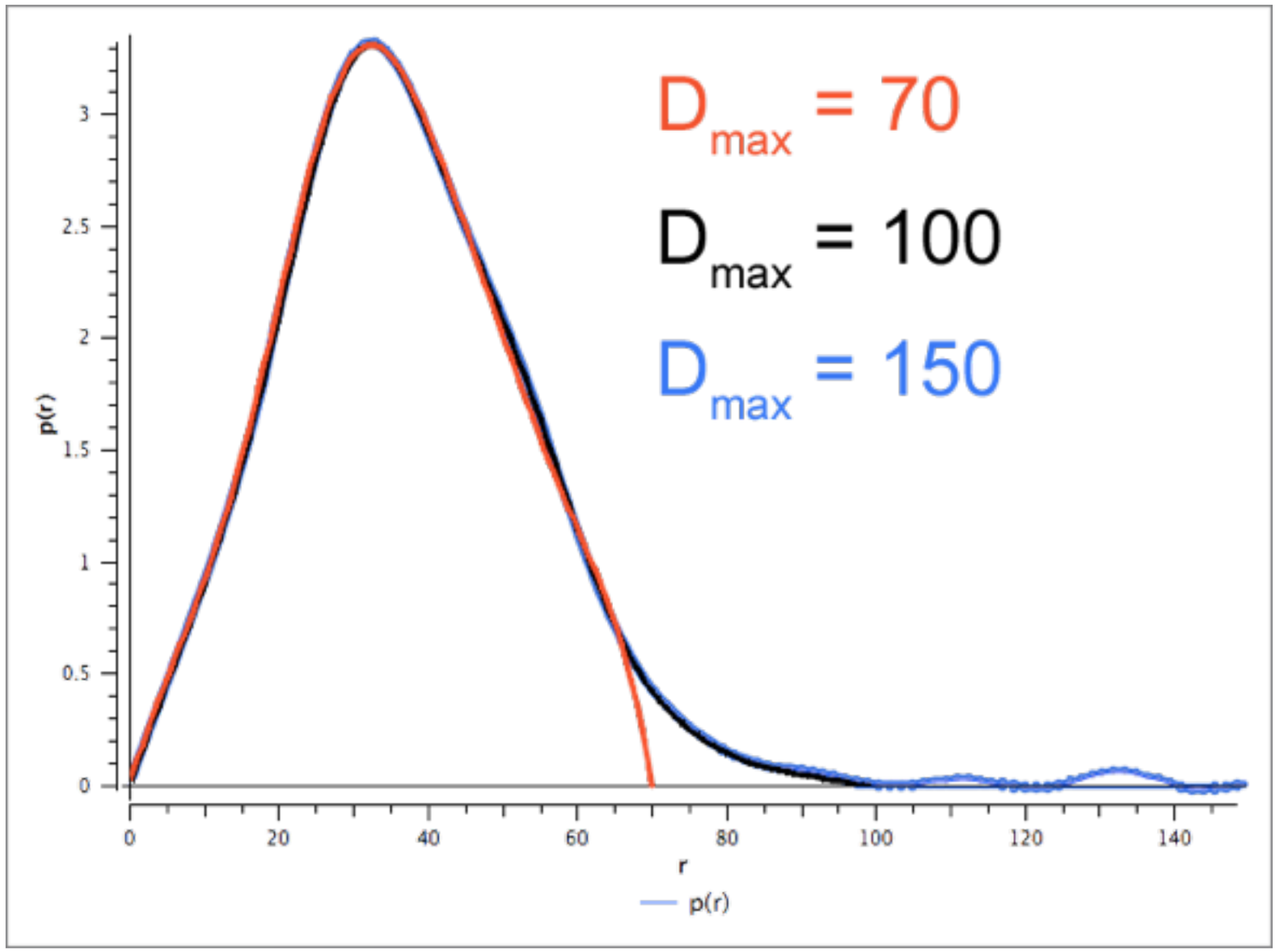
$$R_G^2 = \int_0^{D_{\max}} r^2 P(r) dr \Big/ \int_0^{D_{\max}} P(r) dr$$

- Uses entire curve, less sensitive to interparticle effects
- Especially useful for large particles with narrow Guinier region and noisy data
- A good check of data quality against Guinier Rg estimate

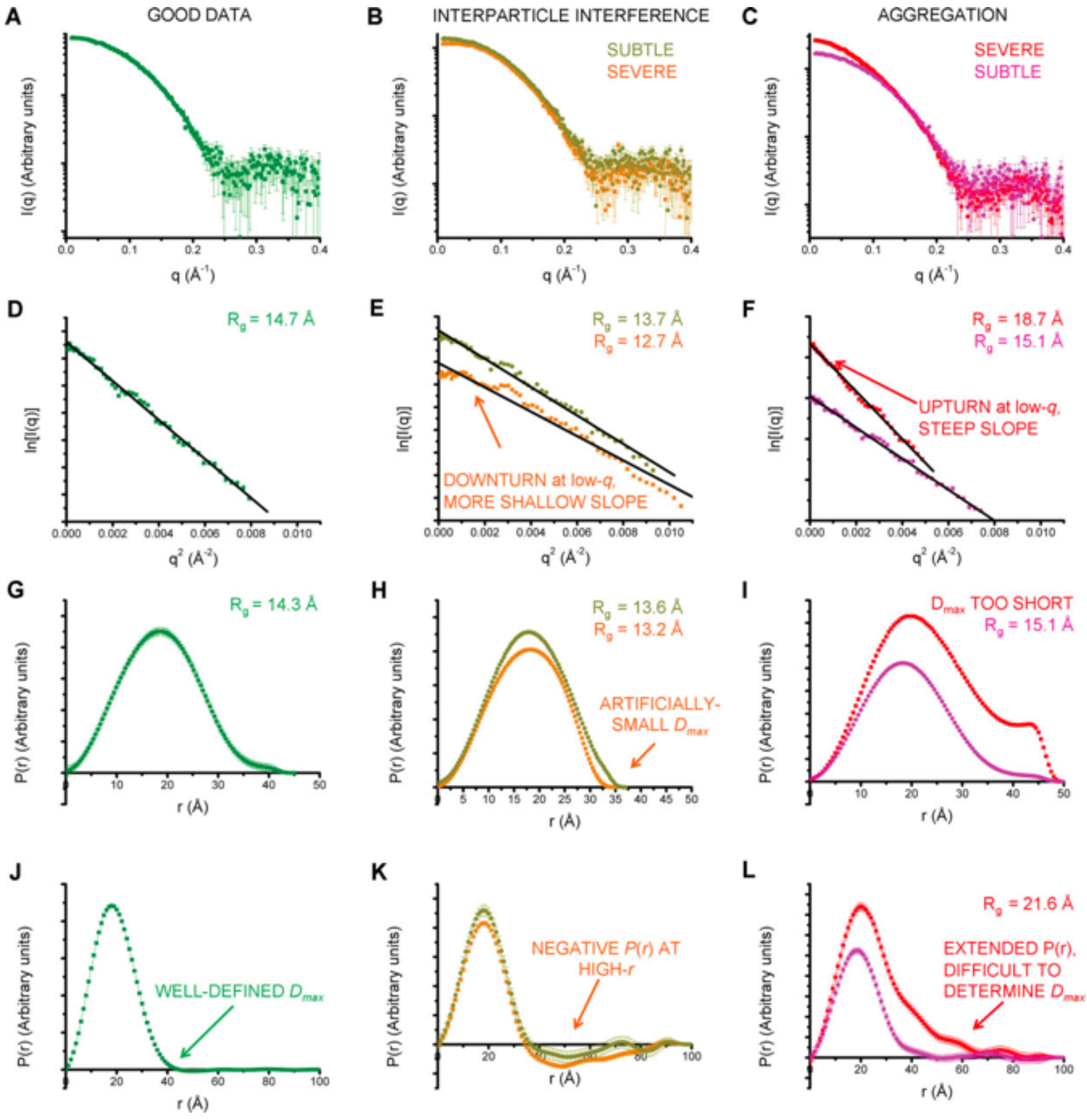


Pair Distance Distribution Function

- Can be used to determine D_{max}
- P(r) should gradually fall to zero at D_{max}
- Underestimated D_{max} appears as abrupt, forced descent to zero
- Starting with large values should identify a decent estimate of D_{max}, given good quality data
- Errors in D_{max} can be large, (~10 - 20%) for good data



Sample quality greatly affects data analysis



Jacques and Trewhella, 2010 *Protein Science* Review

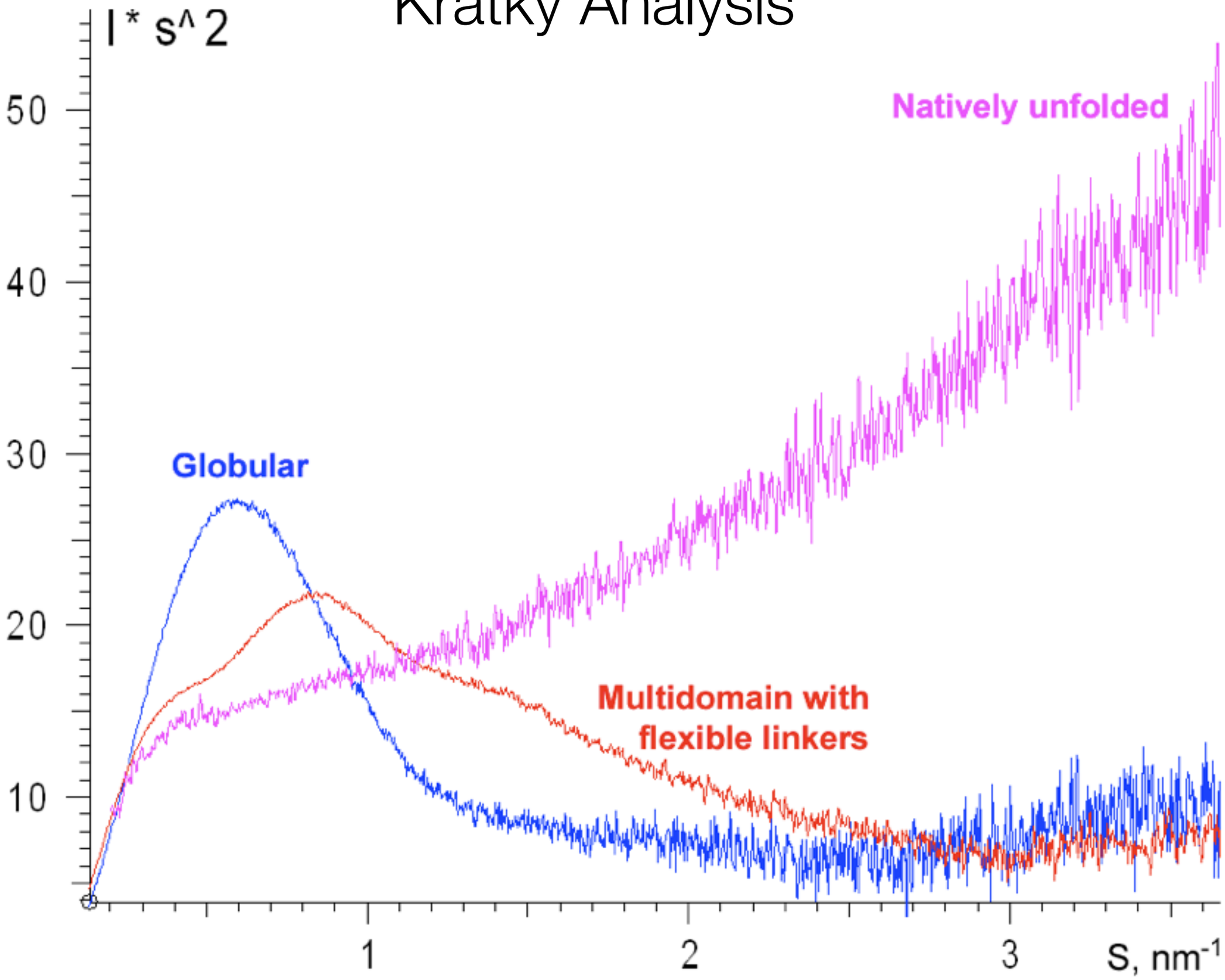
Data Quality

- Use alternate methods (such as MALS, DLS, SEC) to characterize your sample to ensure no aggregation or polydispersity
- High concentrations yield high signal to noise
    - typical concentrations range from 1 - 10 mg/ml
    - smaller particles require higher concentrations than larger particles
    - RNA/DNA scatters more strongly, thus lower concentrations needed
- Check multiple concentrations to ensure no concentration dependence is occurring
- High signal-to-noise is important, but not as important as good sample quality
- Accurate buffer subtraction is essential (dialysis or flow-through buffer)

Multiple Methods to Estimate Molecular Weight

- From I(0) using experimental intensity calibration
  - Intensity at q=0 (extrapolated). Corresponds to square of number of electrons in particle (similar to F(0 0 0) reflection is crystallography)
  - Typical standards include water, BSA, Xylose Isomerase, or Lysozyme (check for interparticle effects in protein standards also)
- From particle volume
  - Assumes average protein density of 1.37 g/cm^3
- New method (Rambo, et al 2013) accurate even for disordered proteins
- SAXS MoW (http://www.if.sc.usp.br/~saxs/saxsmow.html)
- **Molecular weight estimation methods accurate to about 10%**

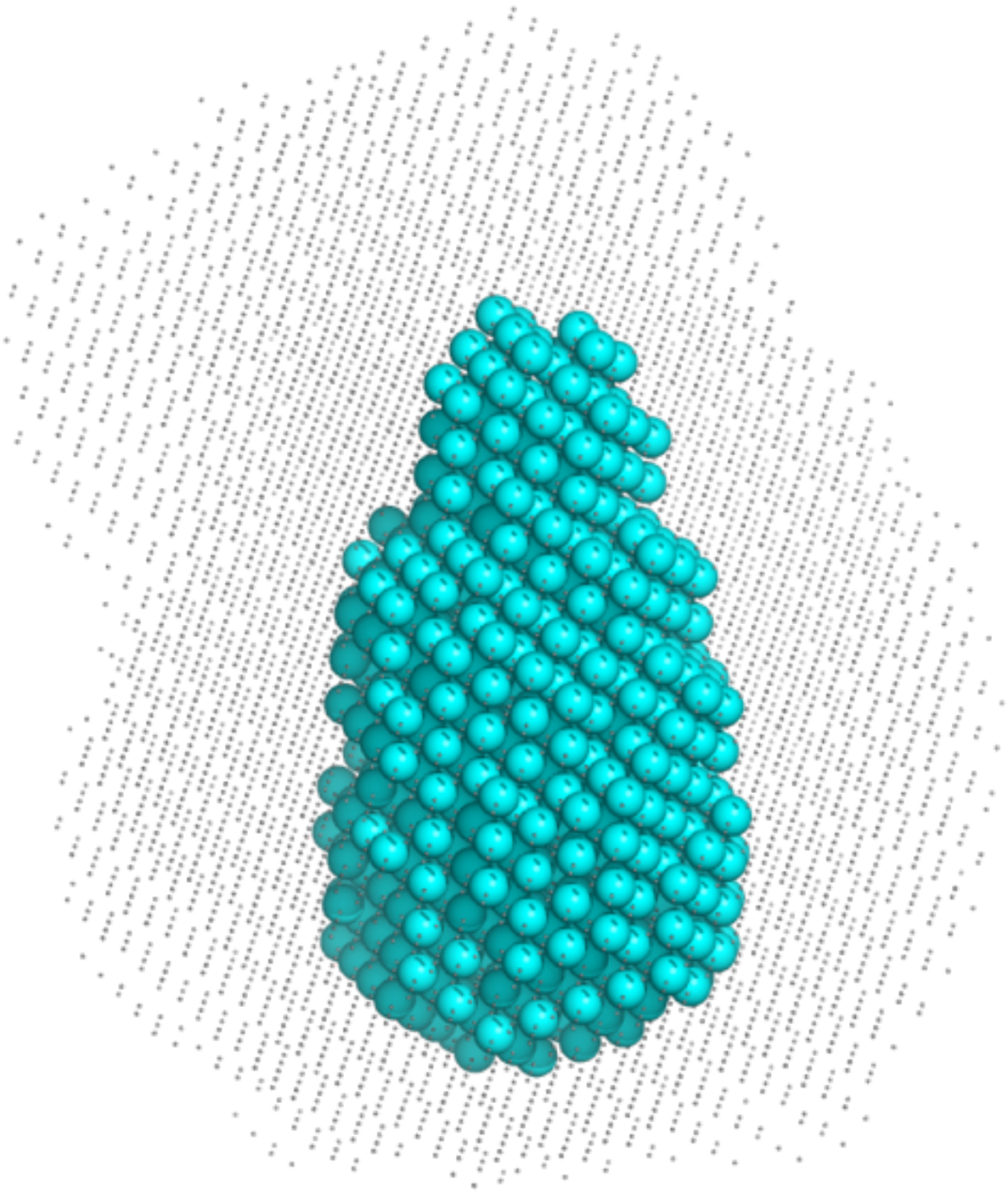
Kratky Analysis



http://www.saxier.org/forum/viewtopic.php?t=337

Envelope Reconstruction

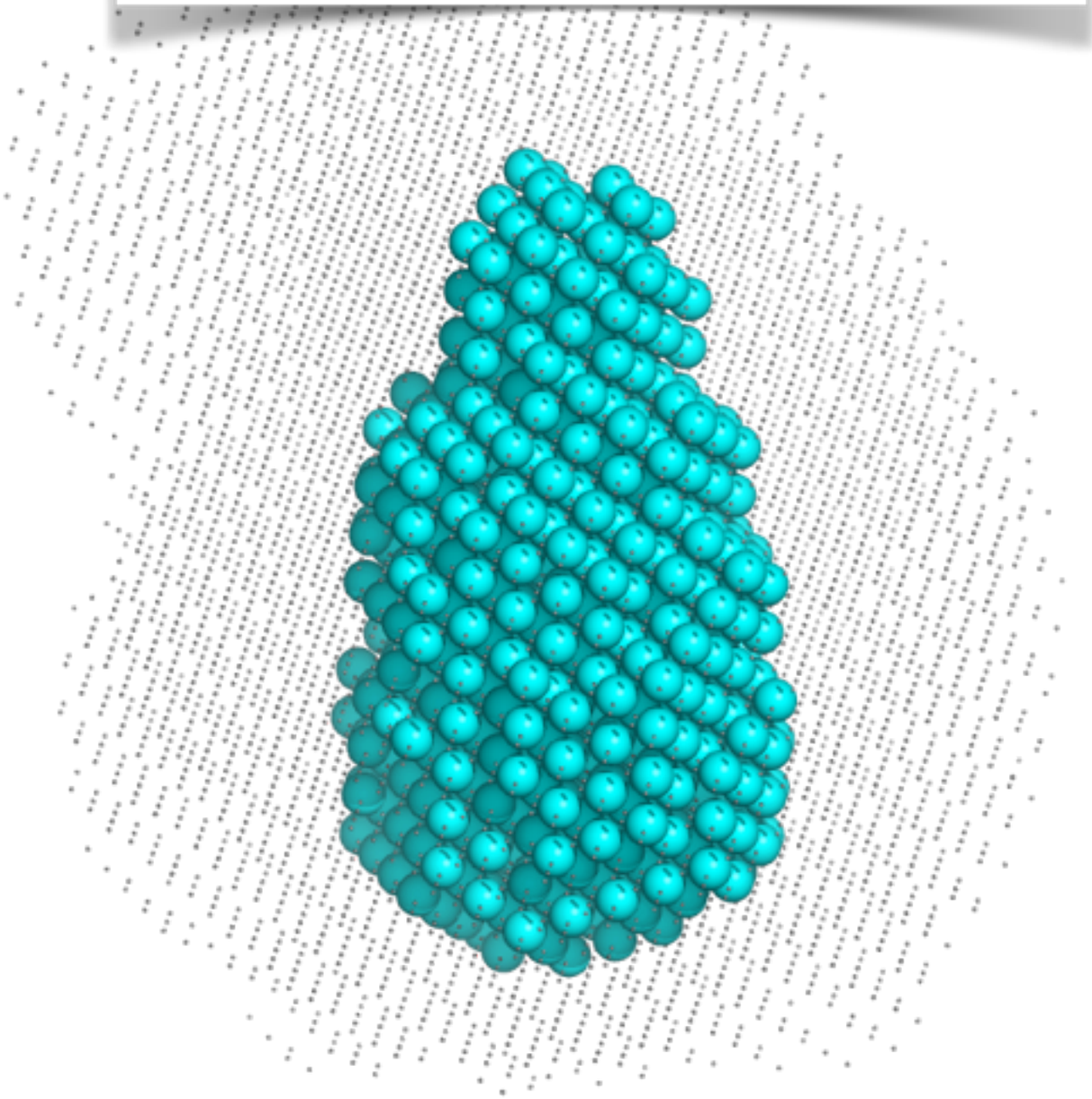
- Several programs exist for *ab initio* envelope reconstructions, most common is DAMMIF
- Possible models for conventional minimization procedures too numerous to be computationally feasible (2^N)
- Monte-Carlo like approaches must be used
- Can easily fall into local minima
- Simulated annealing used to find global minimum utilizing random seed generation



Envelope Reconstruction

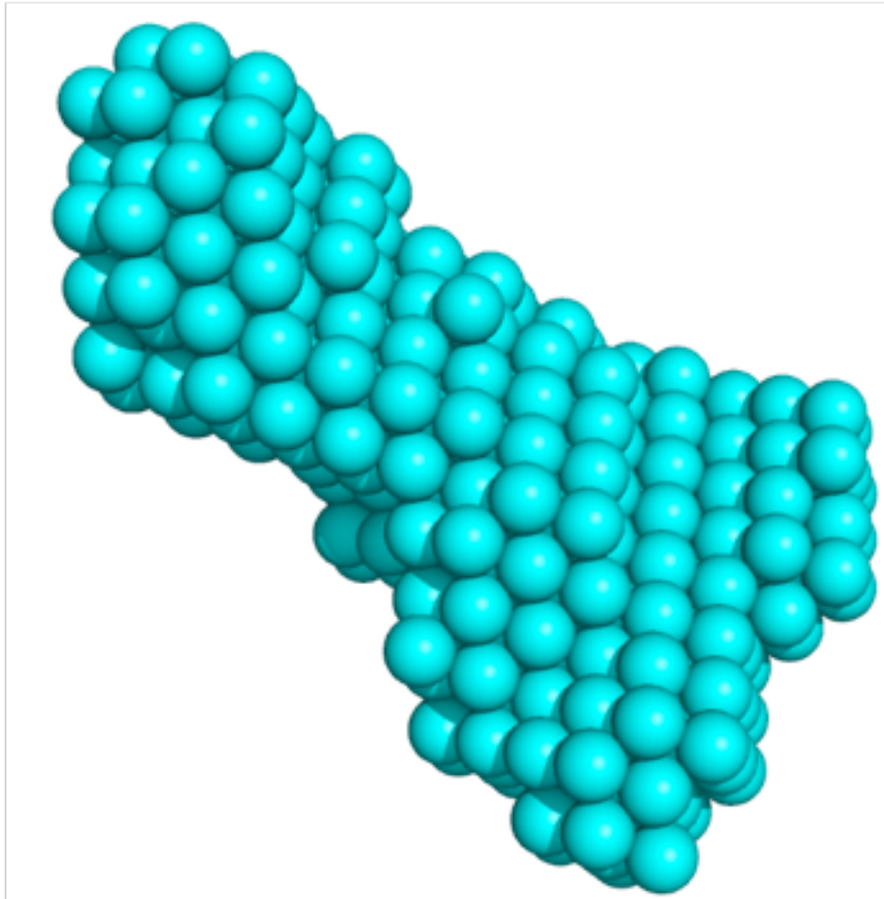
- Several programs exist for *ab initio* envelope reconstructions, most common is DAMMIF
- Possible models for conventional minimization procedures too numerous to be computationally feasible (2^N)
- Monte-Carlo like approaches must be used
- Can easily fall into local minima
- Simulated annealing used to find global minimum utilizing random seed generation

Avoid over-interpretation of envelopes



Envelope Reconstruction

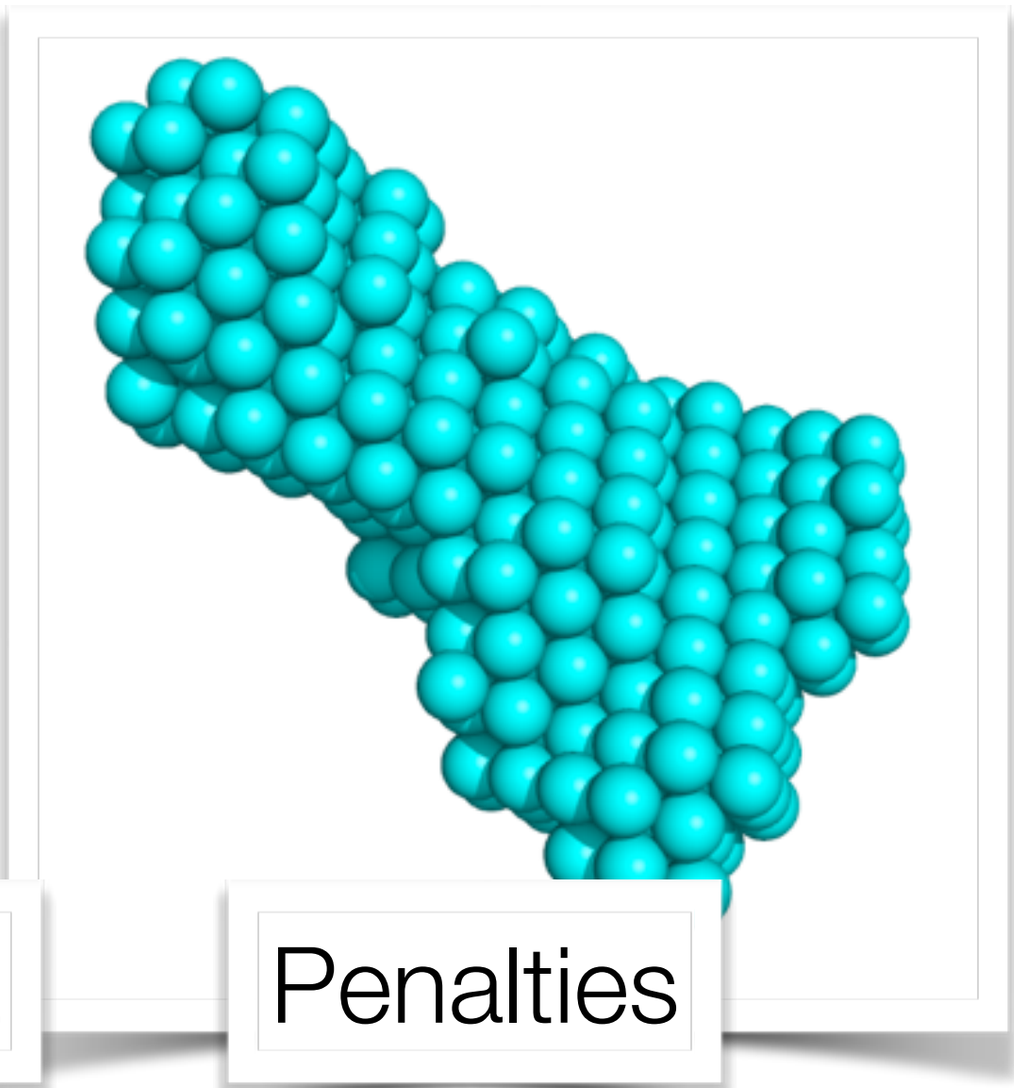
- DAMMIF uses a dummy atom “bead” modeling approach
- 3D model must not only fit the data, but also conform to physical constraints
- DAMMIF utilizes additional “penalties” to discourage the production of envelopes that are loose, not compact, or disconnected
- Due to simulated annealing protocol, multiple DAMMIF runs will produce slightly different models each time



$$Score = \chi^2\,[I_{exp}(s), I_{calc}(s)] + \alpha P(x)$$

Envelope Reconstruction

- DAMMIF uses a dummy atom “bead” modeling approach
- 3D model must not only fit the data, but also conform to physical constraints
- DAMMIF utilizes additional “penalties” to discourage the production of envelopes that are loose, not compact, or disconnected
- Due to simulated annealing protocol, multiple DAMMIF runs will produce slightly different models each time



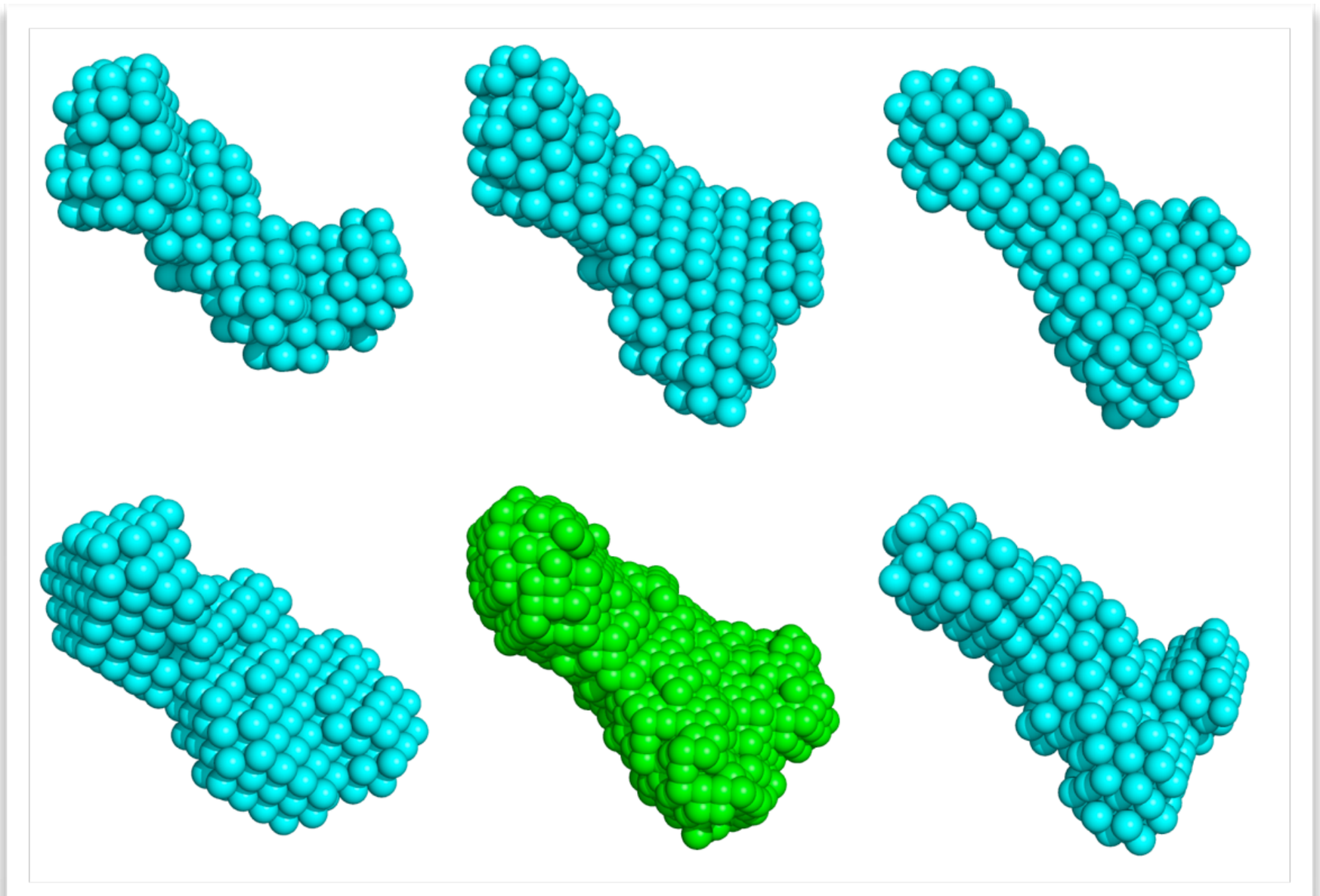
Fit to data

Penalties

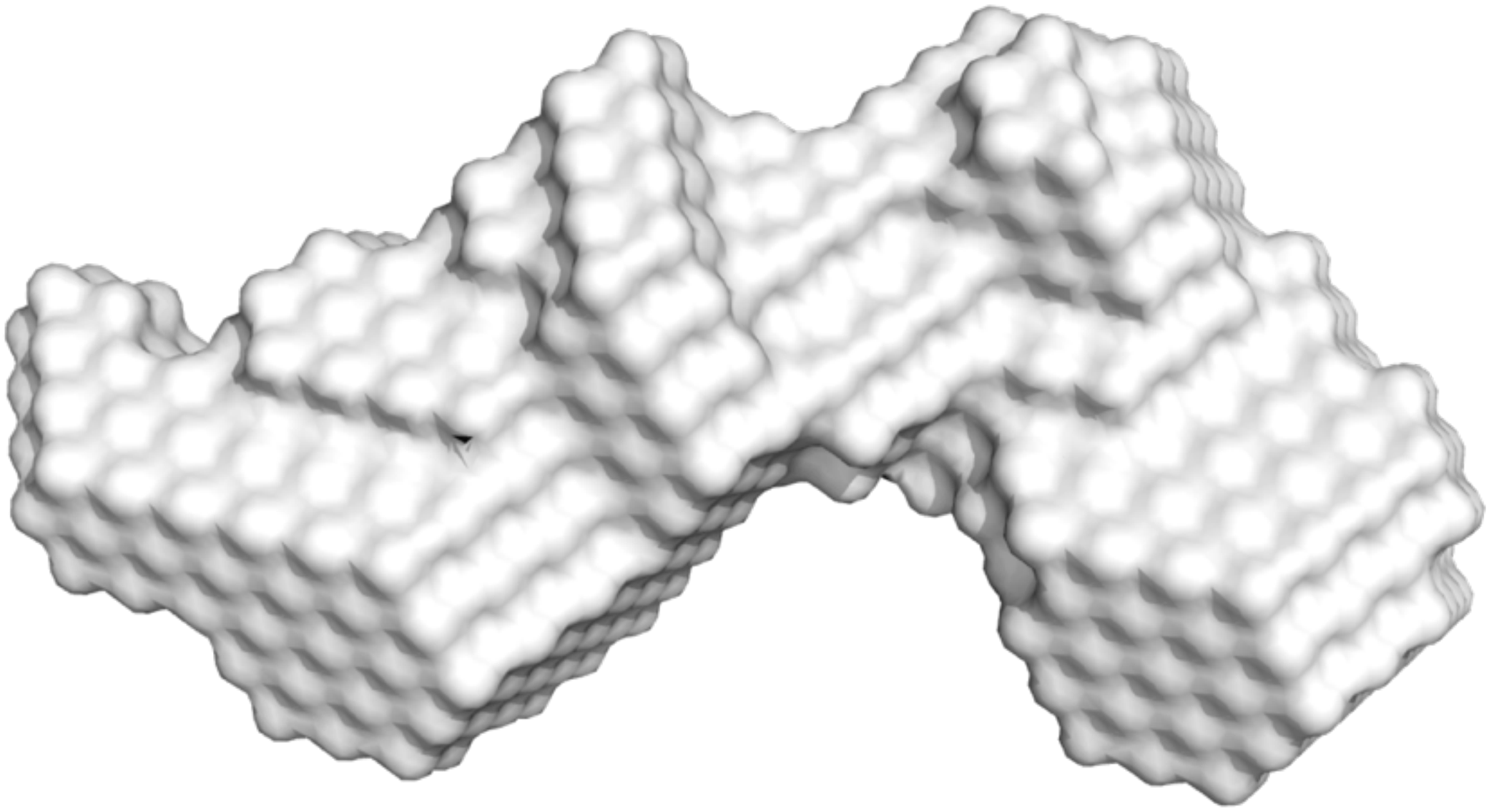
$$Score = \chi^2\,[I_{exp}(s), I_{calc}(s)] + \alpha P(x)$$

Envelope Reconstruction

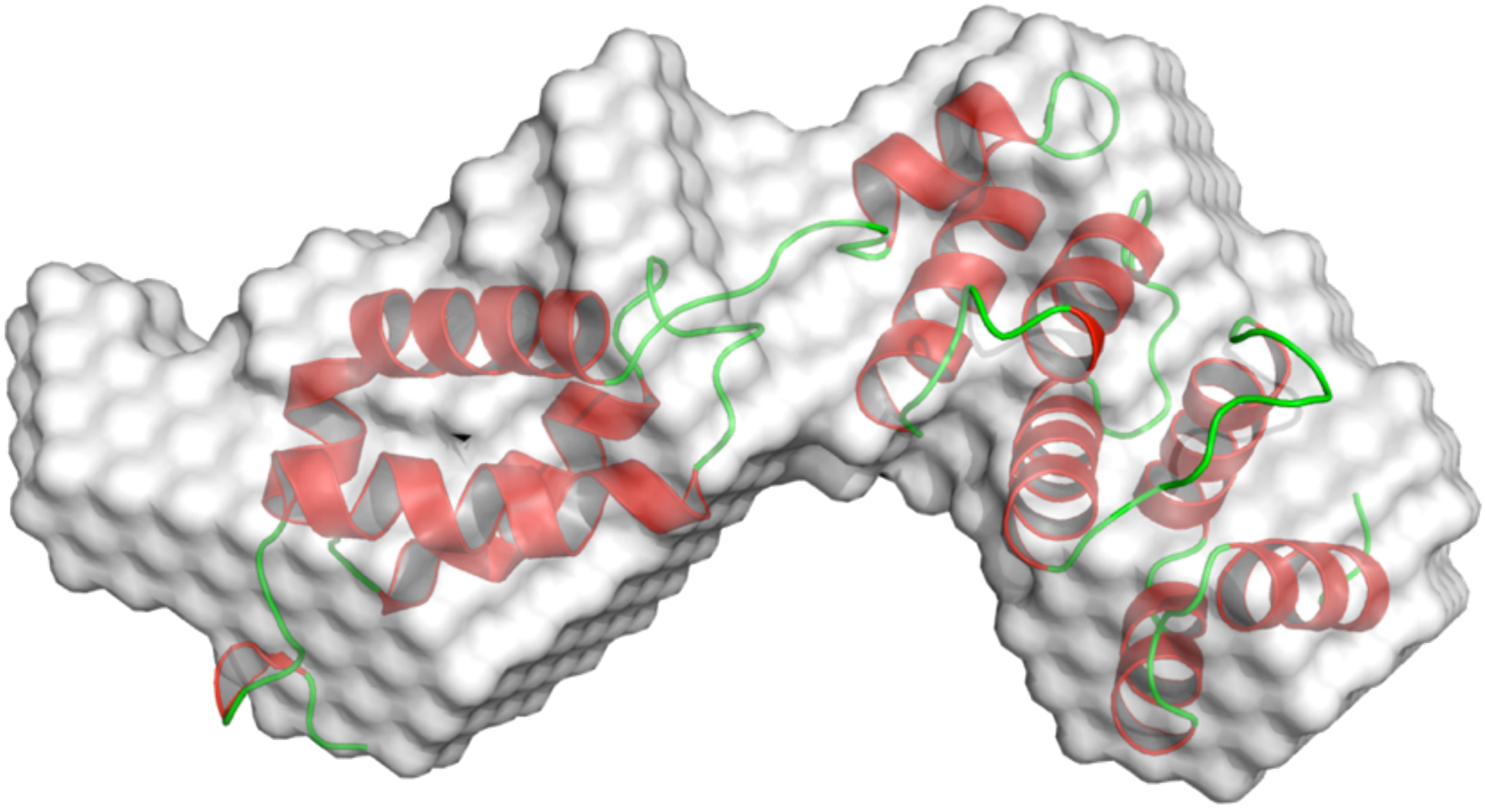
- Averaging with DAMAVER



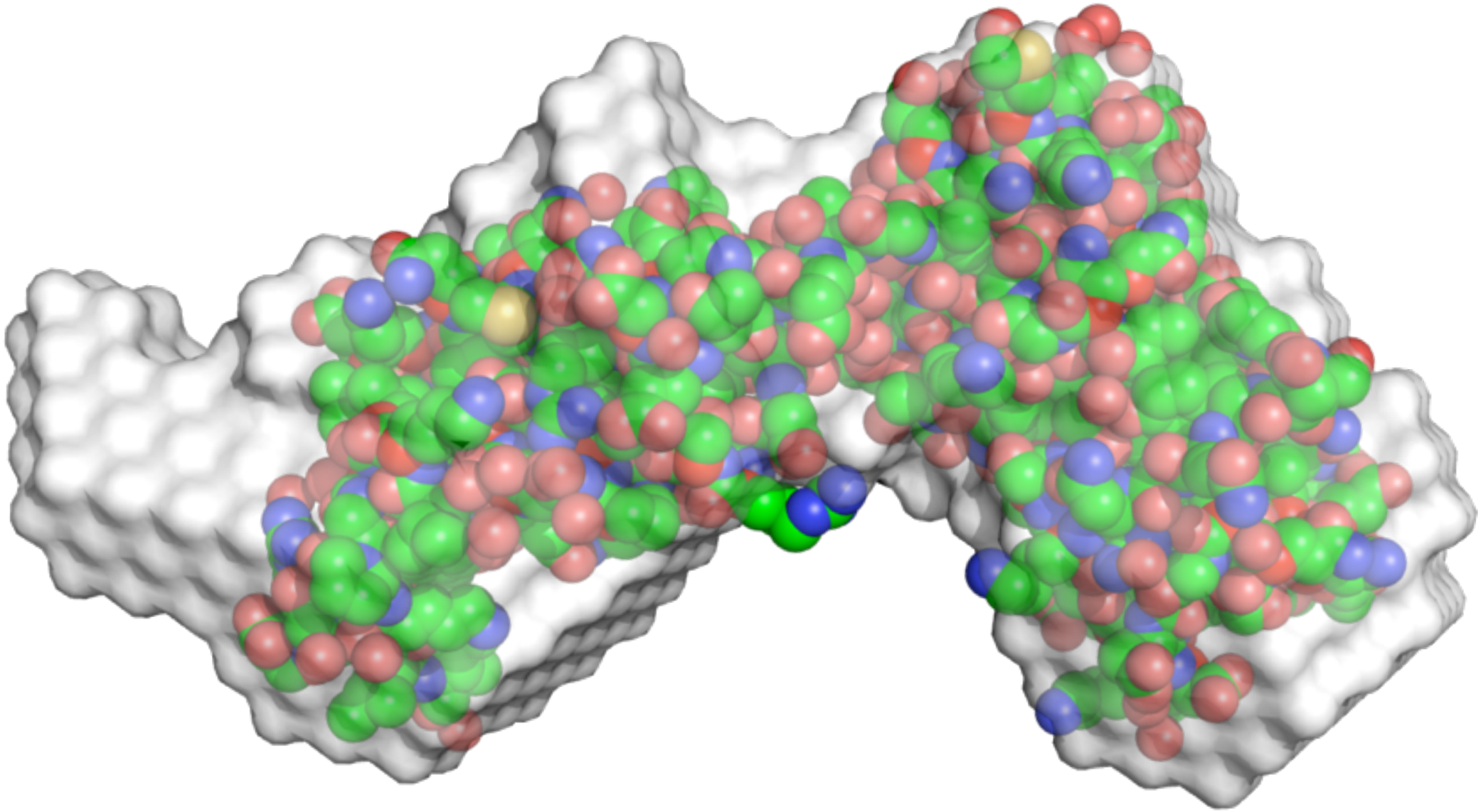
Example



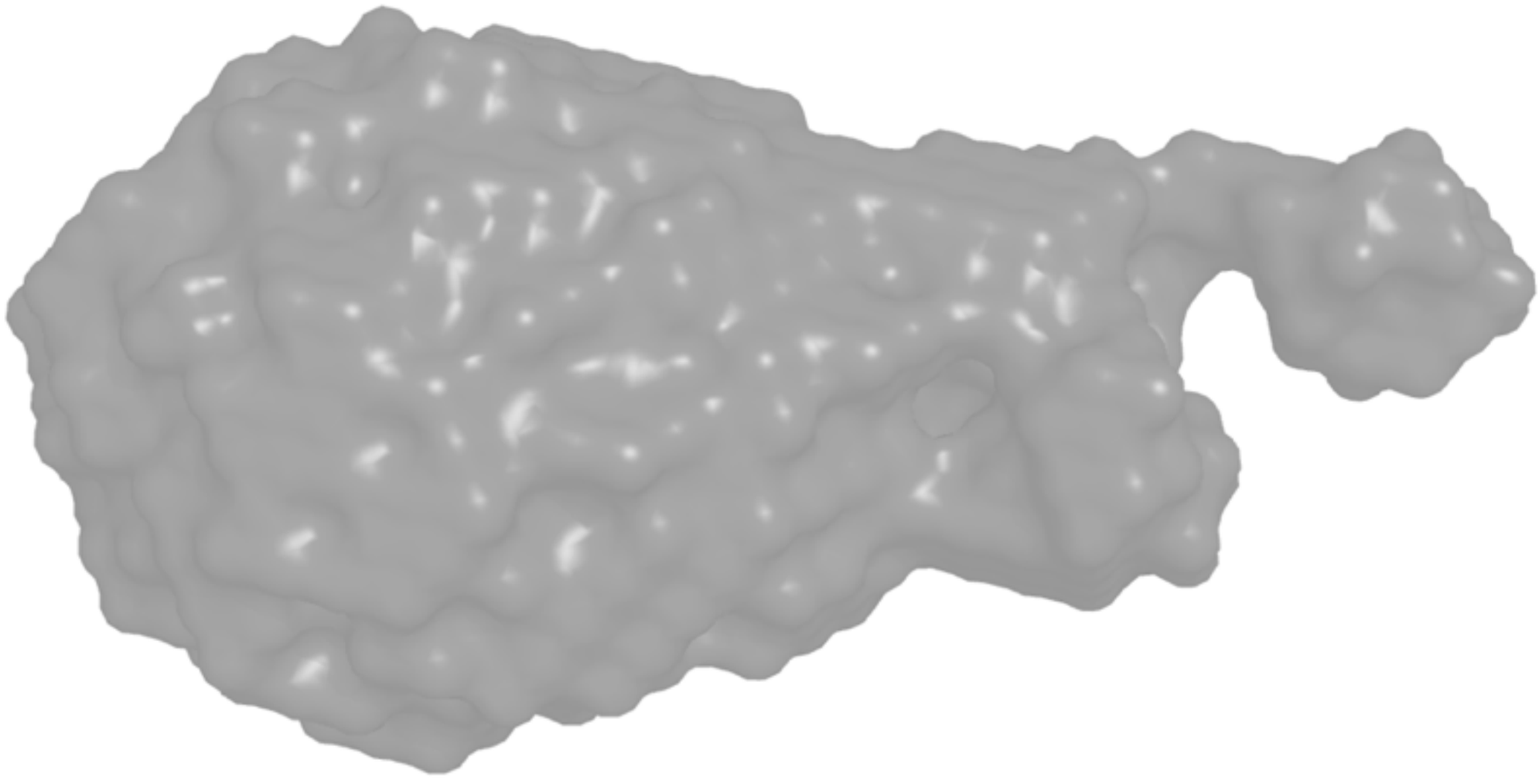
Example



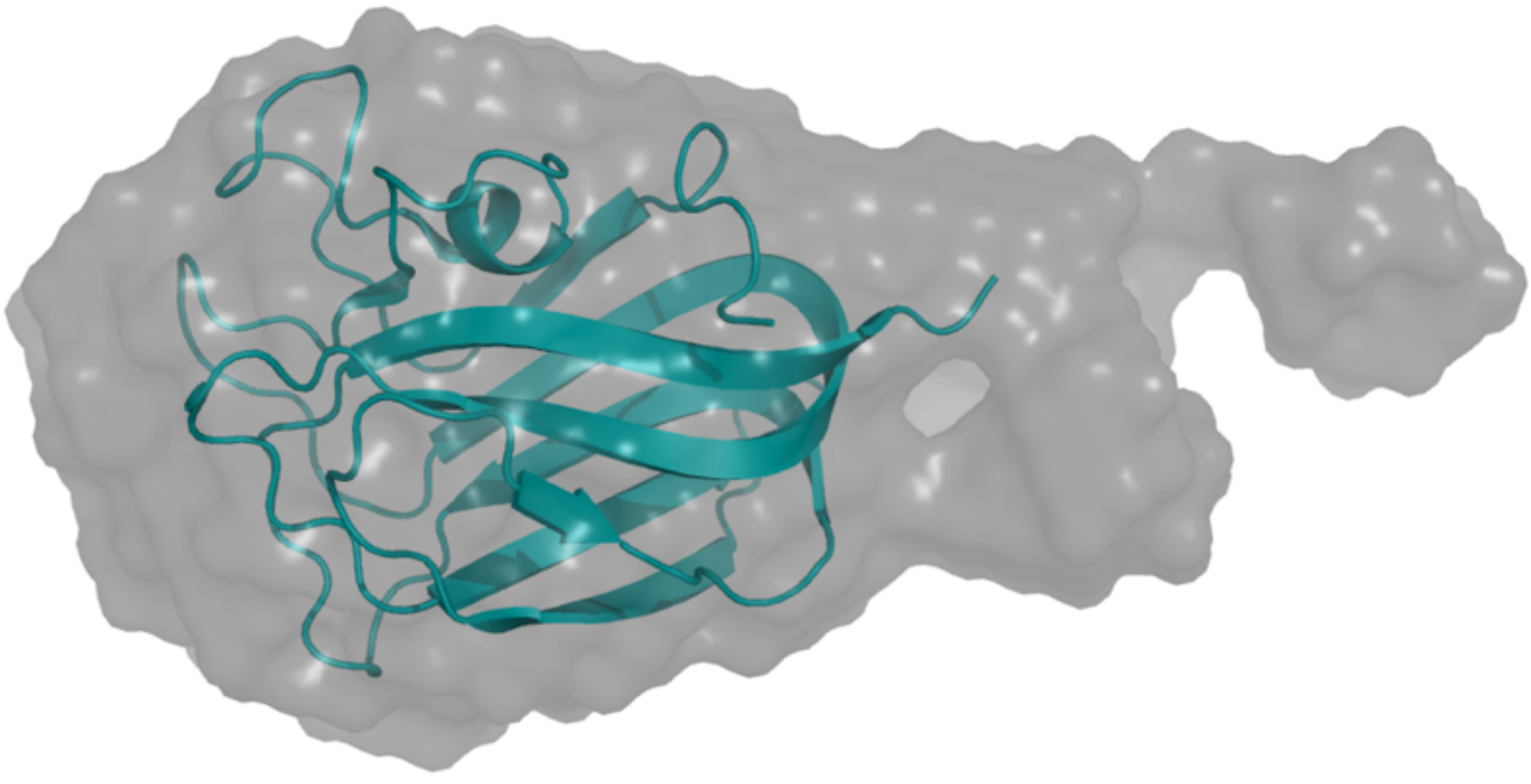
Example



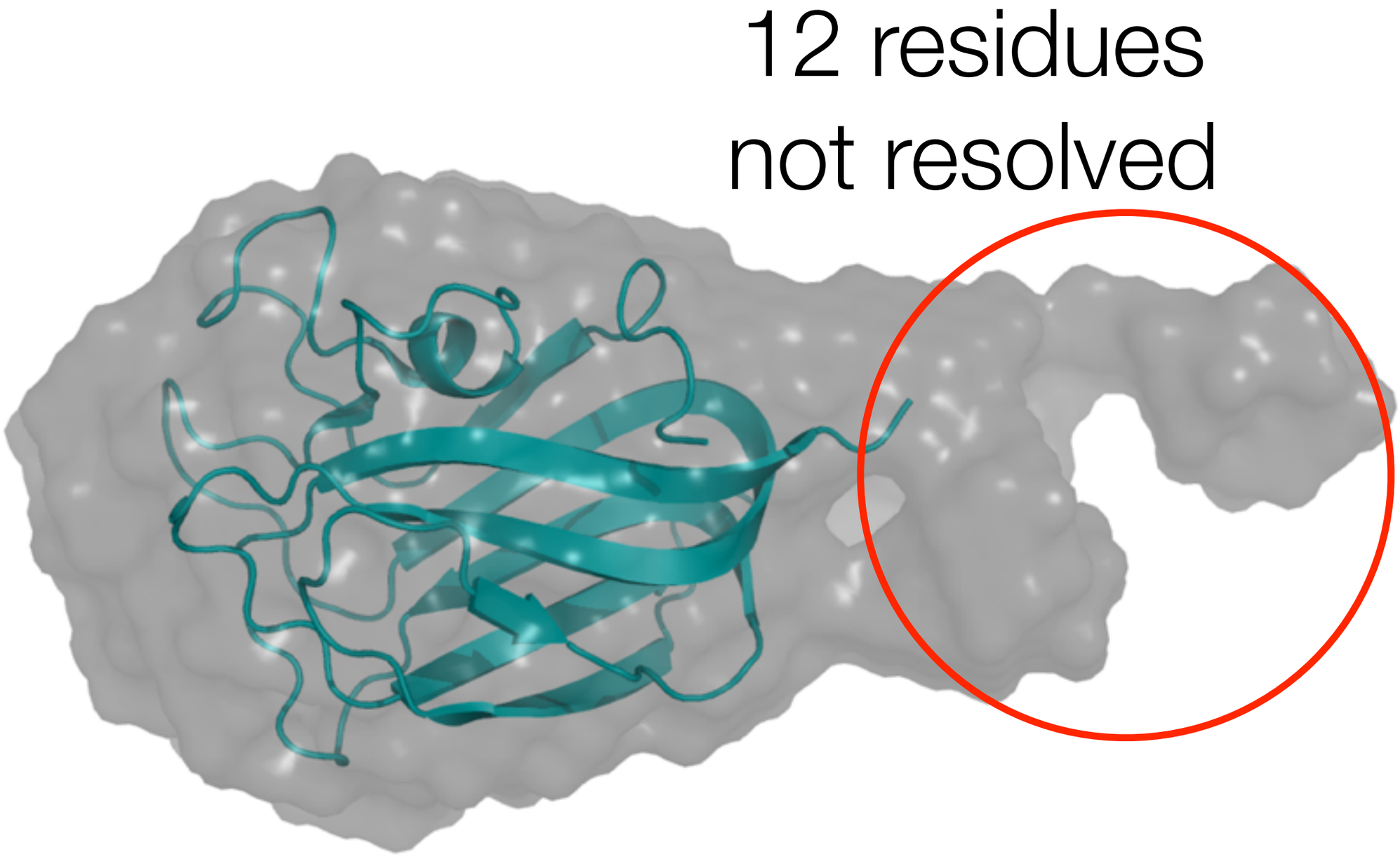
Examples



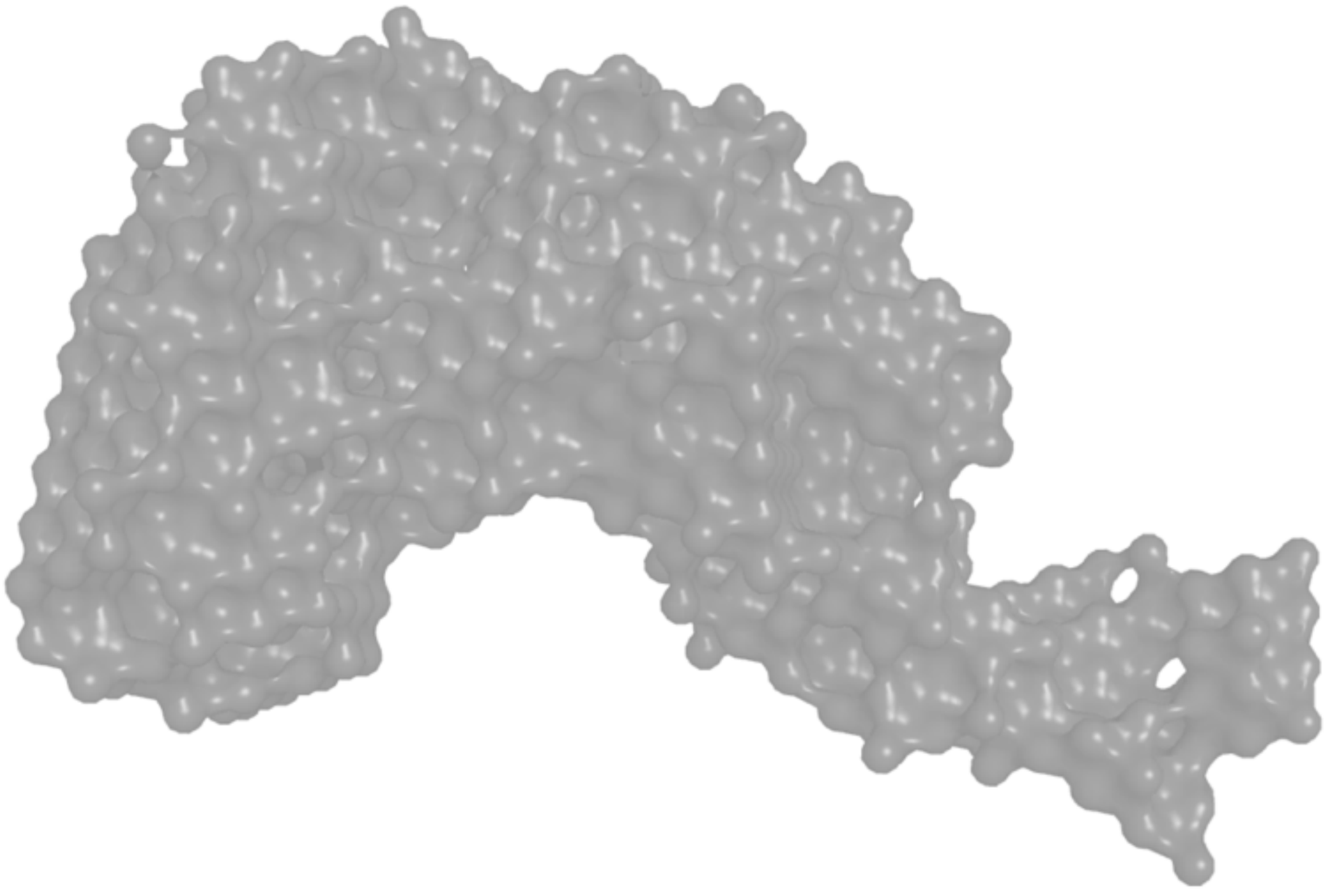
Examples



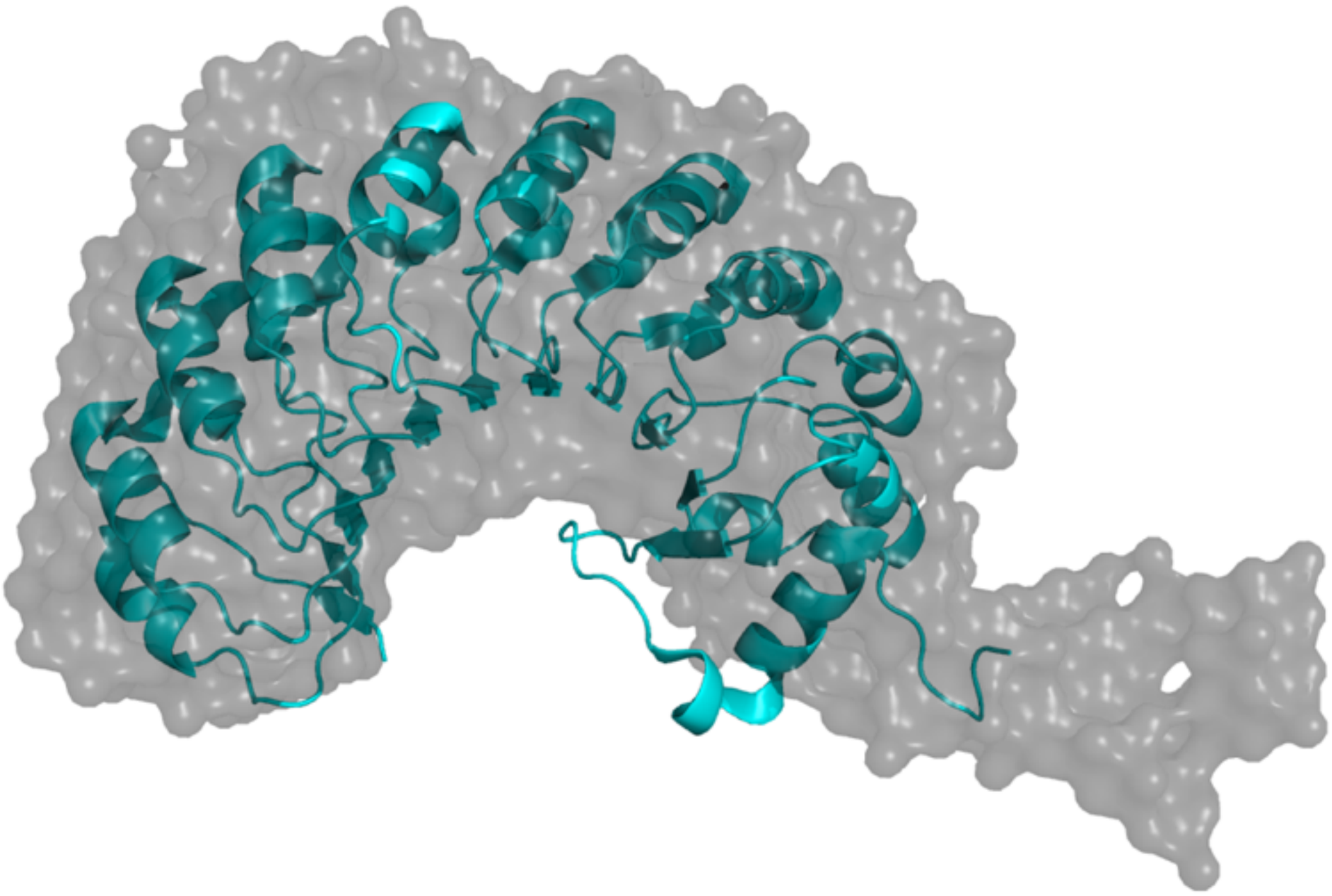
Examples



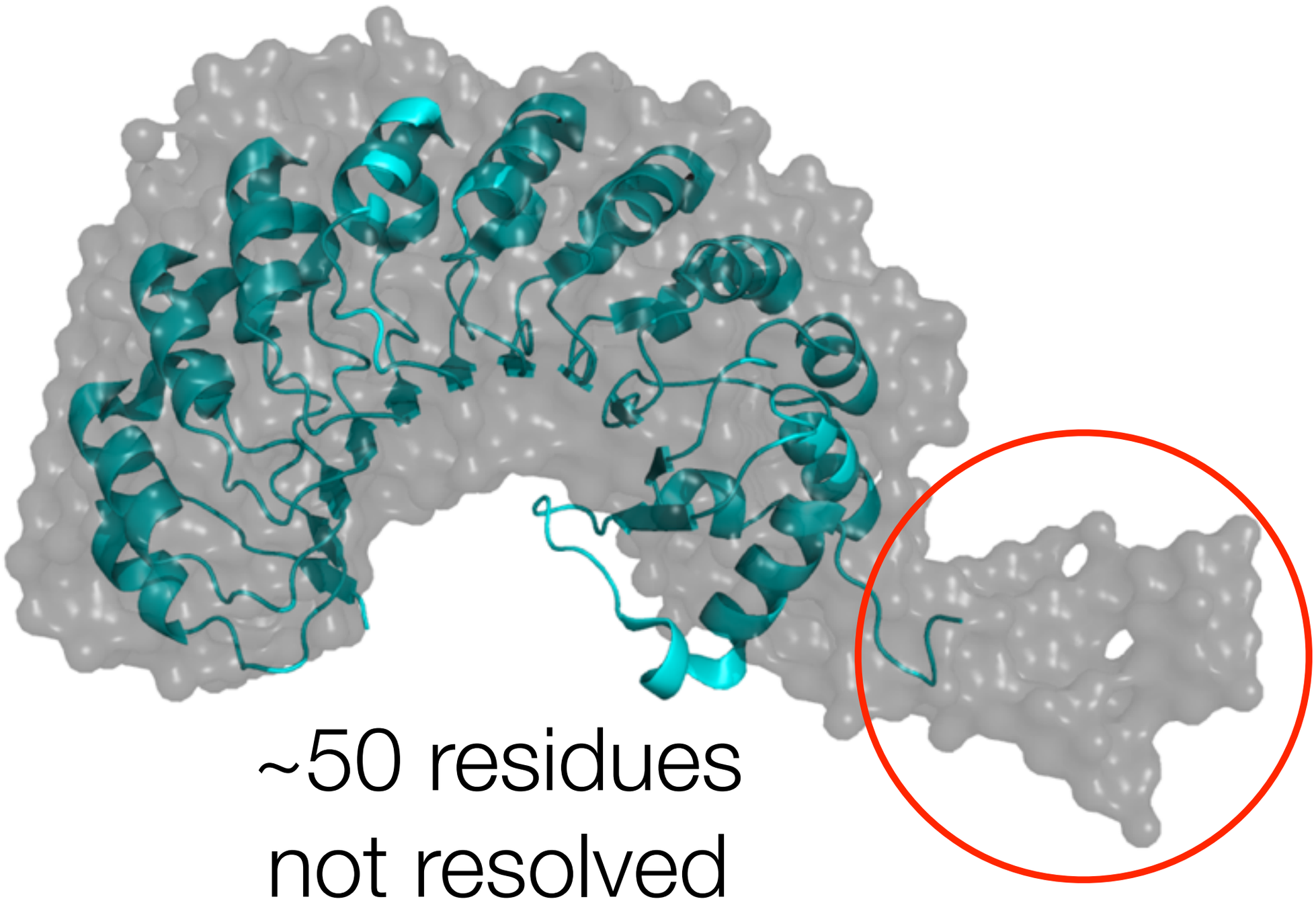
Examples



Examples



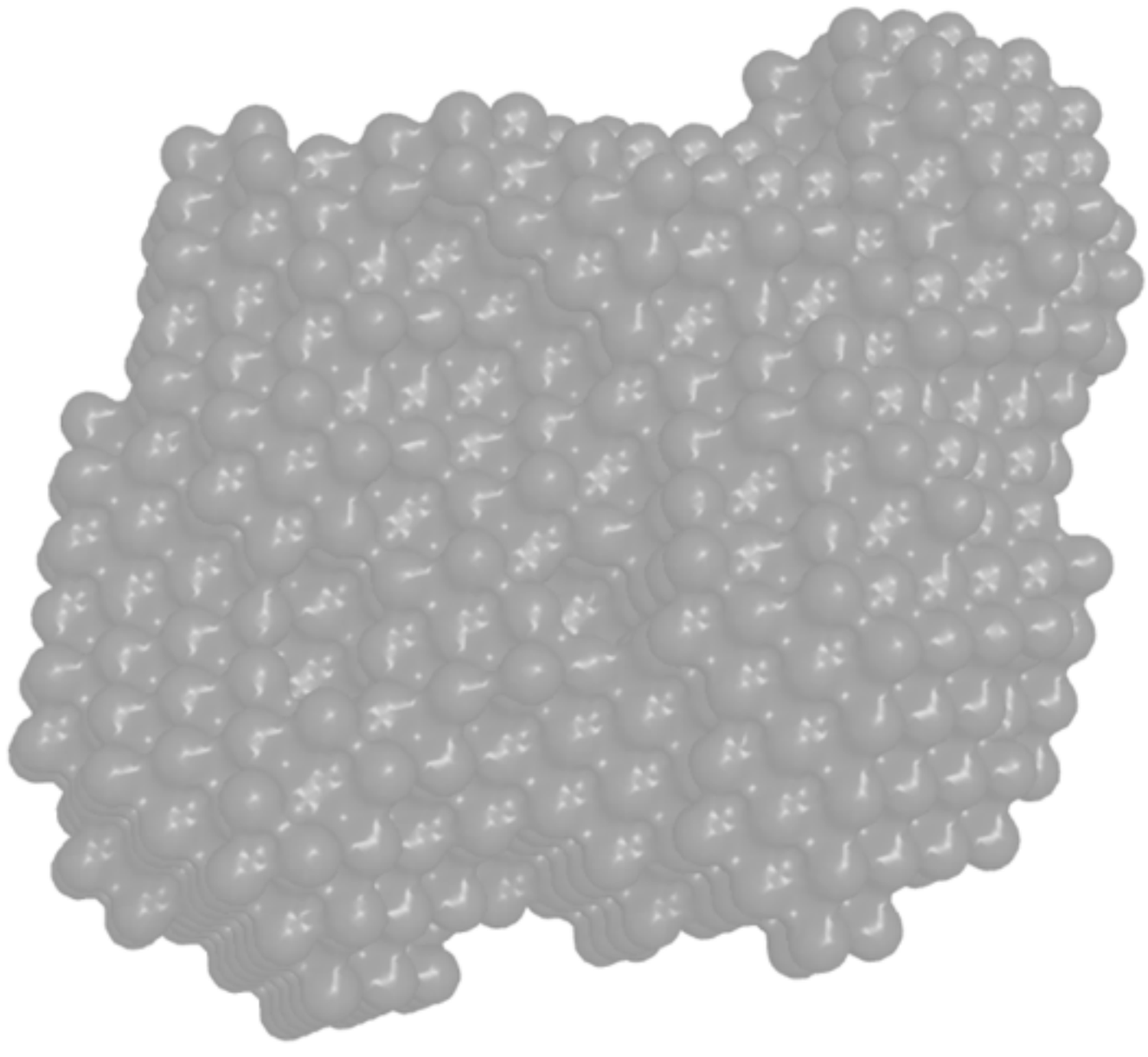
Examples



~50 residues
not resolved

Examples

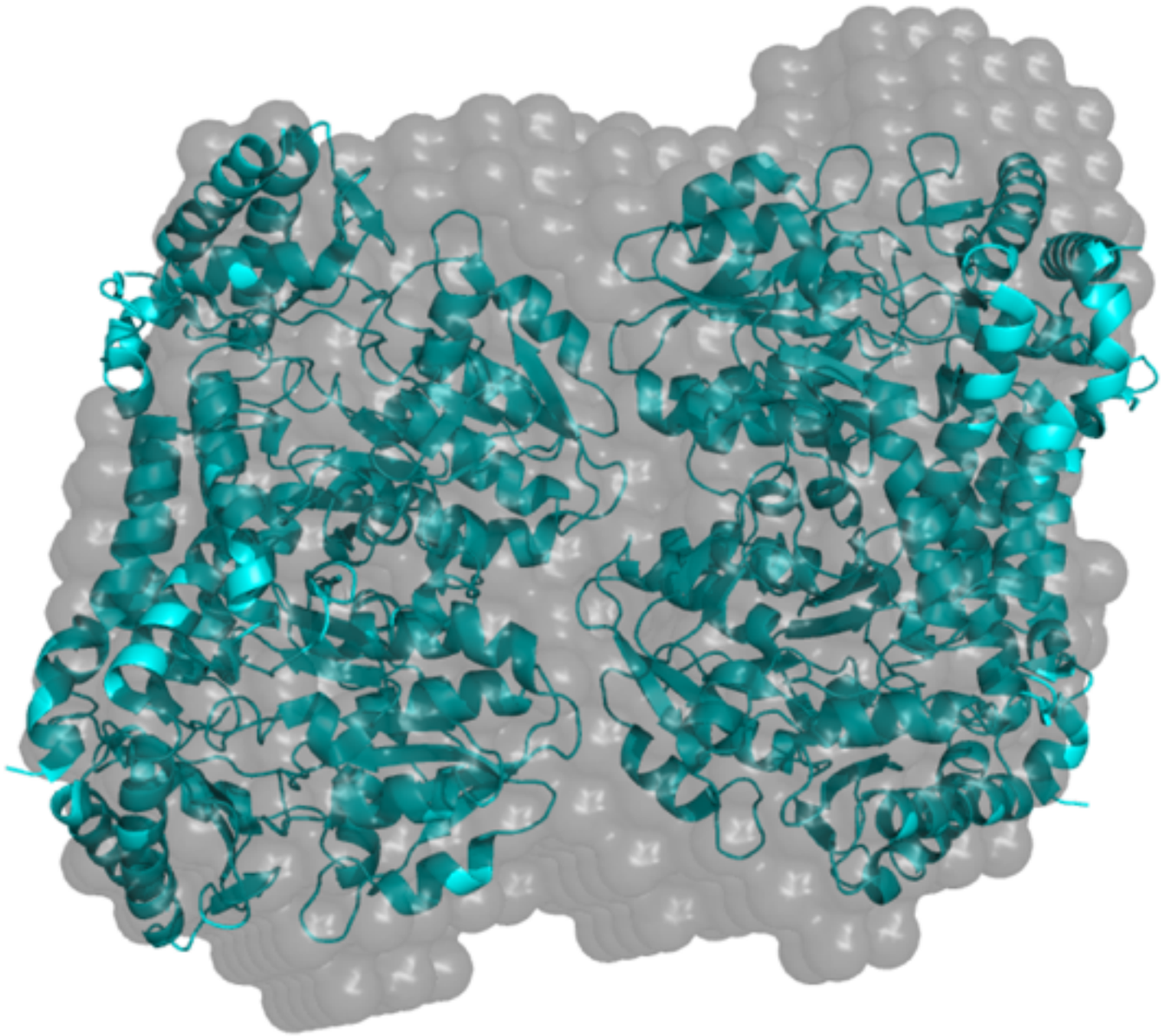
Structure of protein unknown



Examples

Structure of protein unknown

Dimer of homolog fits well



What question do you want SAXS to answer?

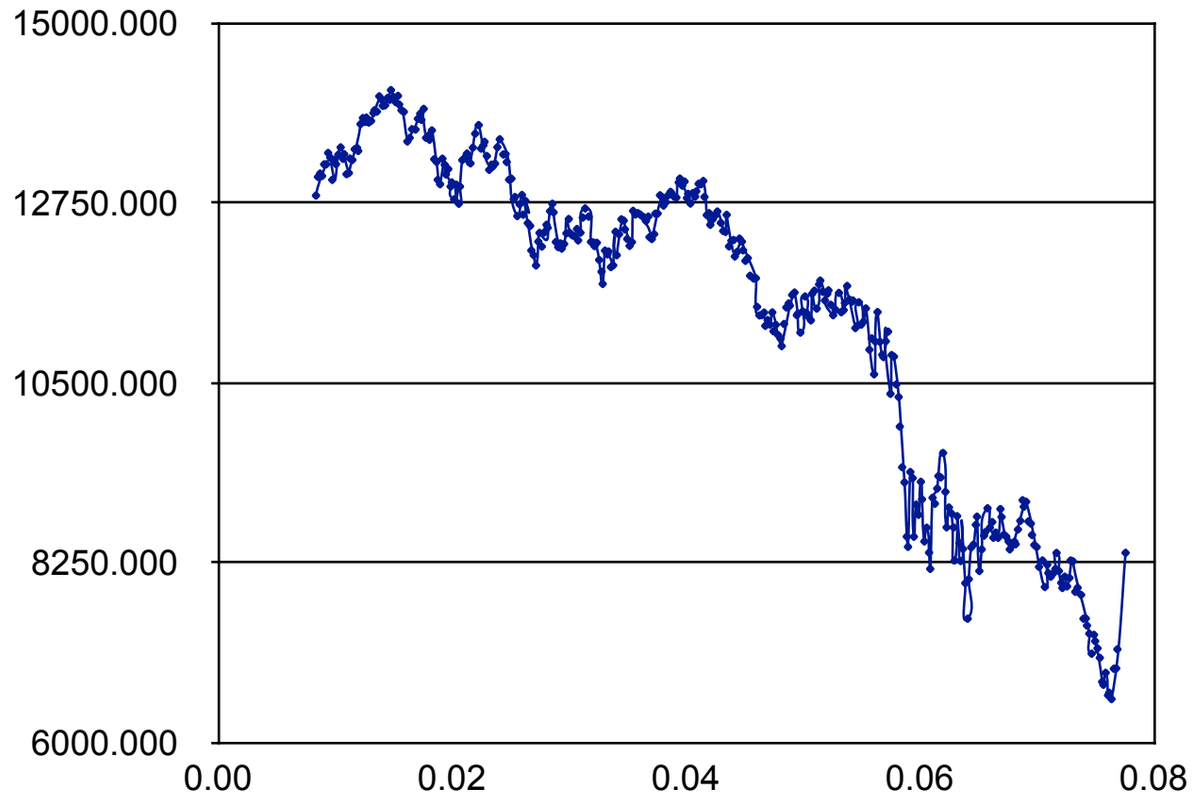
- Defining the question is fundamental to reliable conclusions
- Ask yes or no questions and decide if SAXS can provide an answer
- Question determines resolution and quality of the data that is needed, which can affect experimental setup
  - **Sample-detector distance** - size of particle vs resolution, Oligomers?
  - **Complexes** - molecular weight difference, what resolution?
  - **Effect of solution conditions** - buffer preparation? Dialysis? Number of concentrations? Serial dilution?
  - **Signal to noise** - Concentration? Exposure time?
    - Consider error propagation ($1^2 + 1^2 = 1.4^2$), i.e. twice the exposure doesn't yield twice the signal-to-noise

Wrap-Up

- Many SAXS analyses require monodispersity, so make sure you've got good quality data before trying to draw conclusions.
- SAXS "resolution" is ambiguous, not directly 2π/q. Resolution is really the ability to discriminate between models.
- While useful, don't read too much into envelopes.
- SAXS is a solution technique, so what's in solution is very important. Temperature, pH, or additives can alter your solution structure.
- Be sure to back up any conclusions you draw with other experimental evidence before publishing SAXS data.

Can we use X-ray solution scattering?

Slide courtesy of Eddie Snell

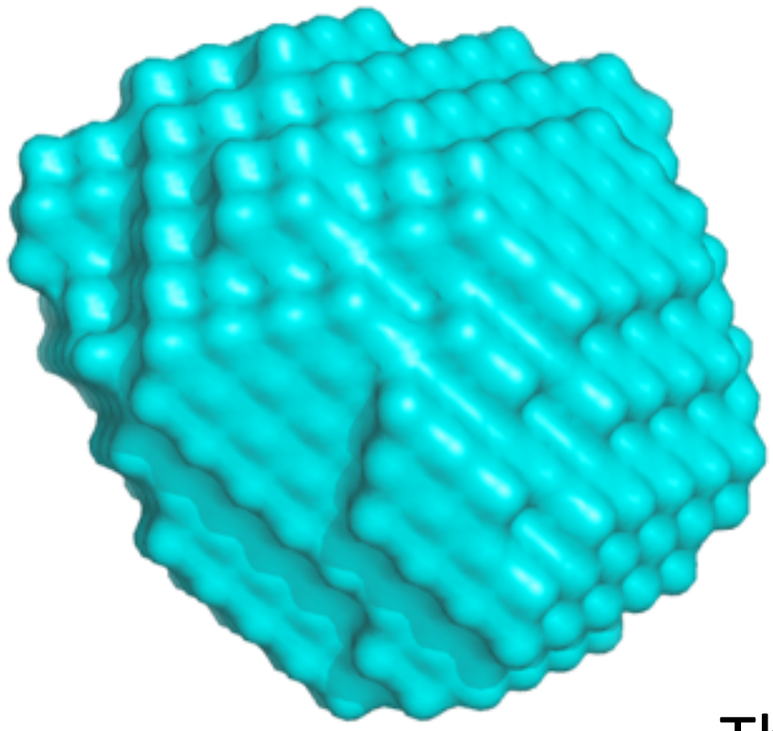


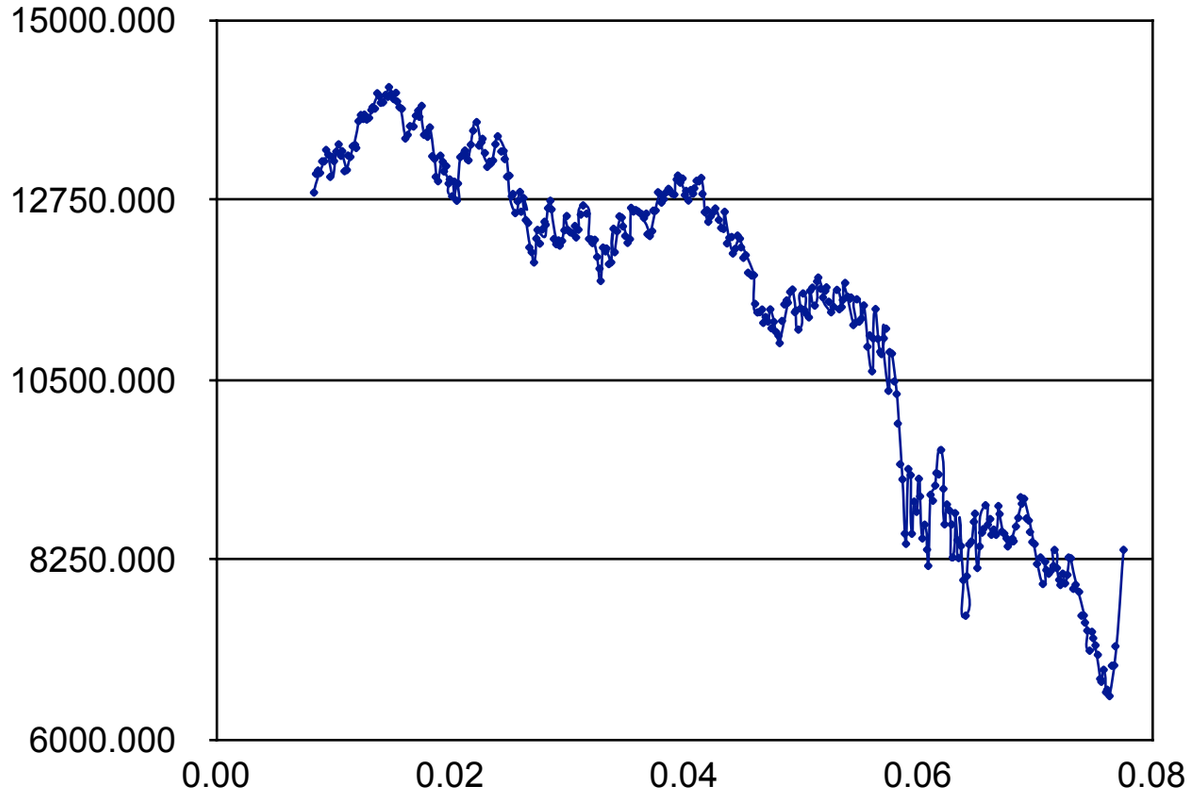
The scattering data from SAXS provides a 1D Fourier transform of the envelope of the particle.

It’s possible to fit multiple envelopes to the data.

Can we use X-ray solution scattering?

Slide courtesy of Eddie Snell



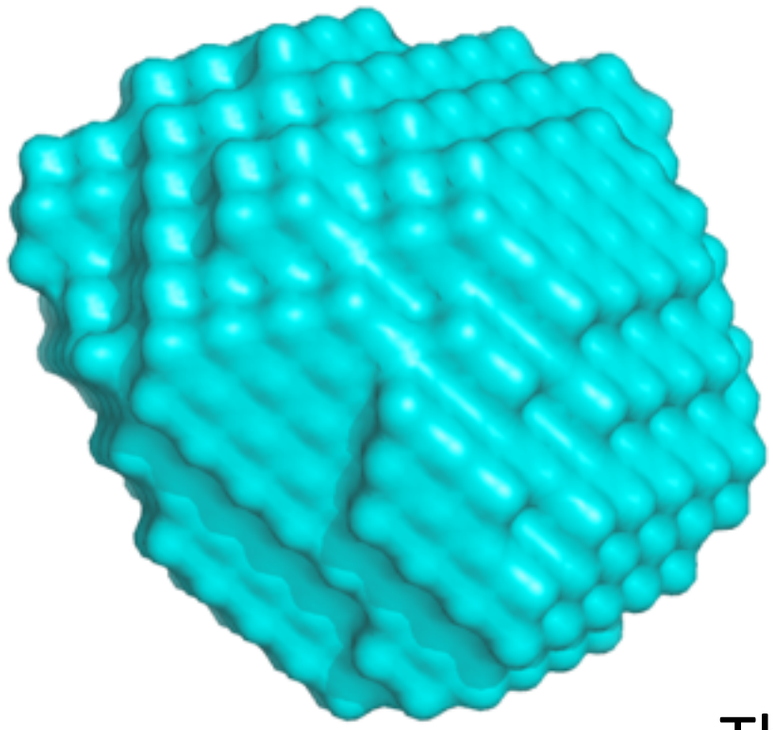


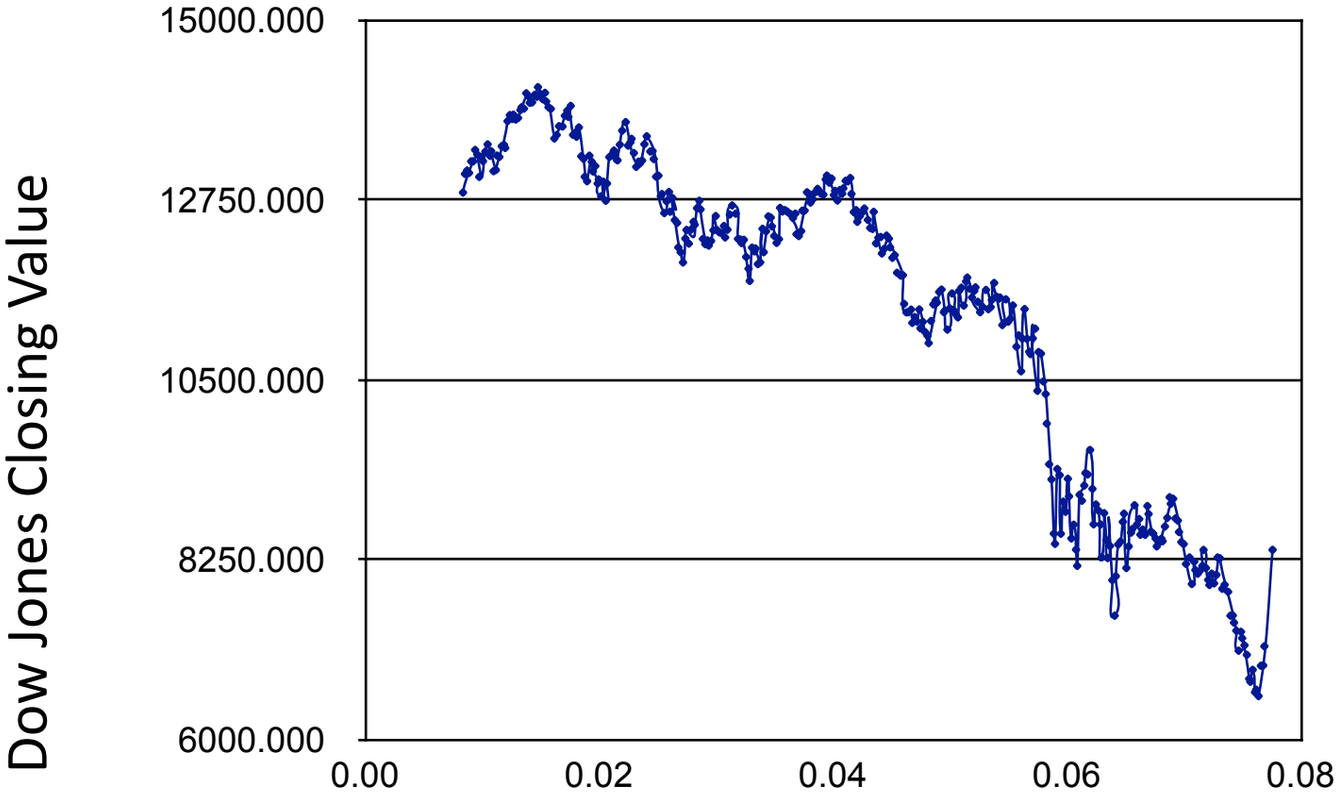
The scattering data from SAXS provides a 1D Fourier transform of the envelope of the particle.

It’s possible to fit multiple envelopes to the data.

Can we use X-ray solution scattering?

Slide courtesy of Eddie Snell





Aug 2007 March 2009

The scattering data from SAXS provides a 1D Fourier transform of the envelope of the particle.

It's possible to fit multiple envelopes to the data.

You will always get an envelope despite the data!