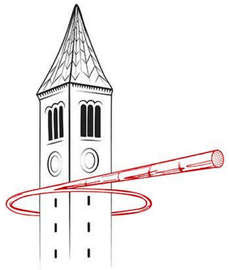
CHESS
CORNELL HIGH ENERGY
SYNCHROTRON SOURCE



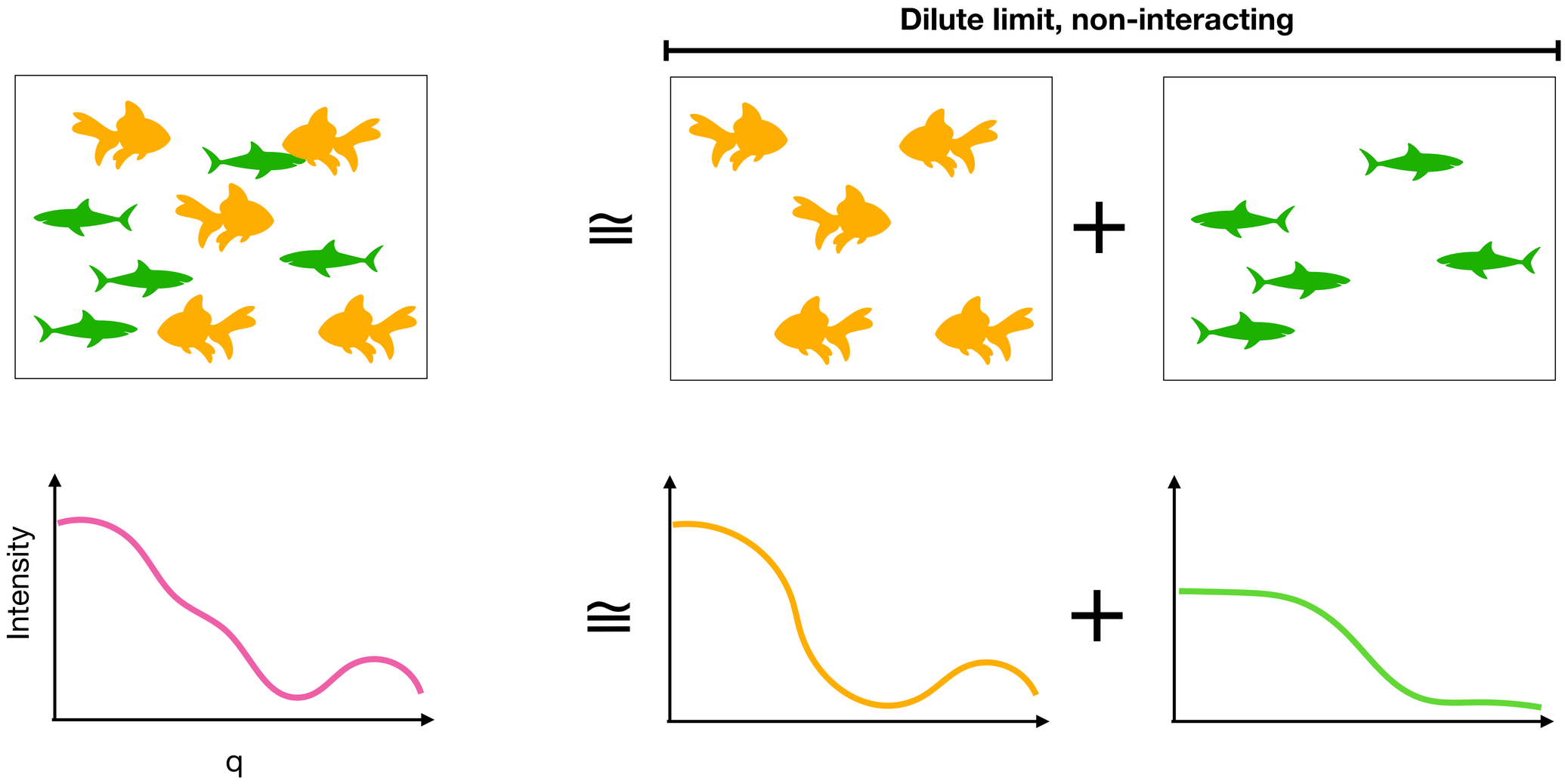
Extreme
Biophysics
RCN

# Advanced analysis

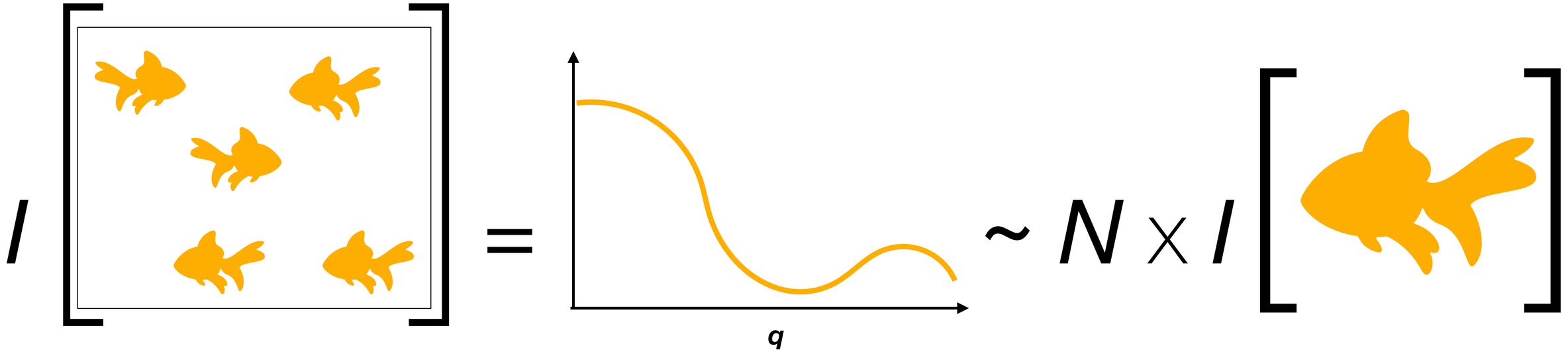
## Atomistic modeling, shape reconstruction, and SEC-SAXS

**Steve Meisburger — Ando Lab @ Cornell — CHESS HP Bio Workshop 2021 — April 29 2021**

# SAXS reports average intensity of a mixture



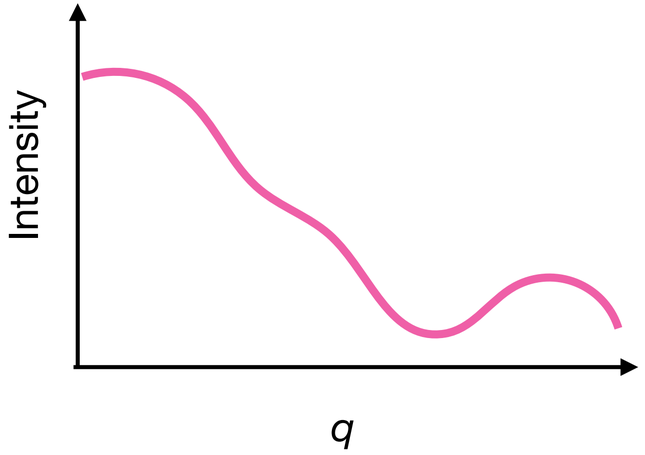
# Advanced modeling requires *monodispersity*



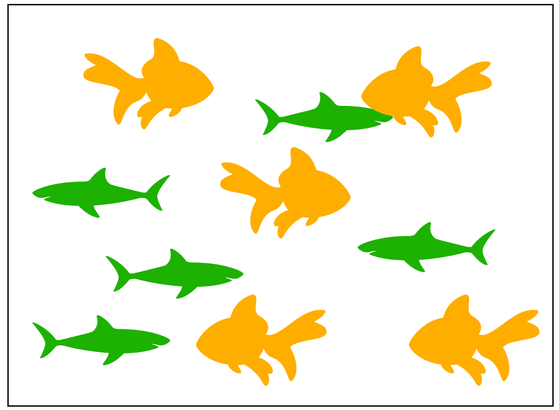

**A monodisperse, dilute solution of *N* fish**

**One fish**

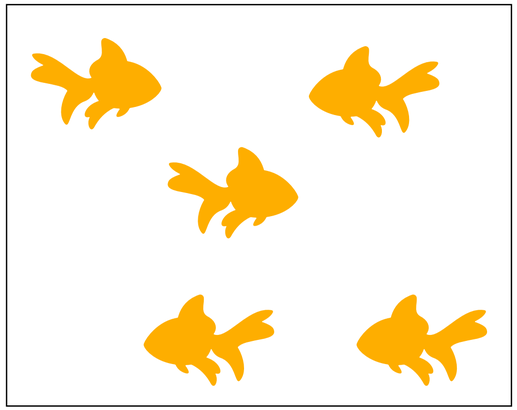
# Advanced modeling requires *monodispersity**




*Is it*



*or*



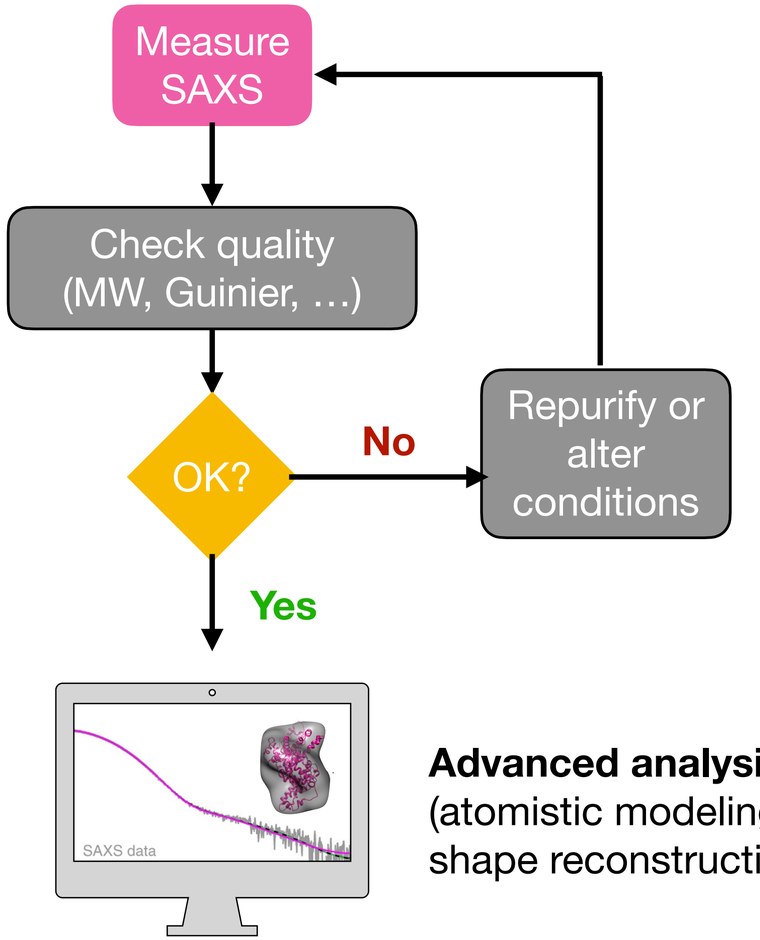
*?*

- MW estimates
- Guinier plot
- Concentration series
- Complementary biophysical methods

***methods exist for modeling multiple structures simultaneously from mixtures, but must take extra care to avoid overfitting.**

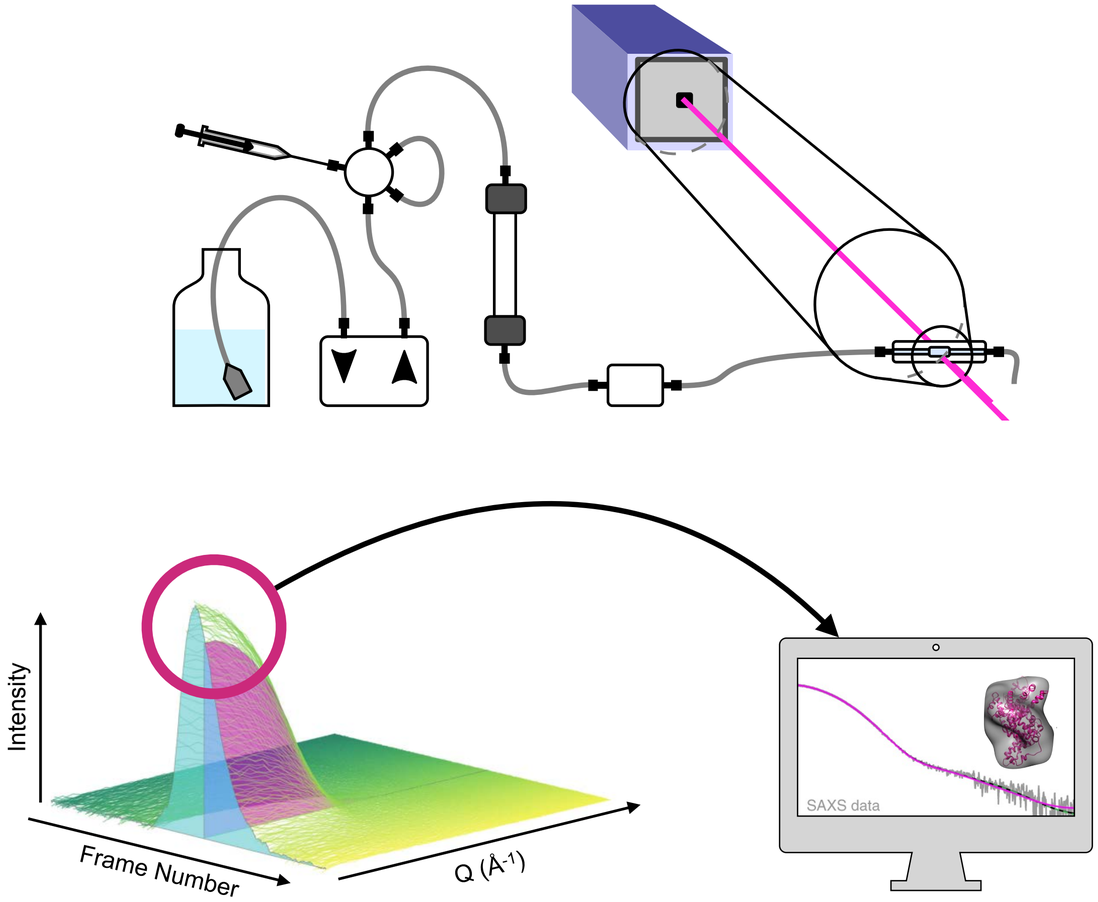
# Two routes to monodispersity

## Batch mode

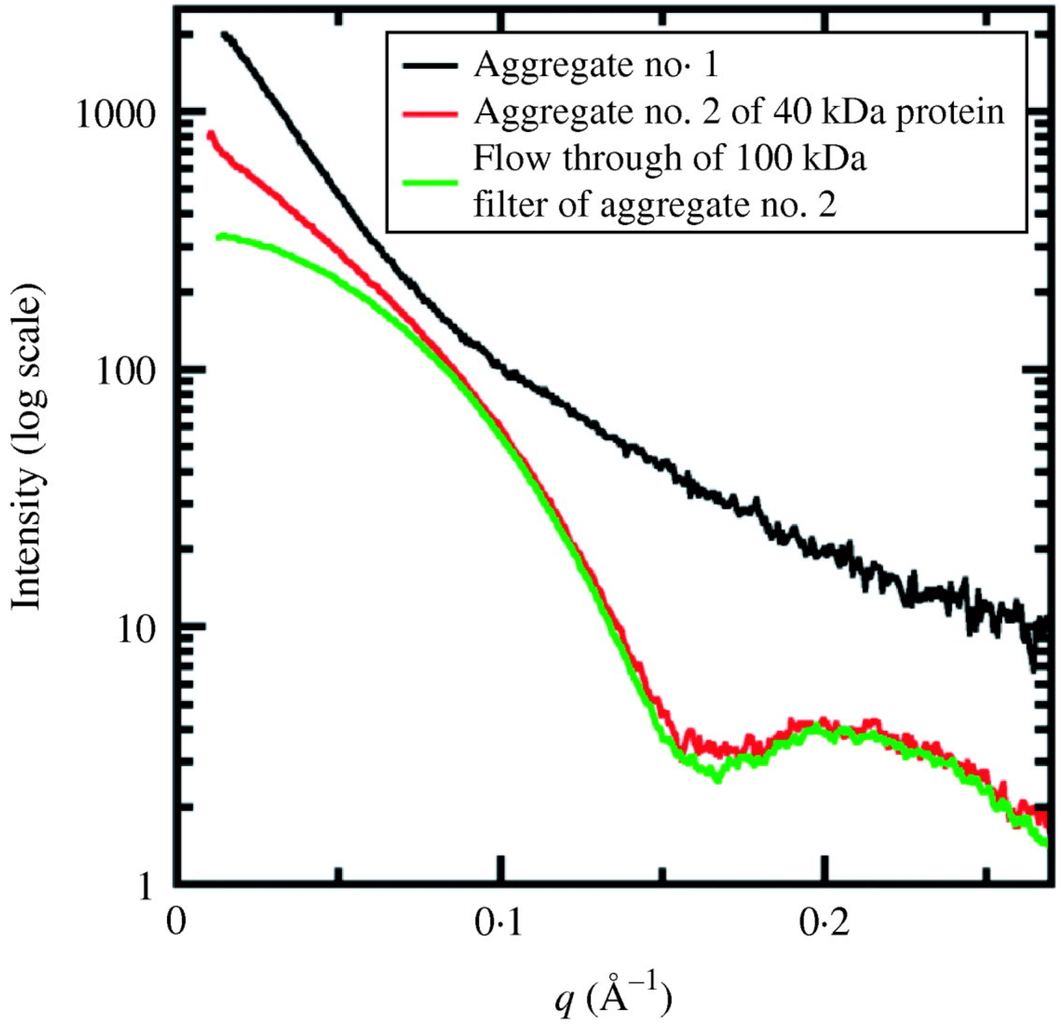



**Advanced analysis**
(atomistic modeling,
shape reconstruction, …)

## In-line chromatography (SEC-SAXS)



# Troubleshooting aggregation in batch mode




Putnam, Hammel, Hura, & Tainer. Q Rev. Biophys. 2007

# Advanced analysis overview

1. Compare SAXS data with simulated scattering from atomic models

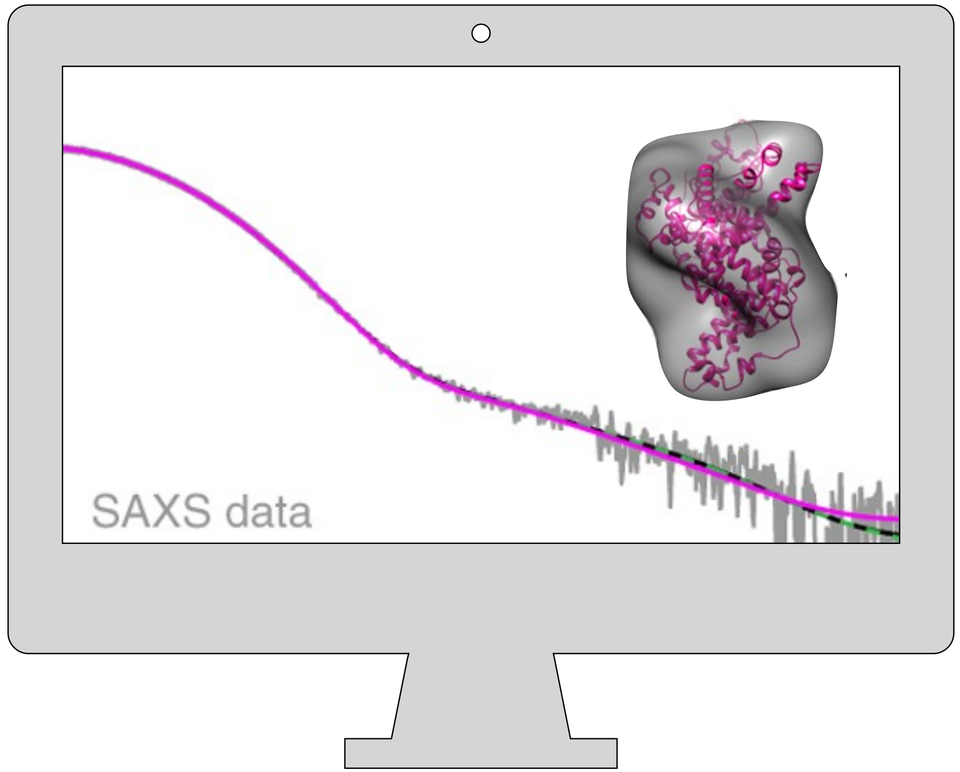
   (CRYSOL, FoXS, …)

2. *Ab-initio* shape reconstruction

   (DAMMIN, GASBOR, DENSS, …)

3. Hybrid techniques: docking, flexible fitting, ensemble modeling, MD simulation, …

   (SASREF, GLOBSYMM, BUNCH, CORAL, …)



# Compare SAXS data with simulated scattering from atomic models

Experimental SAXS profile

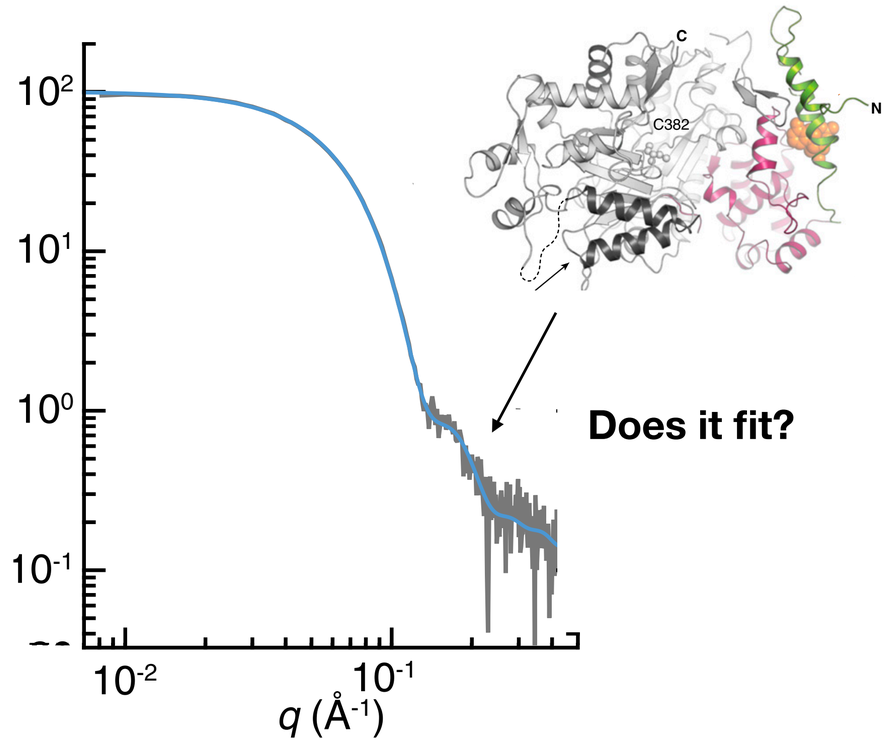
Candidate structure (PDB file)

Fit

**Scattering simulation**
(CRYSOL, FoXS, …)

**hydration layer**

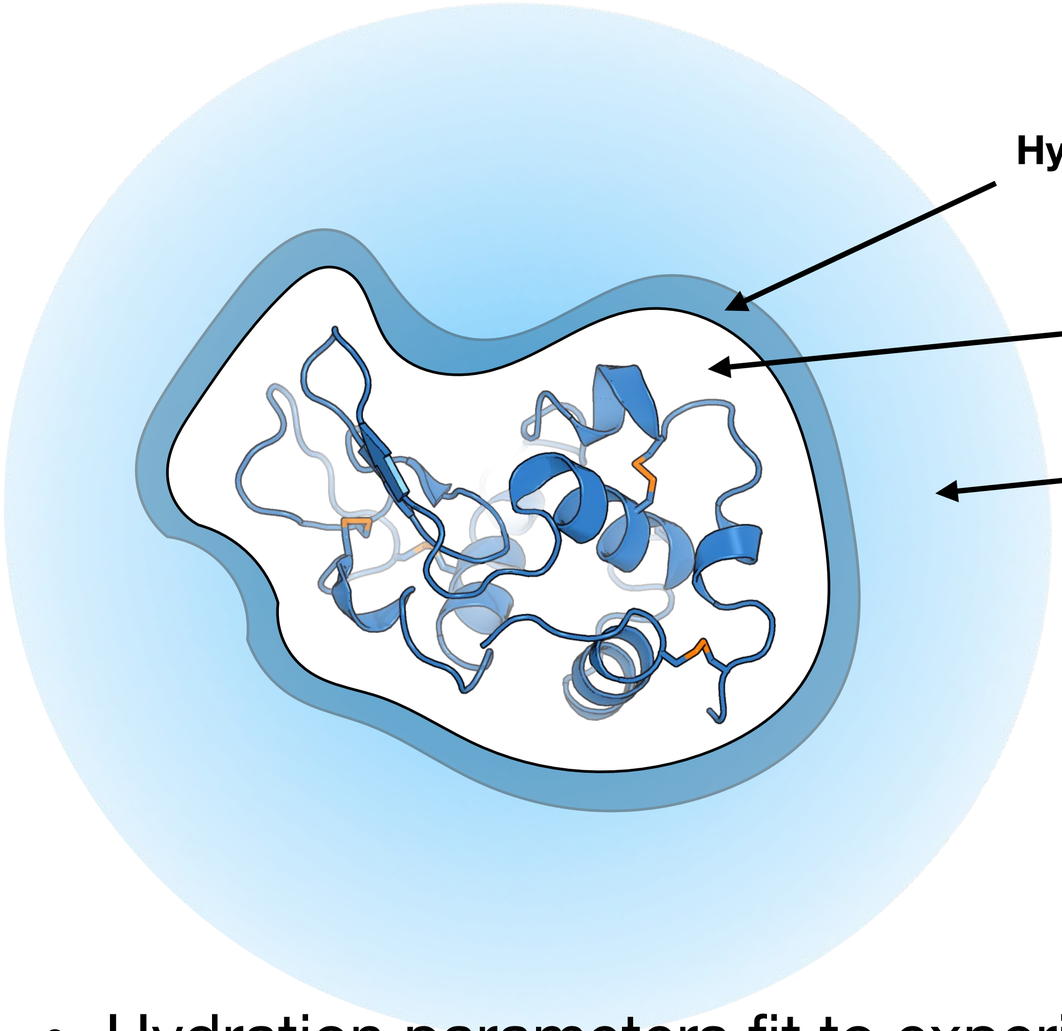
- X-ray crystallography
- Cryo-EM
- homology model
- molecular dynamics

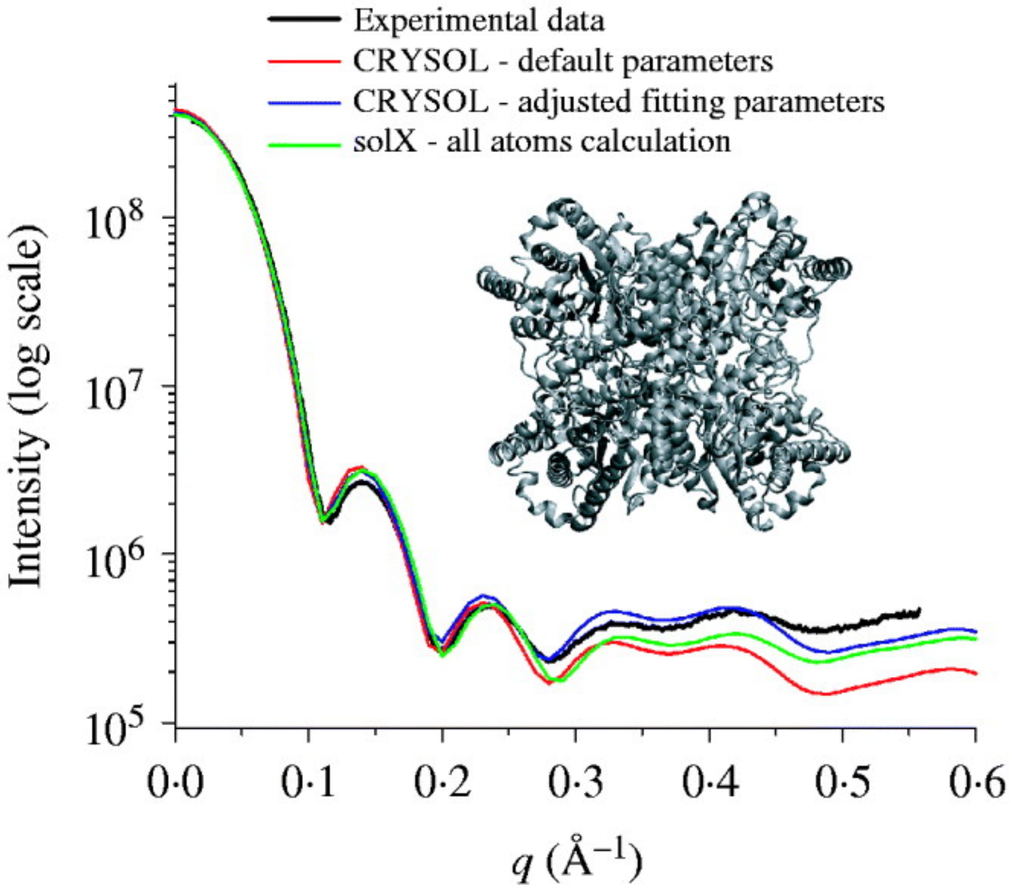



**Important: model must be <u>complete</u> (need to add missing loops, termini, his tags, etc)**

Parker, M. J. et al. PNAS 115, E4594–E4603 (2018).

# Accurate scattering simulation requires hydration layer modeling

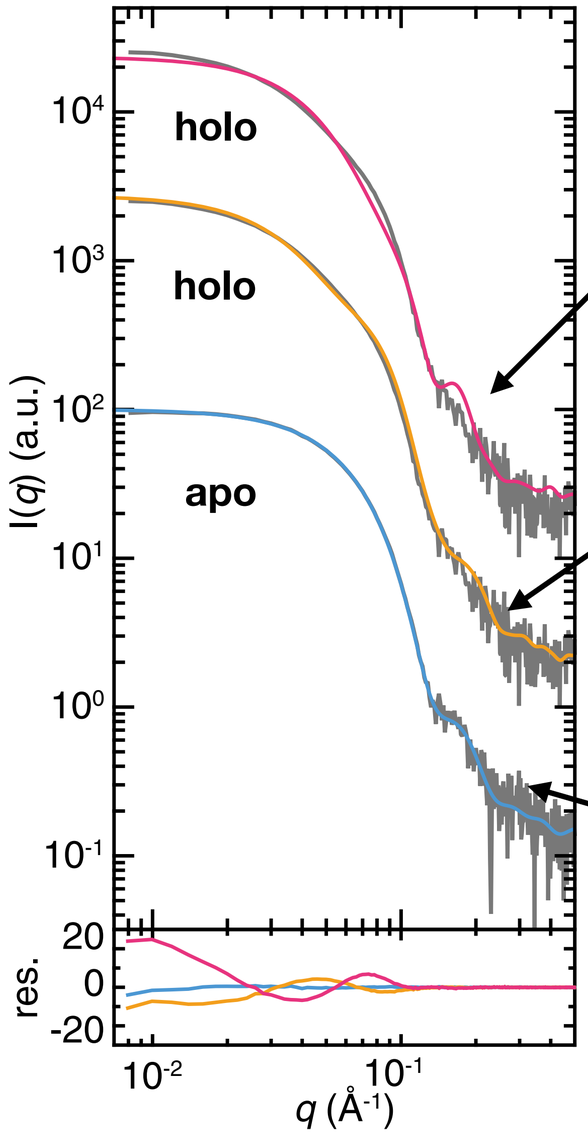


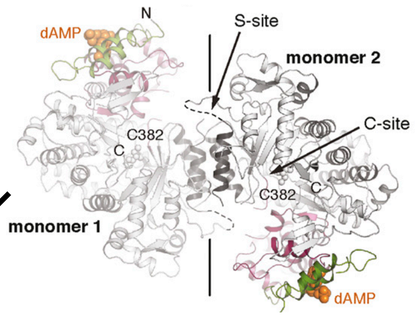




Putnam, Hammel, Hura, & Tainer. Q Rev. Biophys. 2007

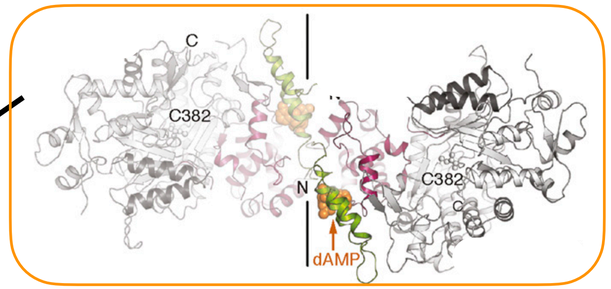
- Hydration parameters fit to experimental data (CRYSOL, FoXS, …)
- WAXS requires greater accuracy (all-atom MD, …)

# Simulation is best for ranking multiple candidates

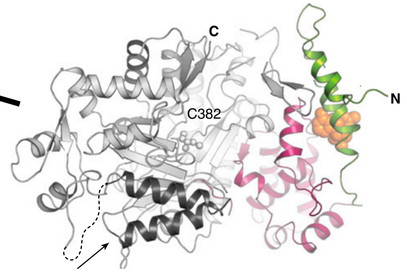



**Canonical dimer (poor fit)**




**Non-canonical dimer (good fit)**




**Monomer (good fit to apo)**

**Examine residuals!**

Parker, M. J. et al. PNAS 115, E4594–E4603 (2018).

# Software for scattering simulation

- **CRYSOL** — fit uniform hydration model
  - Download ATSAS package: https://www.embl-hamburg.de/biosaxs/software.html
- **FoXS** — fit bead-based hydration model
  - Web server: https://modbase.compbio.ucsf.edu/foxs/
- **WAXSIS** — short all-atom MD simulations (no fitting)
  - Web server: http://waxsis.uni-goettingen.de/
- [**HyPred**, **3D-RISM**, **AquaSAXS**, and many more …]

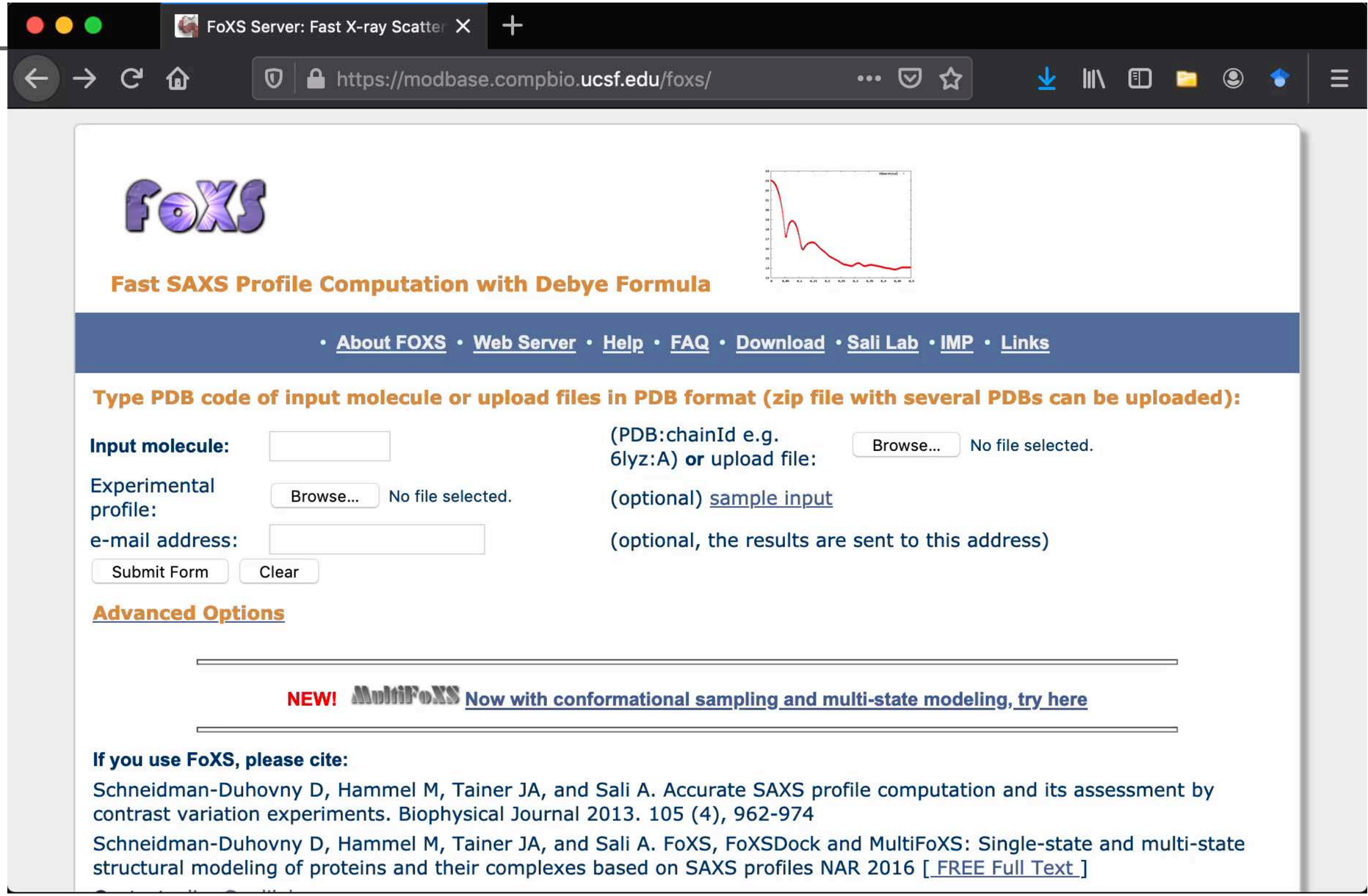
CRYSOL — Svergun, Barberato & Koch. J Appl Cryst, 1995.
foXS — Schneidman-Duhovny, Hammel, Tainer, & Sali. NAR, 2016.
WAXSIS — Christopher J. Knight and Jochen S. Hub. NAR, 2015.
HyPred — Virtanen, Makowski, Sosnick, & Freed. Biophys J, 2011.
3D-RISM — Nguyen, Pabit, Meisburger, Pollack, & Case. J Chem Phys, 2014.
AquaSAXS — Poitevin, Orland, Doniach, Koehl, & Delarue. NAR, 2011.

# FOXS demo

**FoXS reference:**

Schneidman-Duhovny, Hammel, Tainer, & Sali. NAR, 2016.

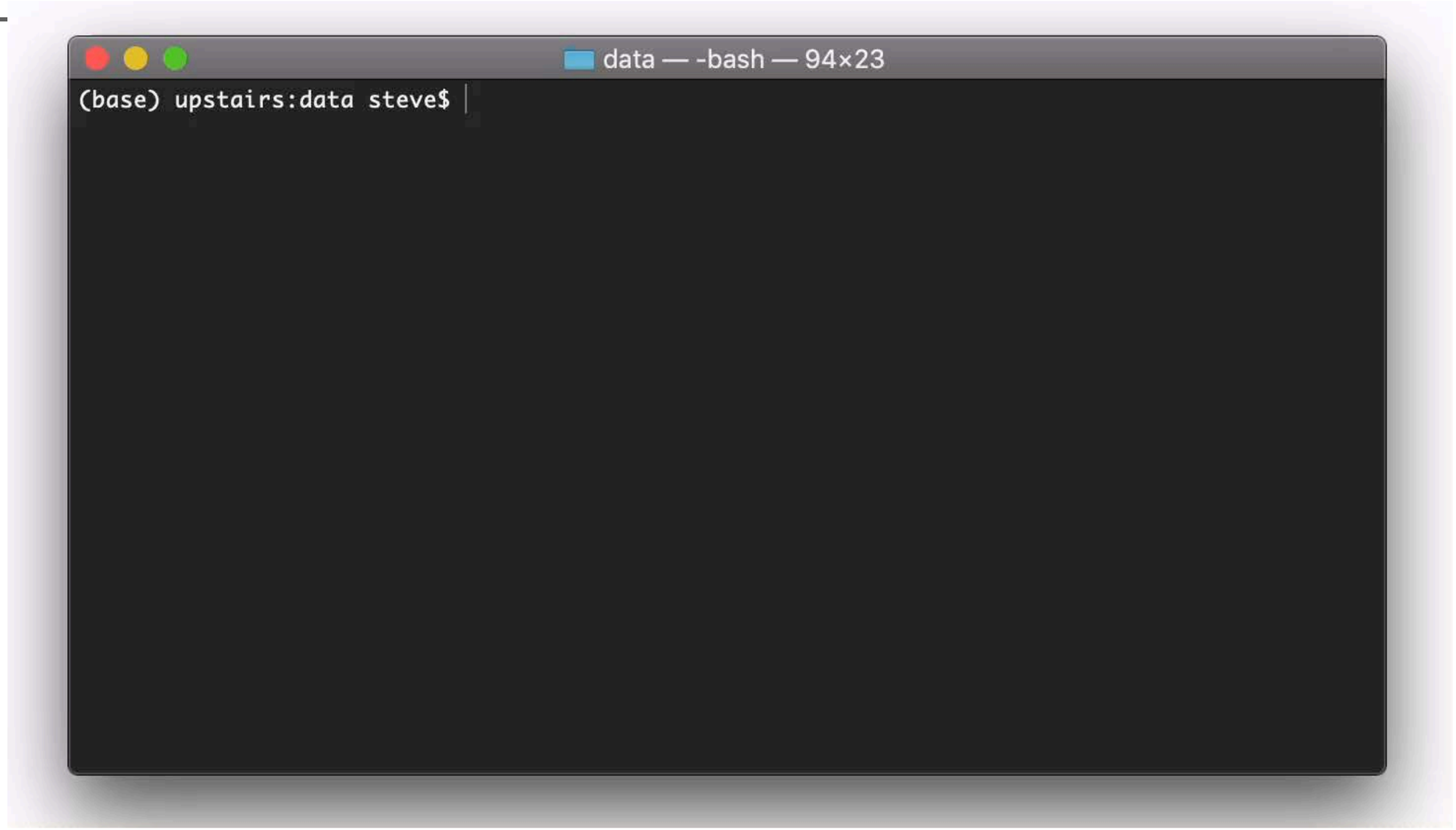
https://modbase.compbio.ucsf.edu/foxs/



# CRYSOL demo

**CRYSOL reference:**

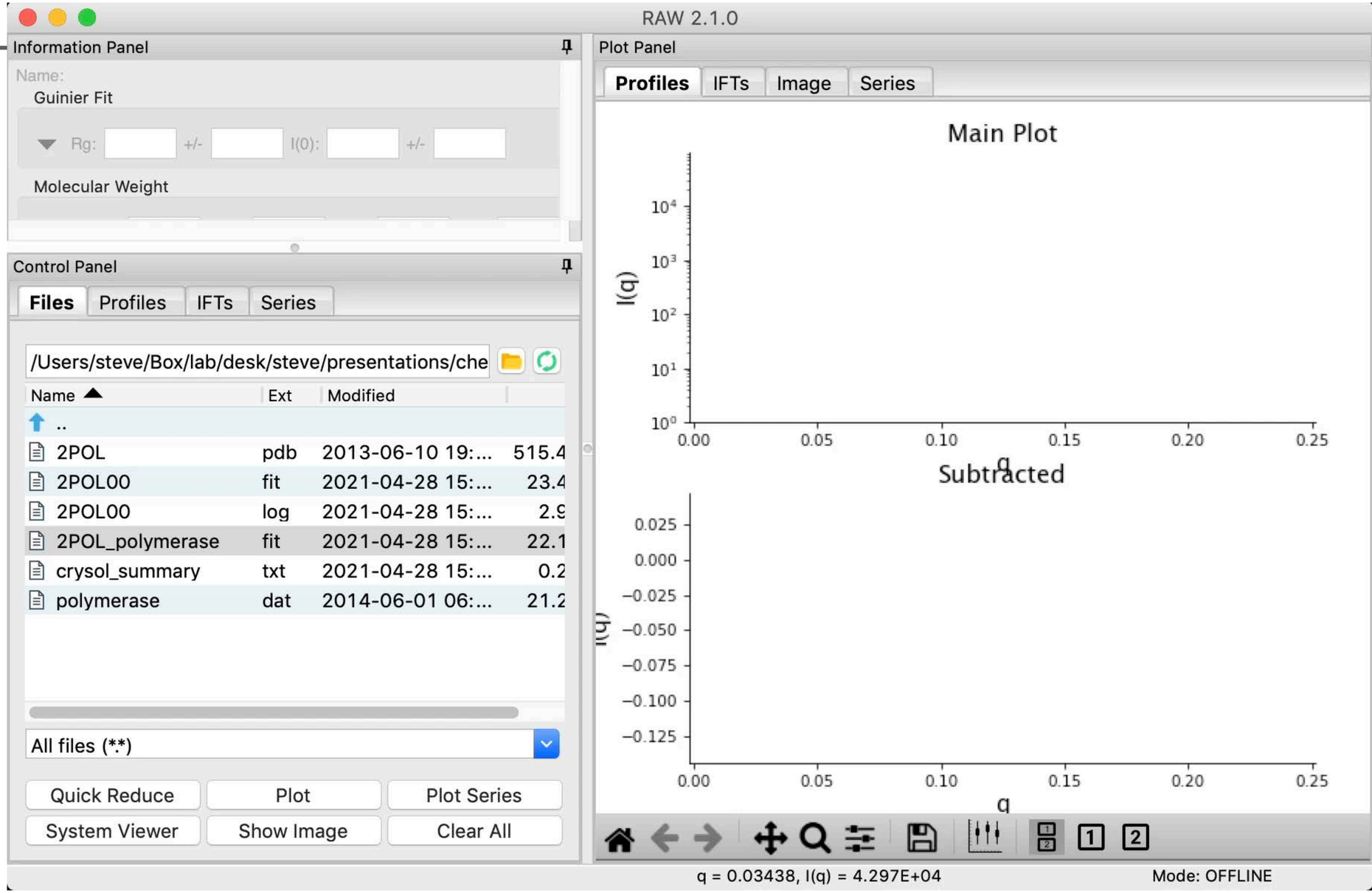
Svergun, Barberato, & Koch(1995)
J. Appl. Cryst. 28, 768-773.



# Plotting fits in RAW

**RAW reference:**

Hopkins, Gillilan, & Skou (2017).
J Appl Cryst 50, 1545-1553.



# SAXS simulation: tips & caveats

- CRYSOL: may need to adjust advanced parameters

  `-lm` number of spherical harmonics.
  affects accuracy at high-*q*

  `-fb` affects accuracy of solvation layer
  I recommend `-fb 18` (maximum)

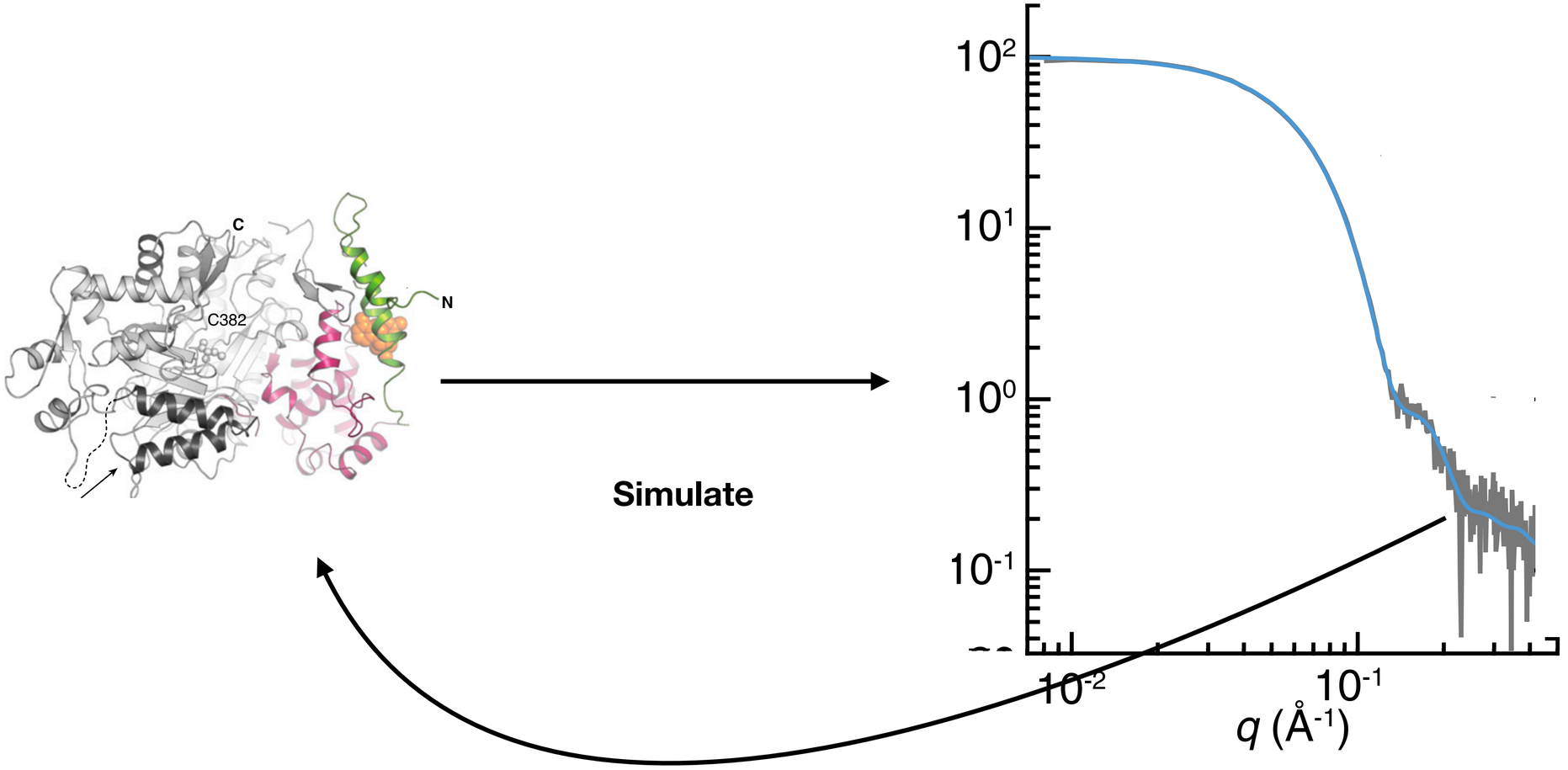
  `-dns` solvent density in electrons / $\mathring{A}^3$
  (default is 0.334 for pure water)

- Always check that fitted parameters make sense
- Both can run in “prediction” mode without input data (set parameters to default or user-input values) — useful!
- Atomic model must be complete
- Make sure atomic model was read in correctly (check number of atoms, chain ID, symmetry, etc)

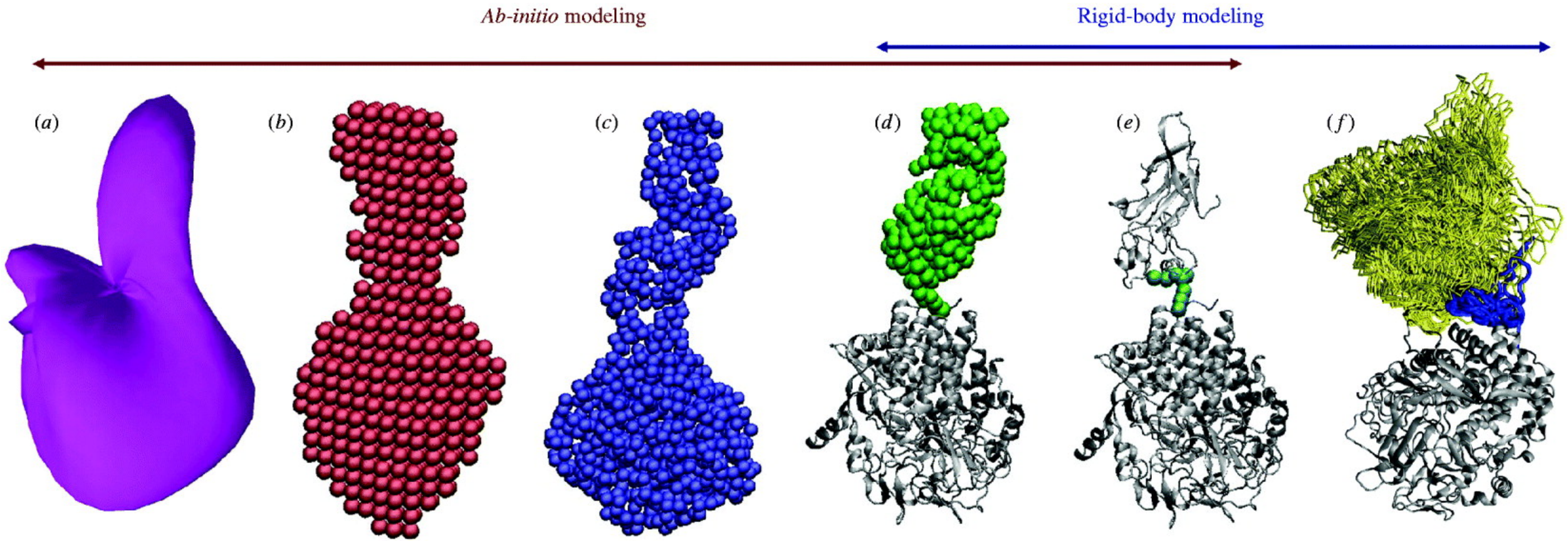
**Getting help**

- CRYSOL
  - `crysol –help`
  - www.embl-hamburg.de/biosaxs/crysol.html
- FoXS
  - modbase.compbio.ucsf.edu/foxs/about.html
  - modbase.compbio.ucsf.edu/foxs/help.html

# Modeling *from* SAXS data



# Multiple approaches to SAXS-based modeling




Envelope representation using spherical harmonics

Envelope from densely packed dummy beads

Envelope from dummy residues forming a chain-compatible model

Missing domain represented by ensemble of dummy residues forming a chain-compatible model

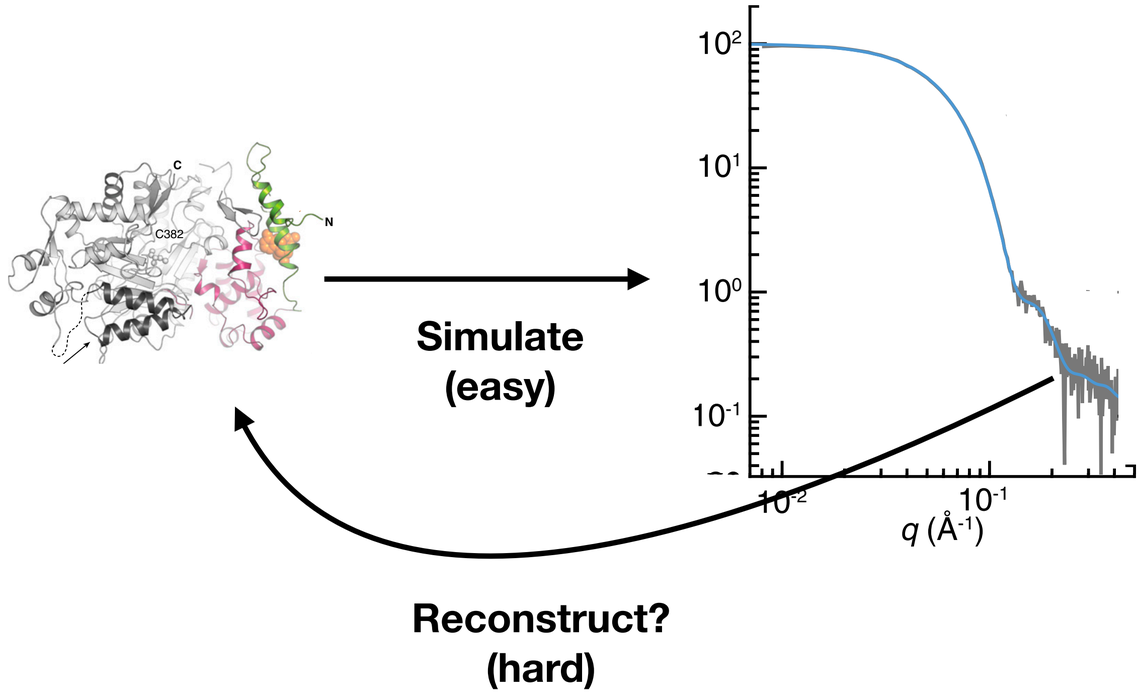
Rigid-body model + missing loop represented by ensemble of dummy residues

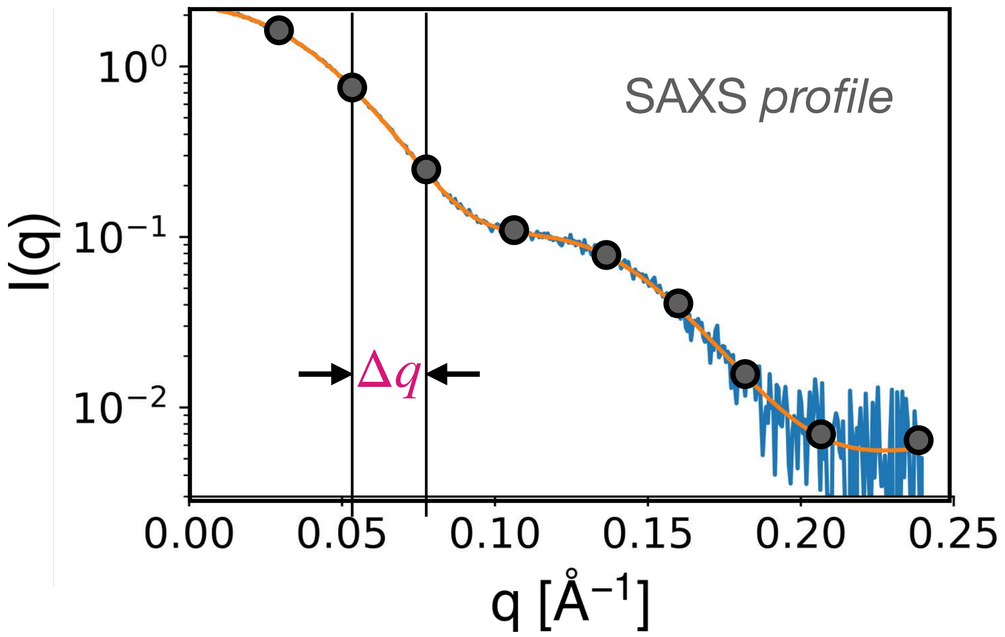
Atomic models derived from rigid body modeling applying conformational sampling

Putnam, Hammel, Hura, & Tainer. Q Rev. Biophys. 2007

# Shape reconstruction is a hard problem

- SAXS is <u>low resolution</u>
  ($q_{\mathrm{max}} = 0.2\ \mathrm{\AA}^{-1} \rightarrow$ resol. $= 30\ \mathrm{\AA}$)
- SAXS is <u>low information</u>
  (Nyquist-Shannon sampling theorem)



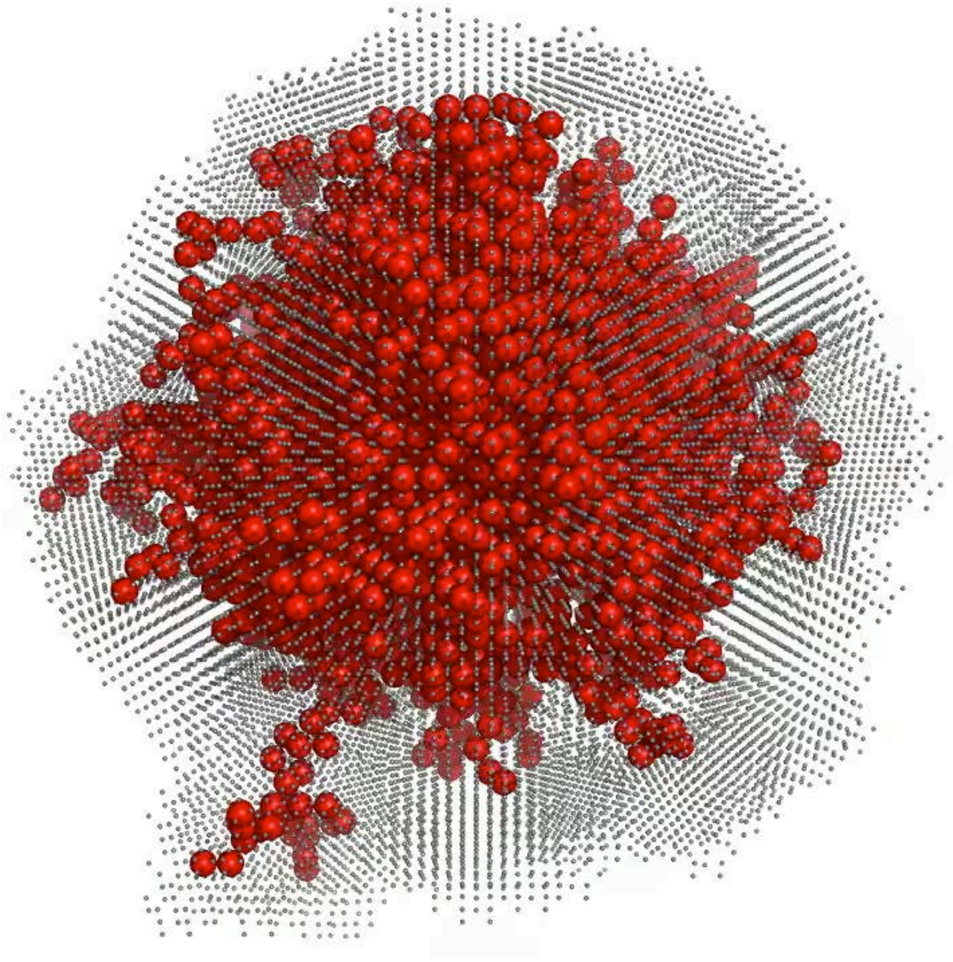

$\Delta q = \pi / d_{\mathrm{max}}$

Amount of information $= (1/\pi) q_{\mathrm{max}} d_{\mathrm{max}}$

Typically, ~ 10-20. Not much!

# Bead modeling approach

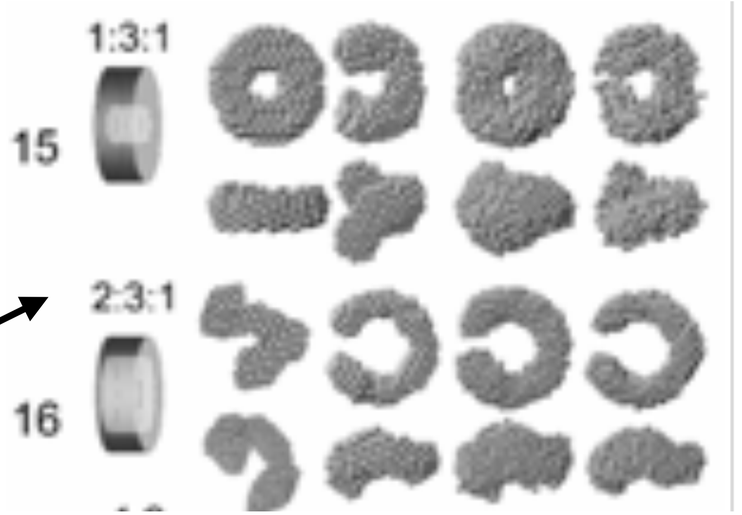
- Software (ATSAS)
  - DAMMIN — original
  - DAMMIF — accelerated
  - GASBOR — chain-like, valid at high *q*
- Optimization using simulated annealing
- A single reconstruction may have high resolution features, this can be misleading.
- Solution is not unique —> repeat many times and average them.

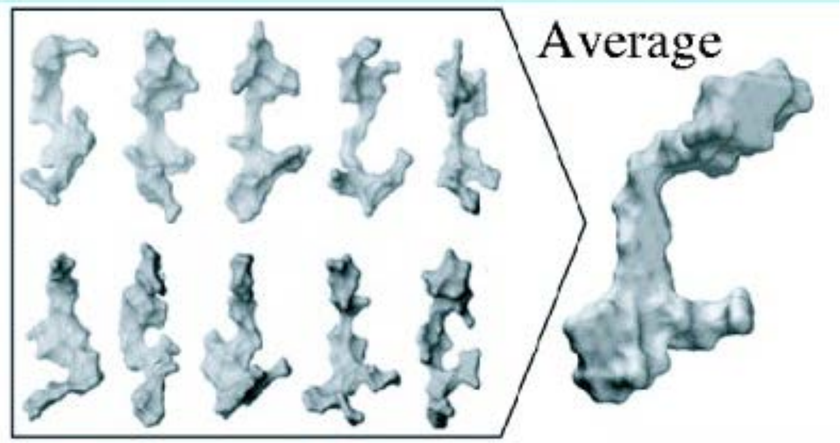


Skou, Gillilan, & Ando (2014) Nature Protocols 9: 1727–1739.

# Unique reconstruction is not guaranteed

- Theoretically, multiple shapes can give the same scattering profile —> *ambiguity*
- Some SAXS profiles are more ambiguous than others.
- AMBIMETER — fit SAXS data to database of simple shapes. How ambiguous?
- Bead models have trouble with certain shapes.
- Best practice: perform multiple, independent reconstructions, align & average (DAMAVER)
  - NSD = “normalized spatial discrepancy”. NSD < 0.7 is similar enough to average
  - Large NSDs —> multiple shapes possible. Cluster first then average (DAMCLUST)




Volkov & Svergun (2003). J. Appl. Cryst. 36(3), 860–864.
M.V. Petoukhov and D.I. Svergun (2015) Acta Cryst. D71, 1051-1058.
Putnam, Hammel, Hura, & Tainer. Q Rev. Biophys. 2007.

# DAMMIF demo

## RAW user interface:

Hopkins, Gillilan, & Skou (2017).
J Appl Cryst 50, 1545-1553.

## ATSAS programs:

DAMAVER — Volkov & Svergun (2003).
J. Appl. Cryst. 36(3), 860–864.

AMBIMETER — Petoukhov & Svergun (2015) Acta Cryst. D71, 1051-1058.

DAMMIF — Franke and Svergun (2009).
J Appl Cryst. 42, 342-346.

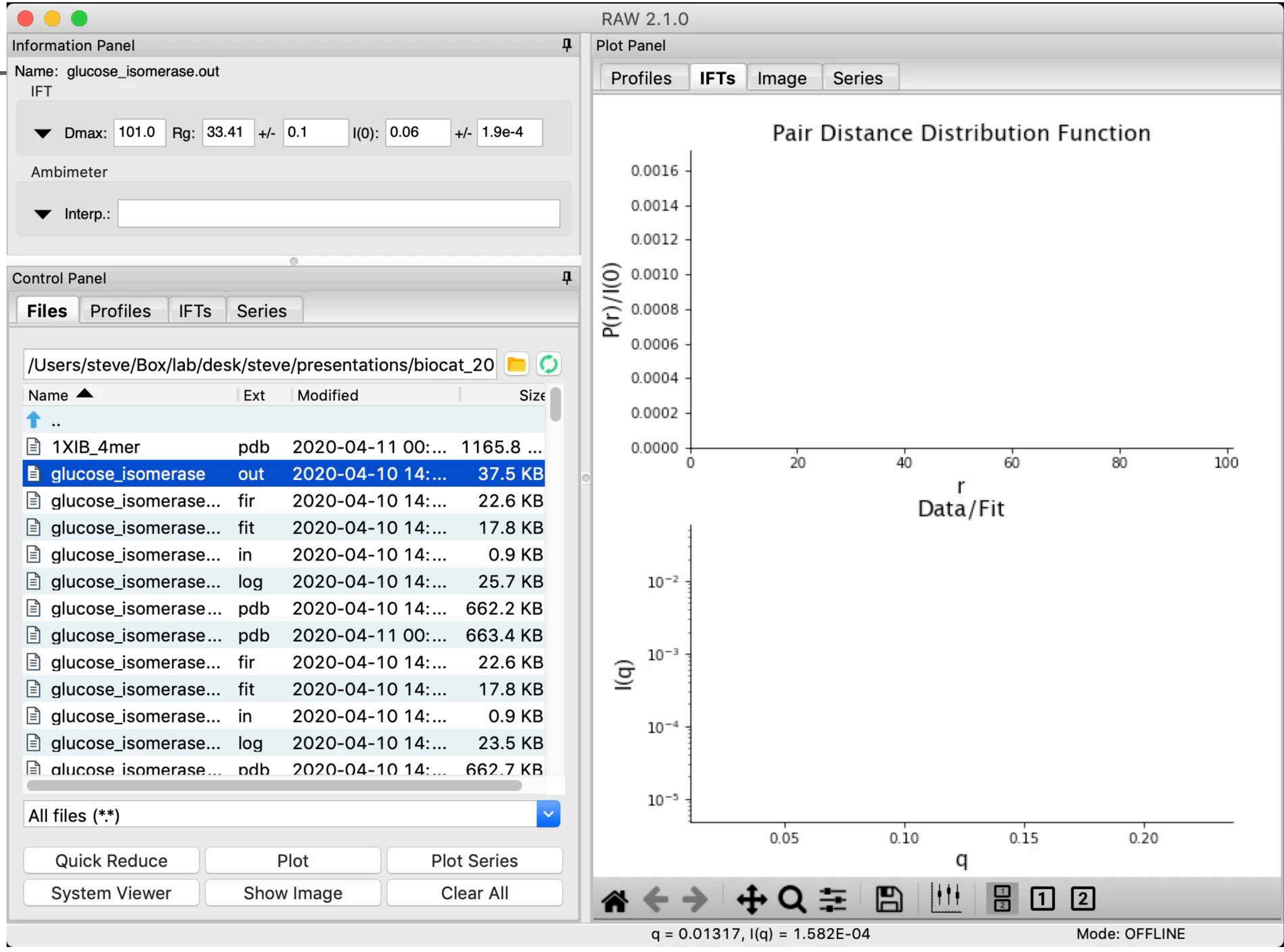
DAMMIN — D. I. Svergun (1999).
Biophys J. 2879-2886.

## Visualization in Chimera X:

https://www.rbvi.ucsf.edu/chimerax/

## Adapted from RAW tutorial:

https://bioxtas-raw.readthedocs.io/en/latest/tutorial/s2_dammif.html



# Advanced modeling summary and advice

- Must have good quality data.
- Monodisperse sample is required for modeling*

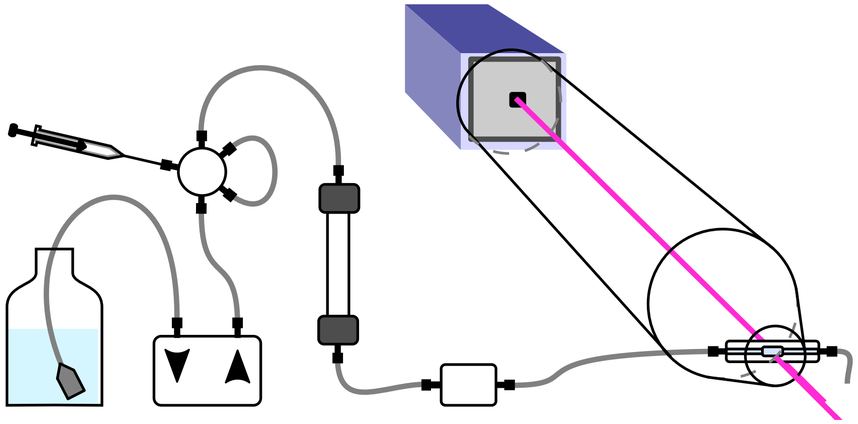
<— SEC-SAXS can help!

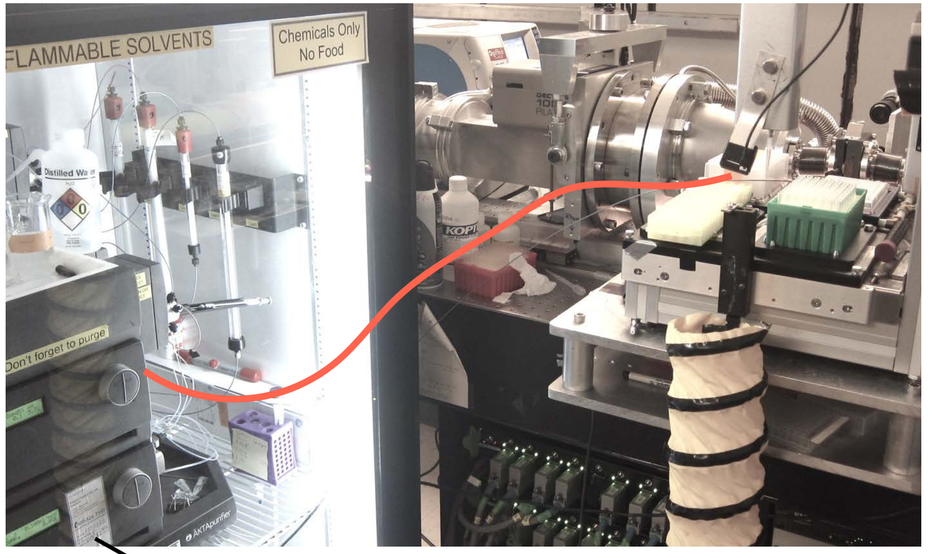
- Lots of software available for *ab-initio* reconstruction
- Reconstruction can be ambiguous. Quantify ambiguity, interpret models with caution.
- If you have an atomic model, fit SAXS simulation to scattering curve.

  Don’t just dock your model in the *ab-initio* density 😒

- Even better: use SAXS simulation to rank alternative models 😀
- *if your sample is a mixture, or is flexible, there is software to fit multiple models. But remember SAXS is low-information. Proceed with caution.

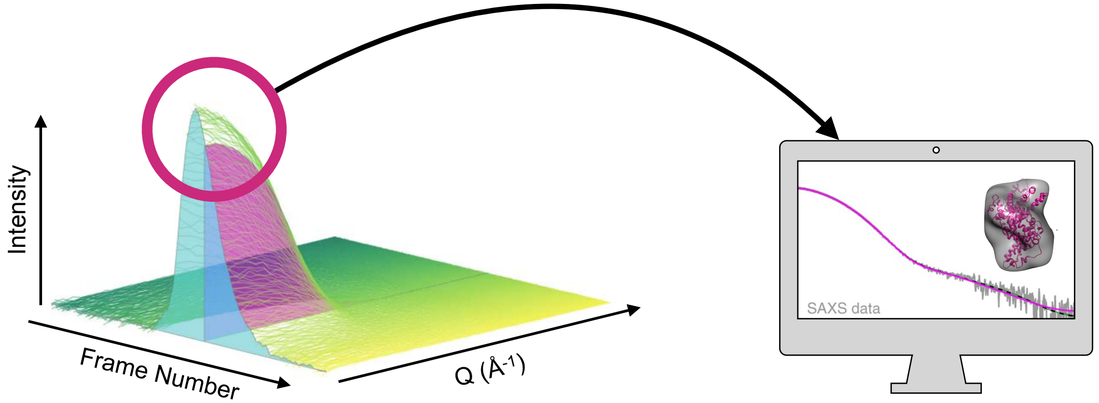
# SEC-SAXS experiment






X-rays

FPLC (AKTA Purifier)




**Advanced data processing**

# Why use SEC-SAXS?

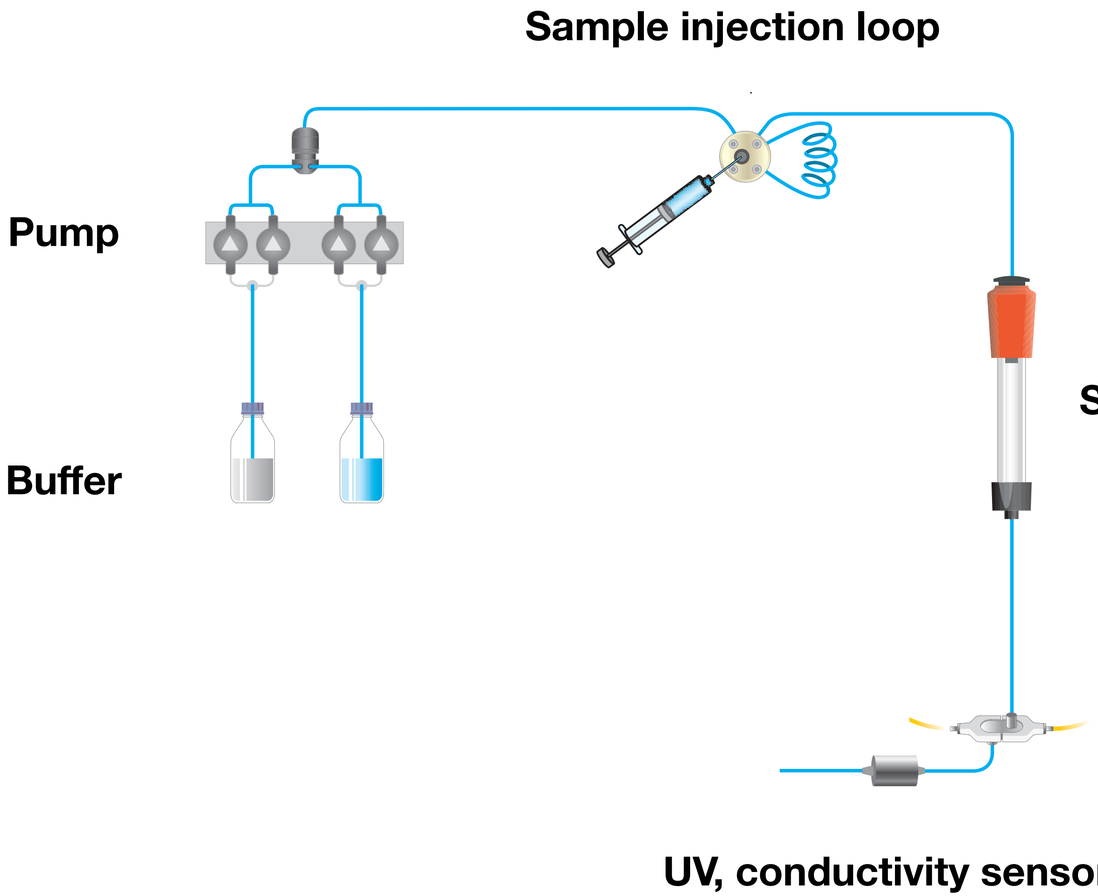
## Pros

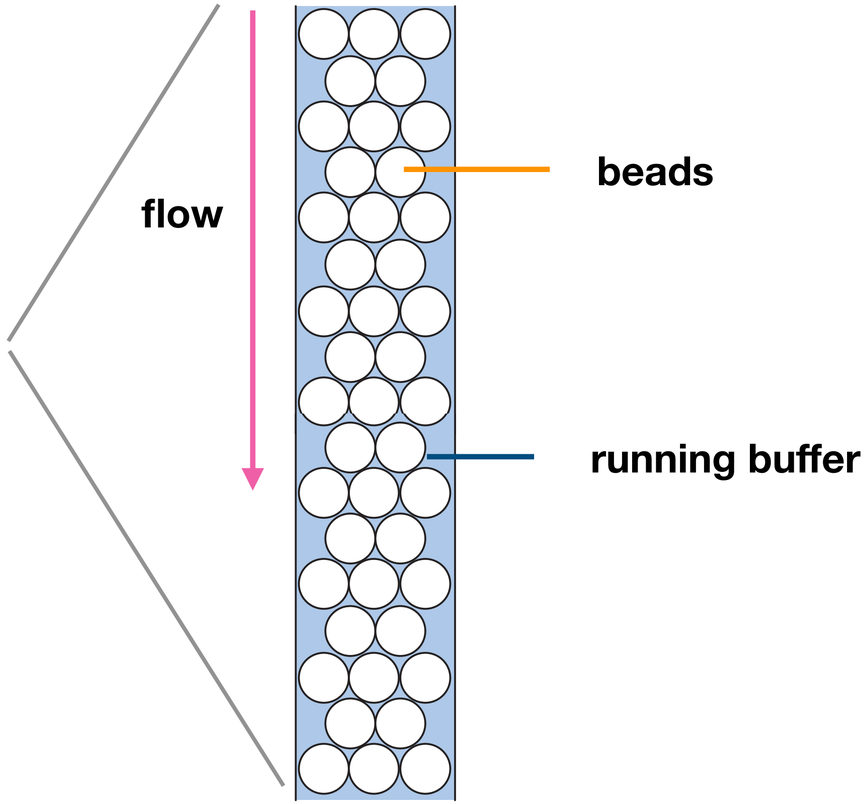
- Exact buffer match
- Remove aggregates
- Confirm monodispersity
- Separate mixtures
- Computationally deconvolve overlapping peaks

## Cons

- Usually uses more material, takes longer, dilutes the sample
- Buffer is fixed (no titrations)
- Protein concentration varies (issue for weak complexes)
- Radiation damage can compromise experiment

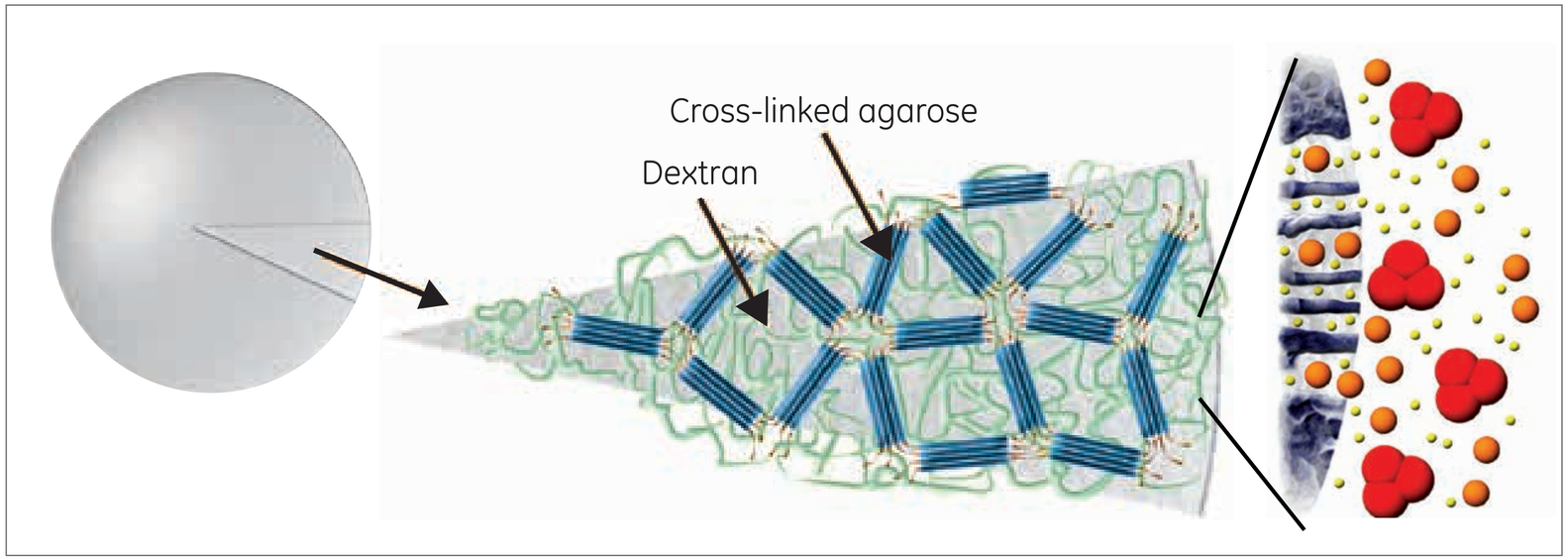
# Size exclusion chromatography (SEC) setup

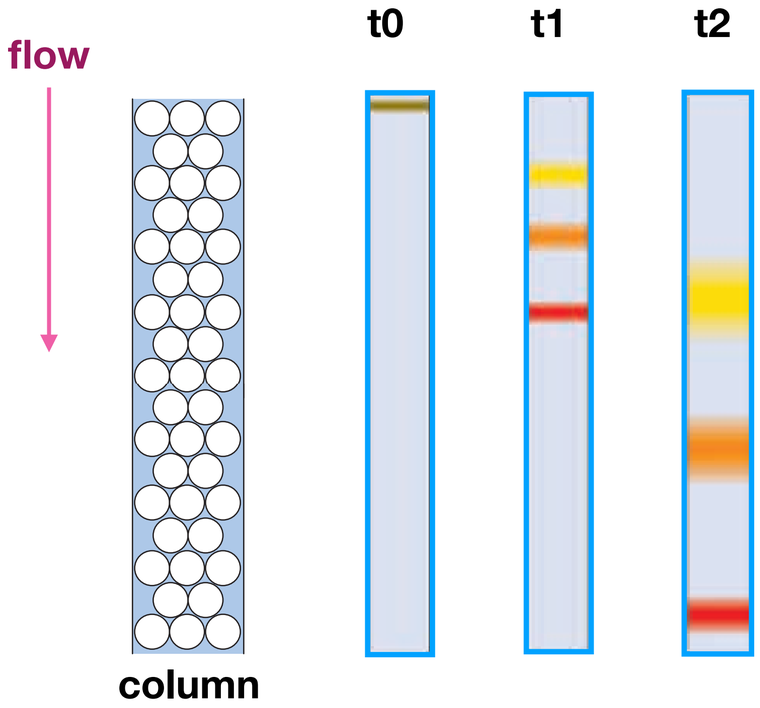



"Size exclusion chromatography: Principles and Methods", GE Healthcare Life Sciences.

# SEC separates particles by size

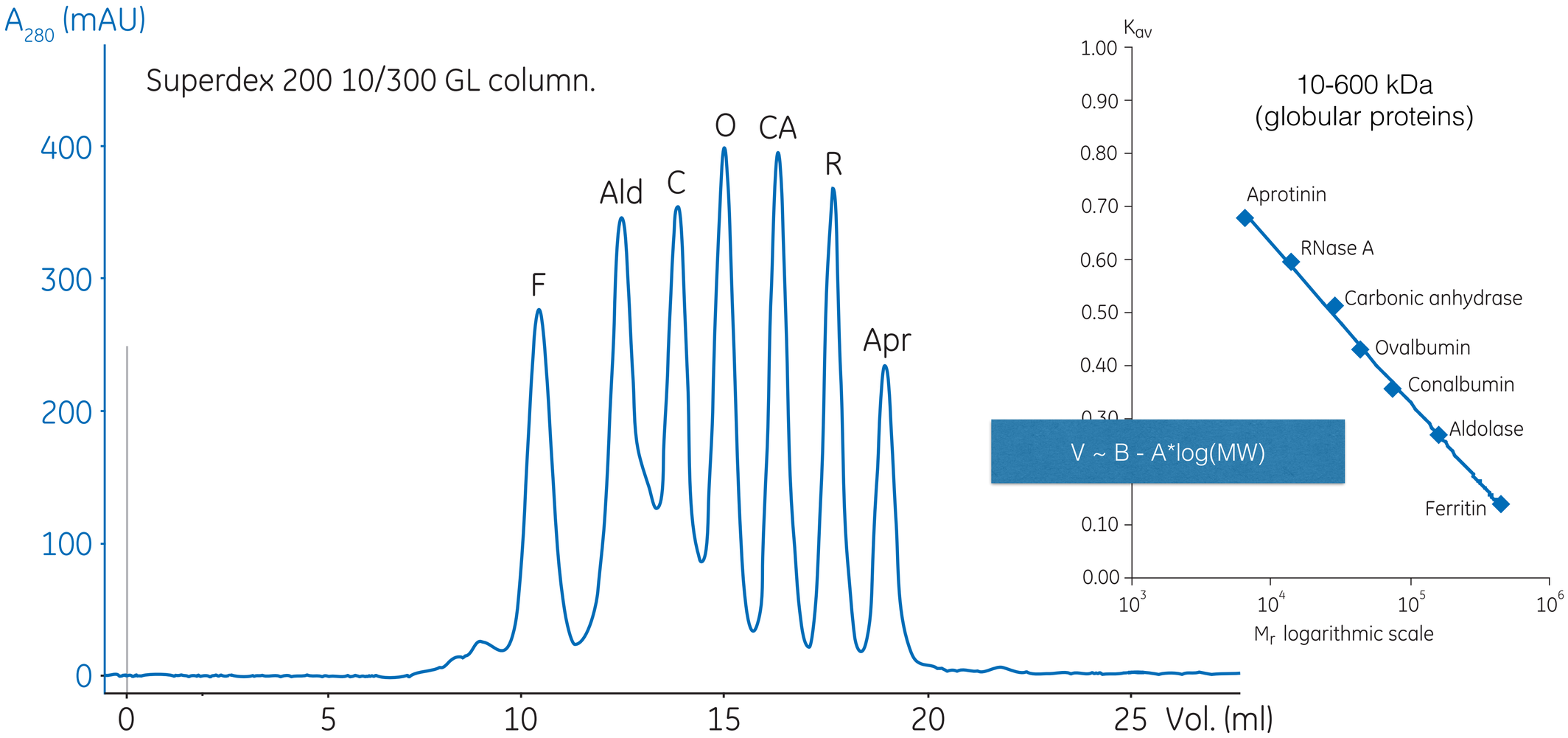



- Large objects are excluded → run quickly / elute first
- Small objects diffuse in the pores → run slowly / elute last

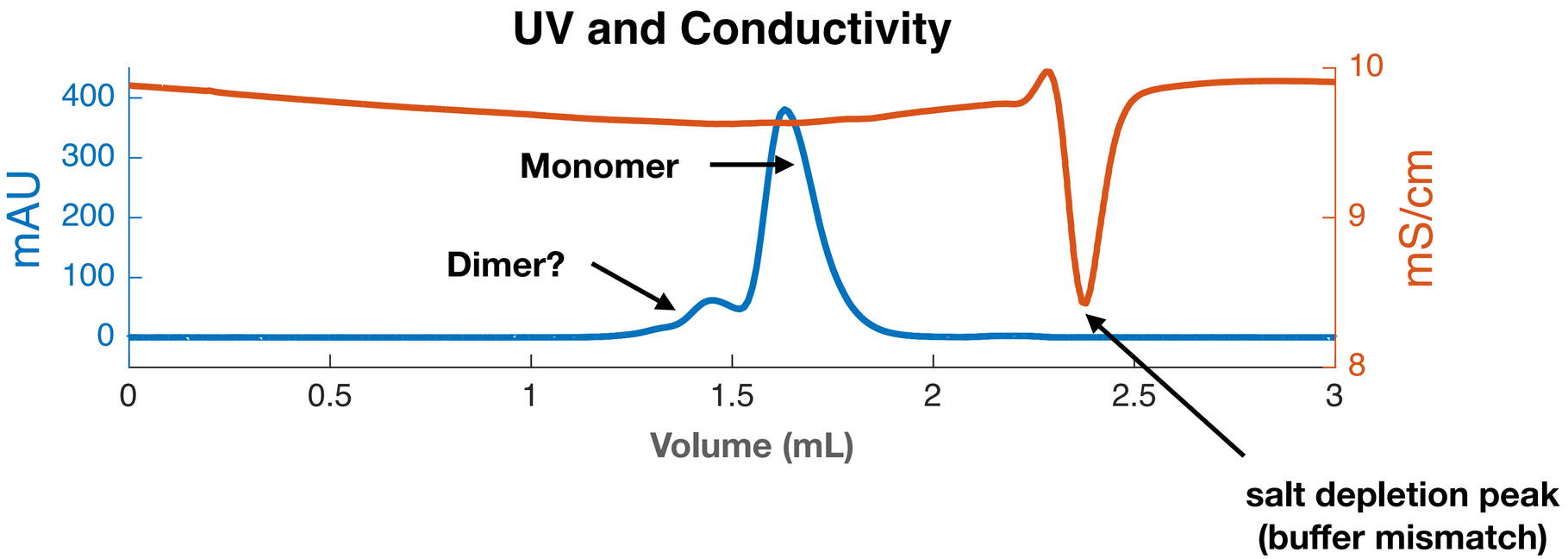
"Size exclusion chromatography: Principles and Methods", GE Healthcare Life Sciences.

# Globular proteins elute according to logarithm of molecular weight




"Size exclusion chromatography: Principles and Methods", GE Healthcare Life Sciences.

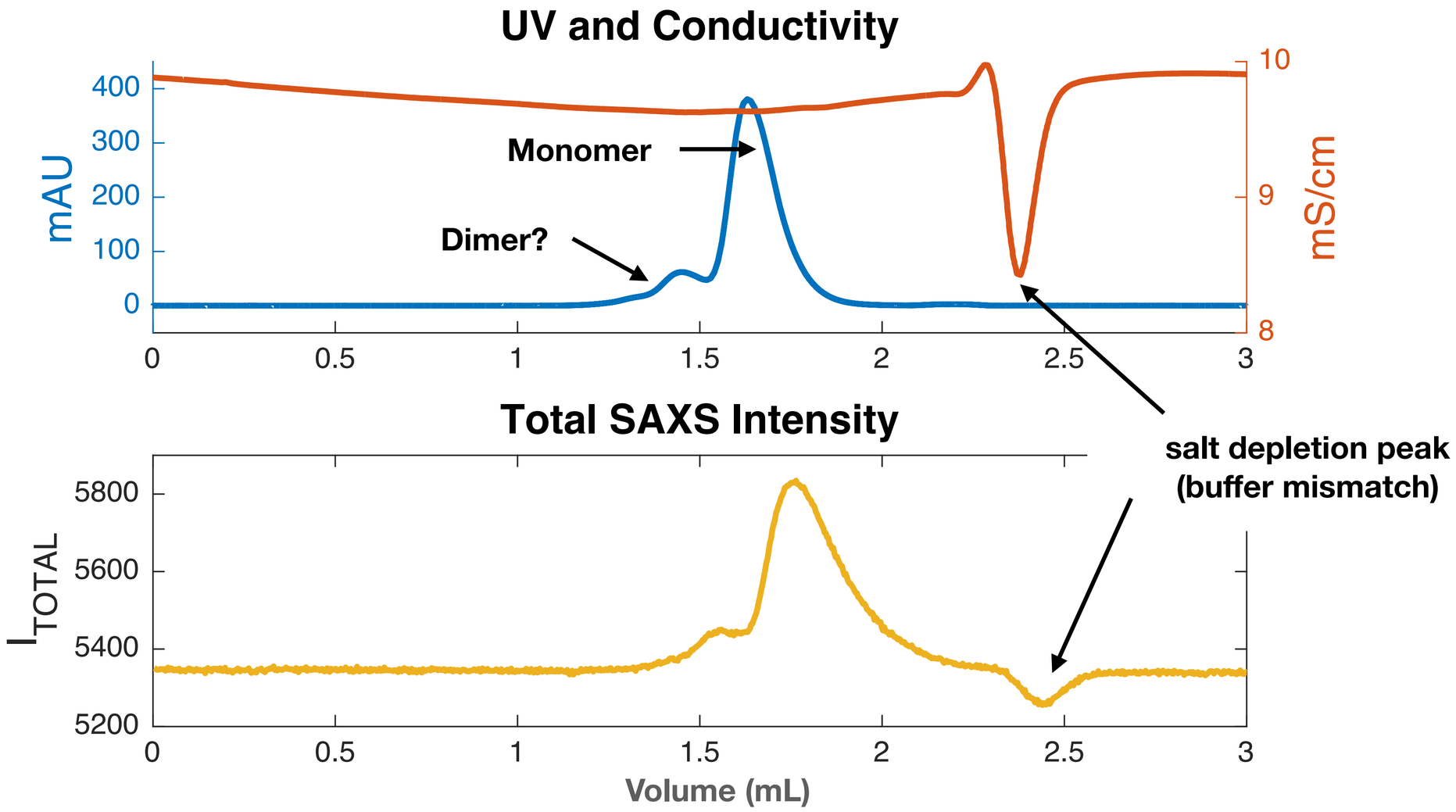
# Example SEC-SAXS dataset from BSA




**Experimental details**

| | |
|---|---|
| **Beamline** | CHESS G1, Nov. 2015 |
| **Sample** | BSA at 11 mg/mL in 50 mM HEPES, pH 7.5 |
| **Column** | Superdex 200 Increase, 3.2/300 |
| **Running Buffer** | 50 mM HEPES pH 7.0, 100 mM NaCl, 5% glycerol |
| **Injection Volume** | 50 uL |
| **Flow Rate** | 0.1 mL/minute |
| **Frames** | 1000 at 2s each |

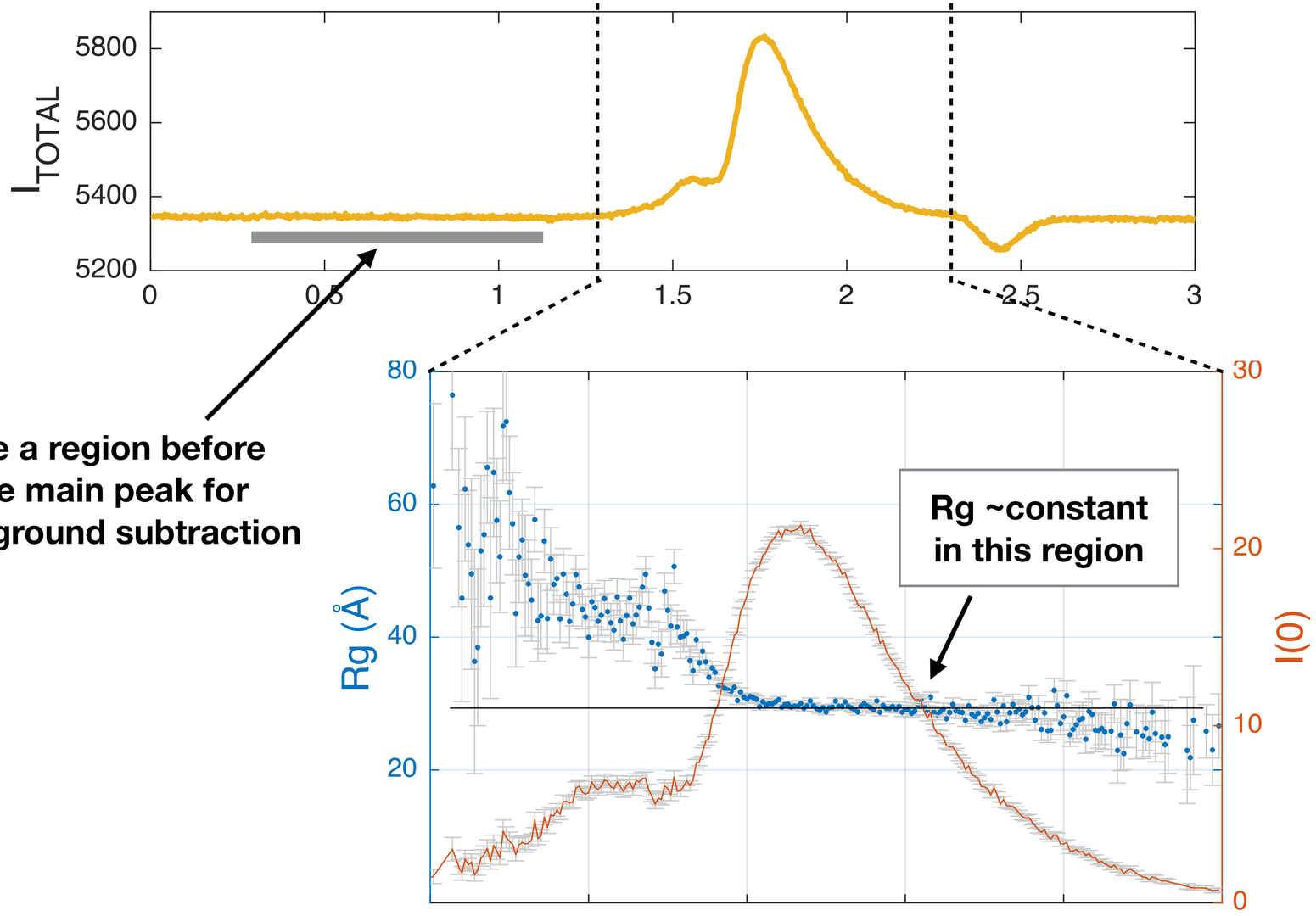
# Example SEC-SAXS dataset from BSA




**Experimental details**

| | |
|---|---|
| **Beamline** | CHESS G1, Nov. 2015 |
| **Sample** | BSA at 11 mg/mL in 50 mM HEPES, pH 7.5 |
| **Column** | Superdex 200 Increase, 3.2/300 |
| **Running Buffer** | 50 mM HEPES pH 7.0, 100 mM NaCl, 5% glycerol |
| **Injection Volume** | 50 uL |
| **Flow Rate** | 0.1 mL/minute |
| **Frames** | 1000 at 2s each |

# Guinier analysis across the peak



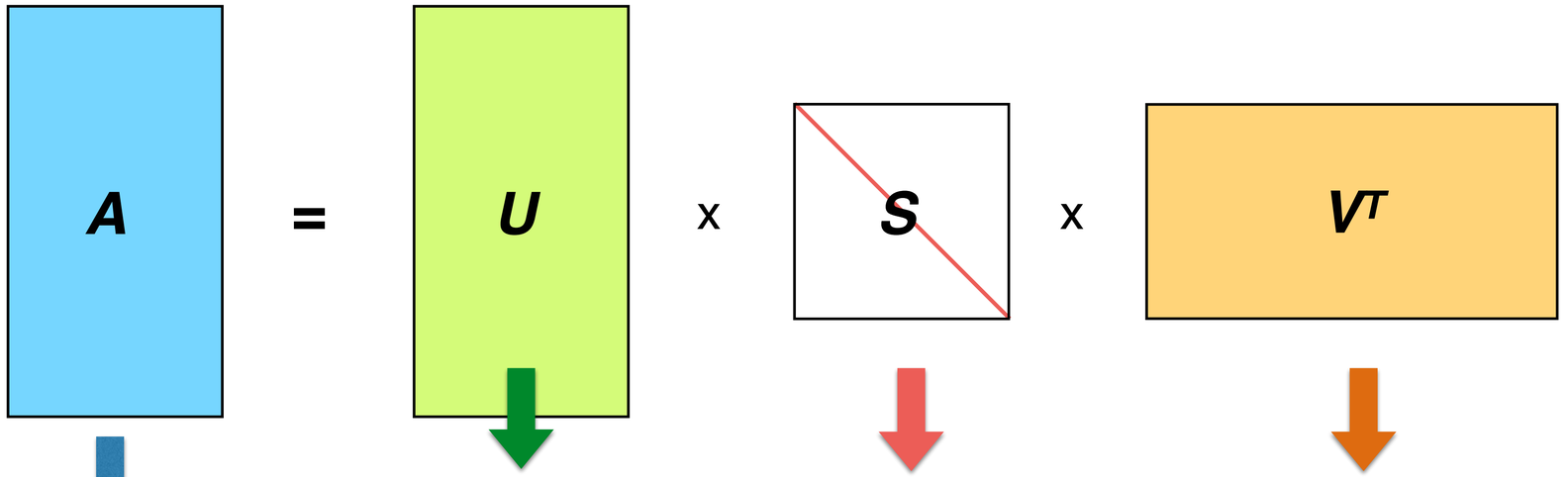

**Use a region before the main peak for background subtraction**

**Is the peak one species?**

**What curves can be averaged together?**

# SVD can help detect “pure components”

- SVD = “singular value decomposition”
- Method from linear algebra (assume SAXS is additive)
- Detect number of significant singular values above noise
- = minimum number of distinct scattering components in mixture

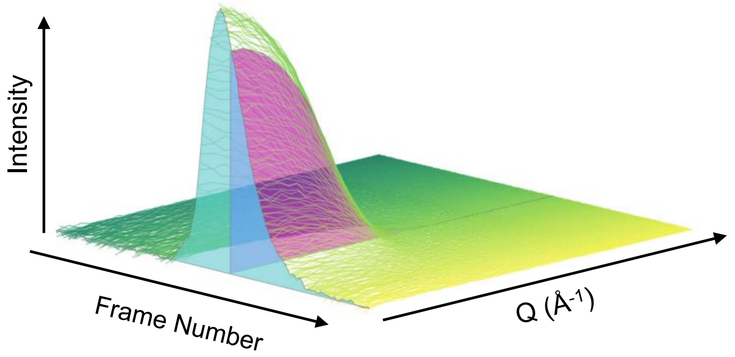


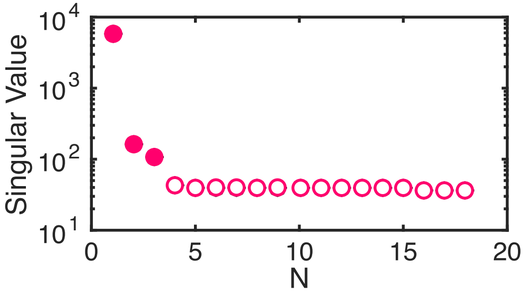

SEC-SAXS dataset

singular vectors along *I* vs. *Q* (SAXS profiles)

singular values

singular vectors along *I* vs. frame (elution profiles)



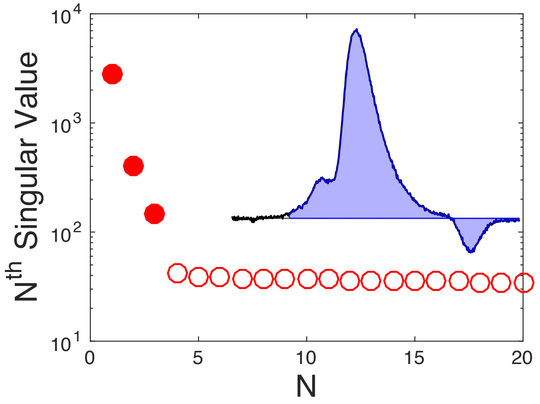



# Application of SVD to SEC-SAXS dataset from BSA

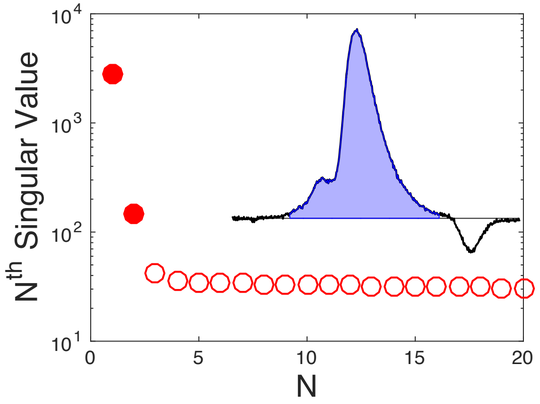
**Is the peak one species?**

**What curves can be averaged together?**

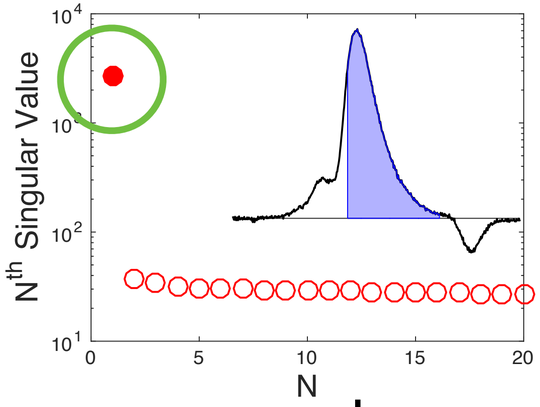
**whole peak: 3 s.v.**



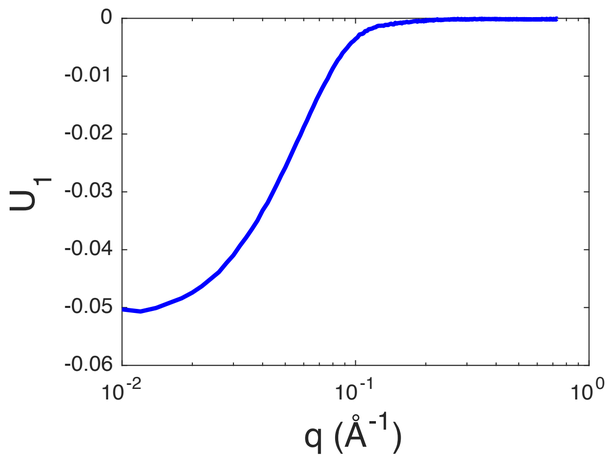

**first part: 2 s.v.**



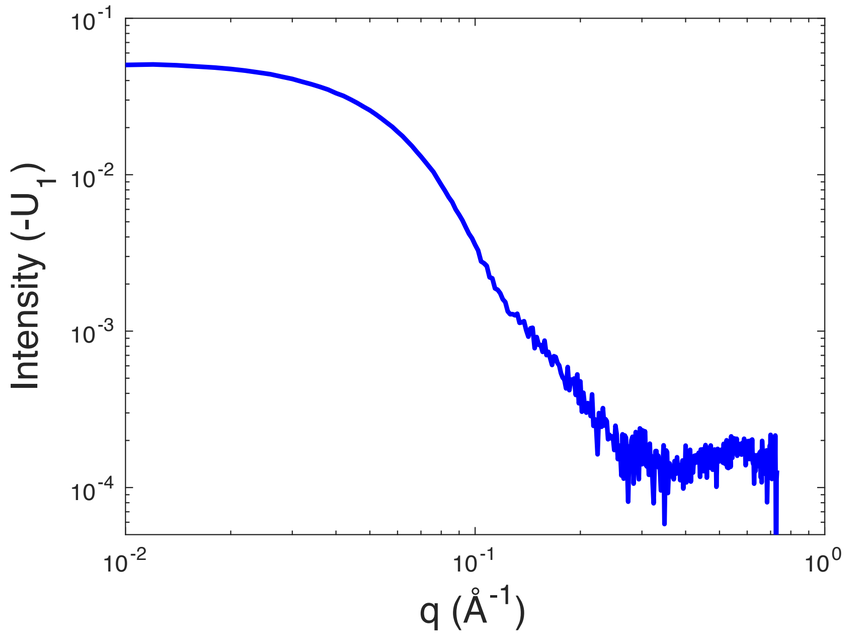

**middle part: 1 s.v.**




**U1 = first left singular vector**

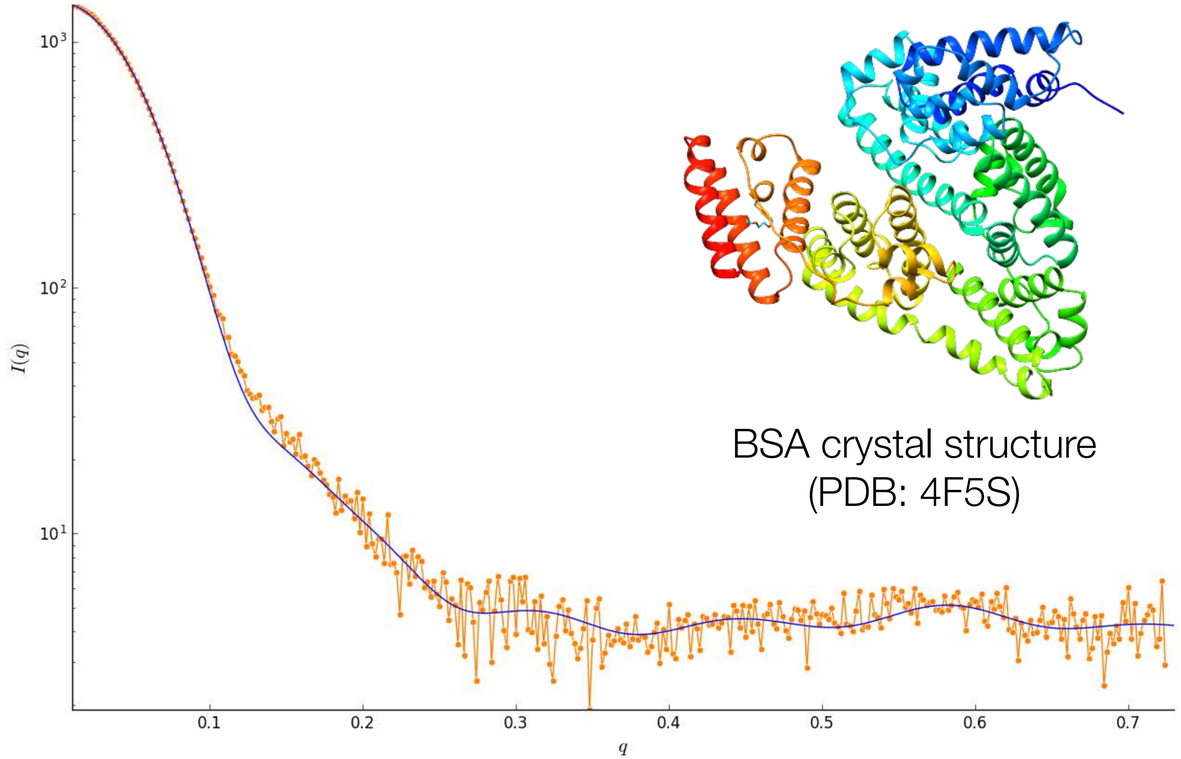



**flip the sign**



# Proceed to advanced analysis with confidence!

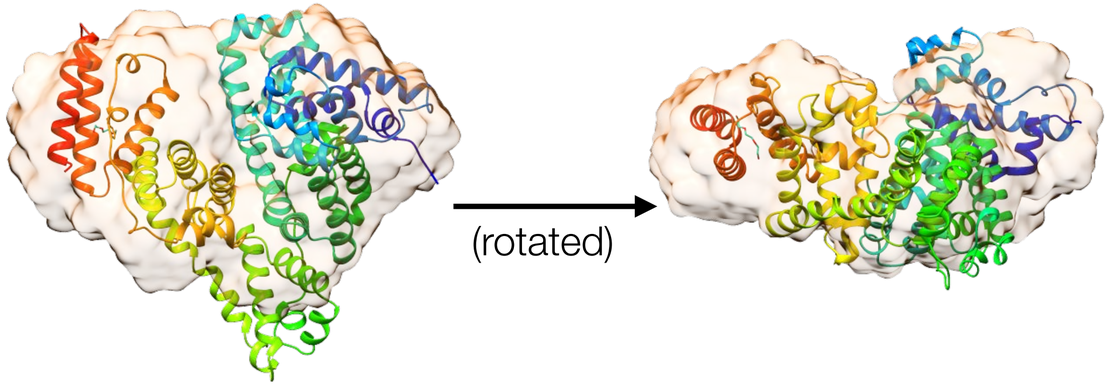
**CRYSOL fit**






BSA crystal structure
(PDB: 4F5S)

**Shape reconstruction**




8x DAMMIF fast mode, mean NSD 0.685

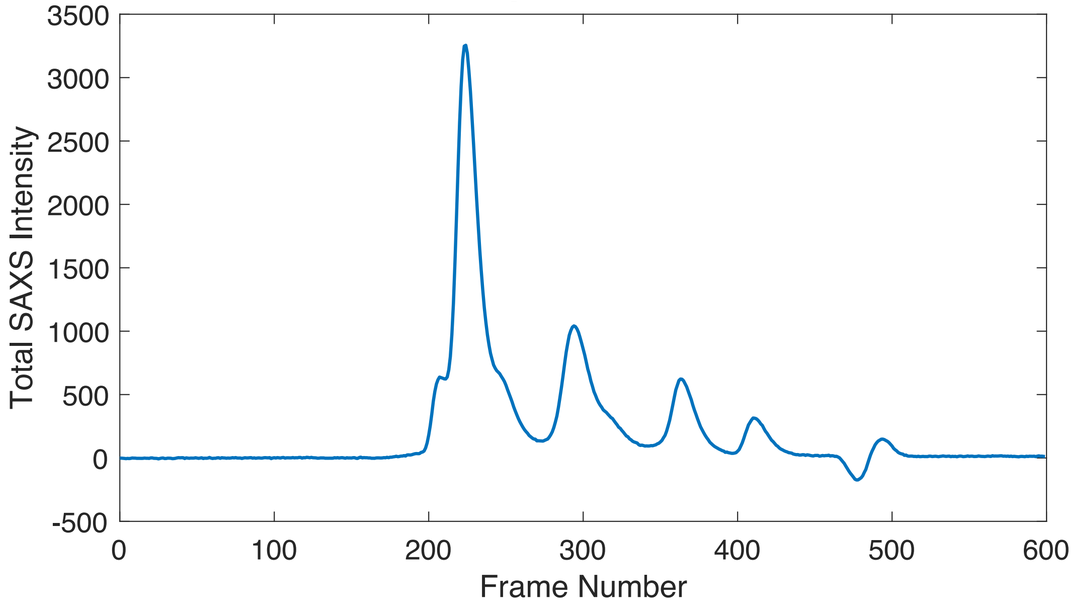
# What if peaks overlap?

- Often, Rg will change across the peak with no constant region
- Possible causes
  - Insufficient column resolution to separate species
  - Peak broadening
  - Re-equilibration of oligomers or aggregation
- Optimize purification as much as possible (buffer components, etc.)
- Advanced computational methods
  - EFA = “Evolving Factor Analysis”
  - REGALS = “Regularized Alternating Least Squares”

Meisburger, S.P. et al. (2016) JACS 138, 6506–6516.
Meisburger, S.P., Xu, D. & Ando, N. (2021) IUCrJ 8(2).

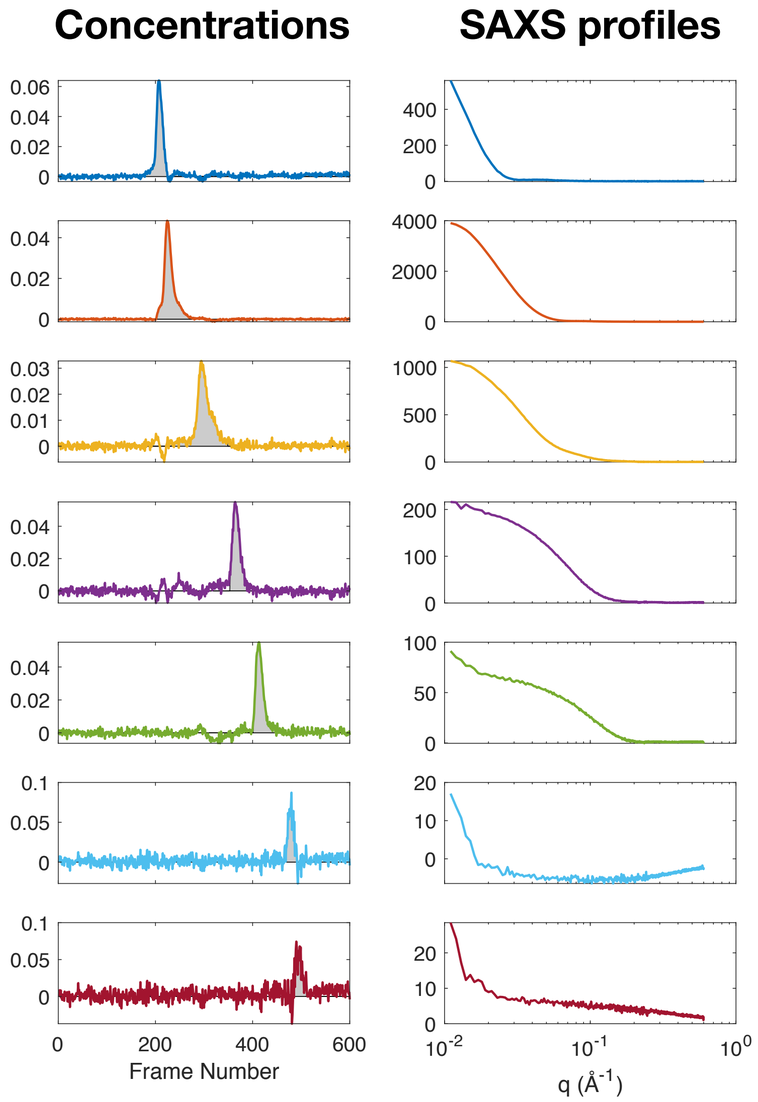
# EFA automatically detects and separates overlapping peaks in SEC-SAXS

**Example dataset**




Bio-rad chromatography standard
(thyroglobulin, γ-globulin, ovalbumin,
myoglobin, vitamin B12)

**EFA**




**Latest method, REGALS, also works on AEX-SAXS, titrations, time-resolved SAXS …**


Meisburger, S.P. et al. (2016) JACS 138, 6506–6516.
Meisburger, S.P., Xu, D. & Ando, N. (2021) IUCrJ 8(2).

# SEC-SAXS Summary

- SEC can separate molecules based on size
- SEC-SAXS is great at removing aggregates, separating oligomers, providing a good buffer match, and increasing overall confidence in data.
- SEC-SAXS typically requires extra sample, time
- Take care: experimental variables, assess data quality
- Powerful deconvolution methods (SVD, EFA, REGALS, …)