Jacobs School of Medicine and Biomedical Sciences
Department of Structural Biology
University at Buffalo

Advanced analysis: 3D reconstructions and fitting

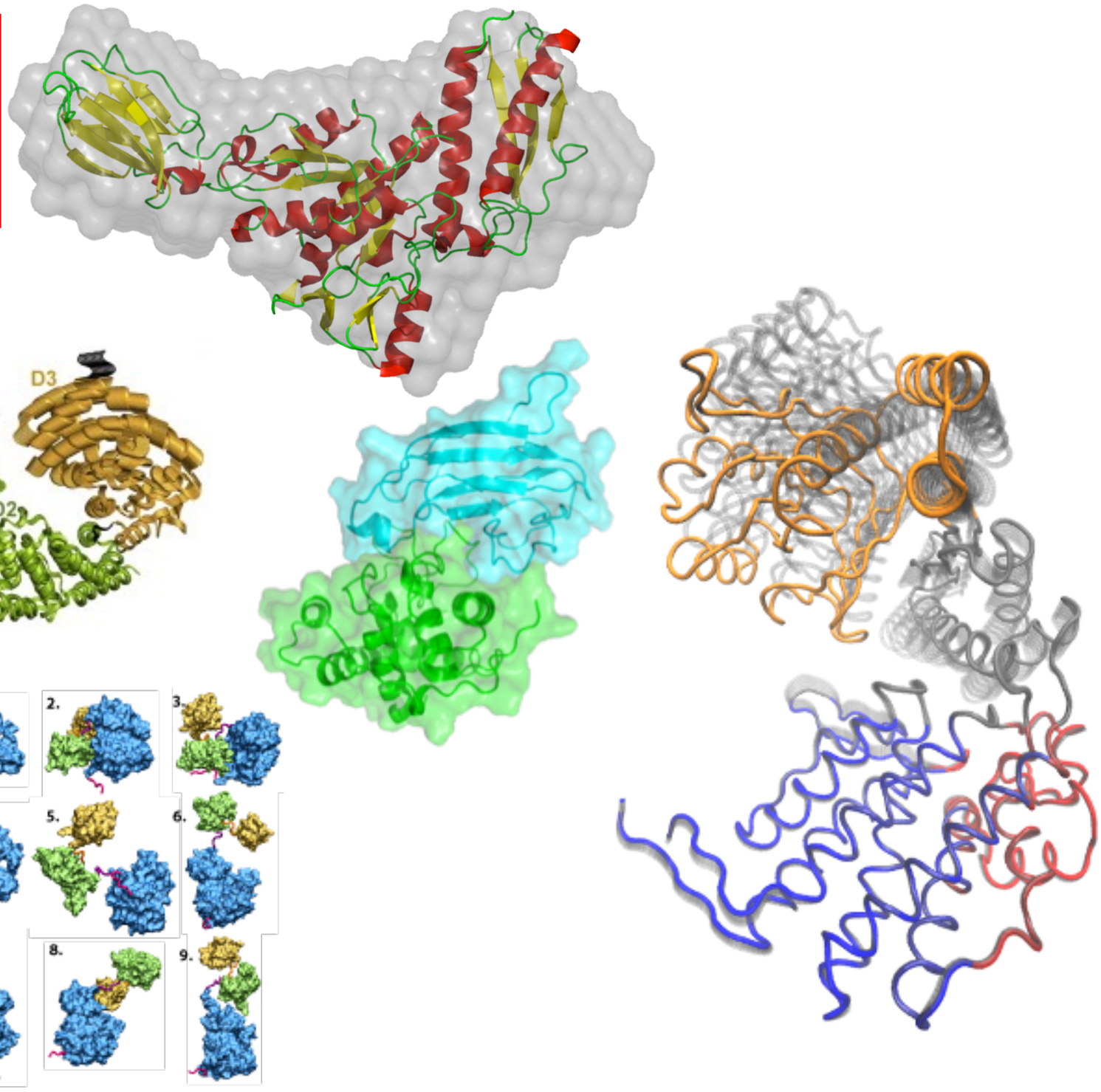
Everything BioSAXS 8 Workshop

June 2022

Tom Grant, Ph.D.
Assistant Professor
Department of Structural Biology
University at Buffalo
tdgrant@buffalo.edu

Advanced Methods of Modeling SAXS Data

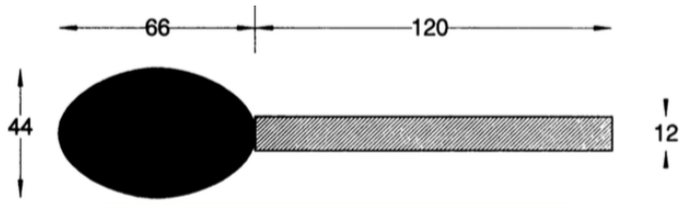
- Structure Fitting
- *Ab initio* 3D Reconstruction
- Rigid Body Modeling
- Docking
- Flexible Fitting
- Ensemble Modeling
- Mixtures
- Hybrid Modeling
- Contrast Matching
- Time-resolved SAXS



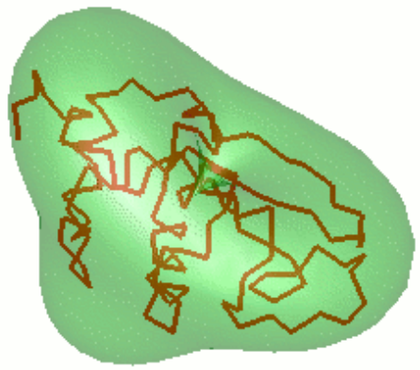
Modeling a Direct Electron Density Determination

Simple shapes

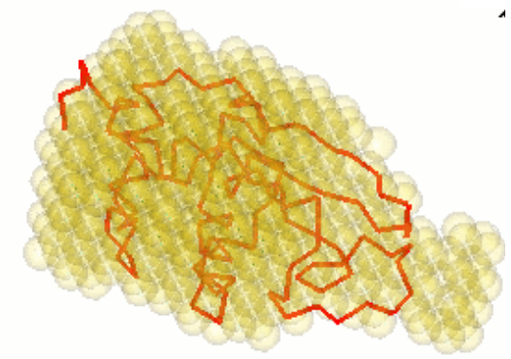
Collections of simple shapes



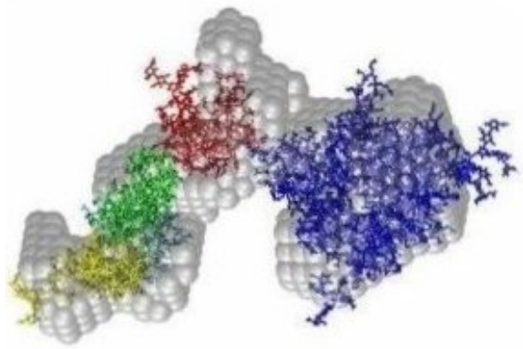
Spherical harmonics envelopes

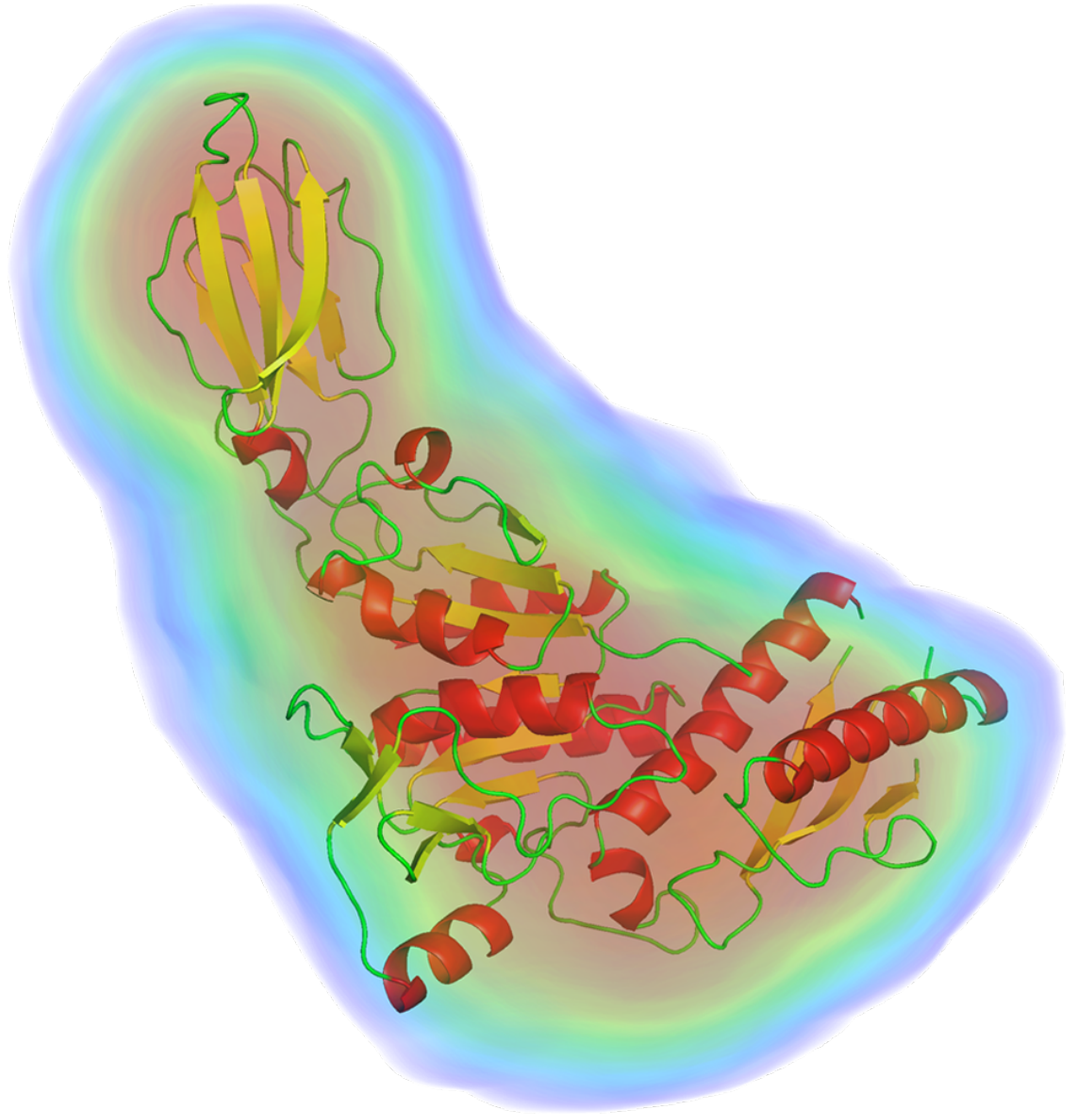


Bead modeling



Hybrid modeling





Calculating SAXS Profiles from Models

- All methods to model SAXS data require accurate calculation of reciprocal space scattering profile from model
- Calculation must not only be accurate, but also be computationally fast to evaluate thousands or even millions of possible candidate models
- Many, many algorithms exist for this purpose
- Methods to calculate scattering primarily based on Debye equation or spherical harmonics approximation
    - Debye equation:

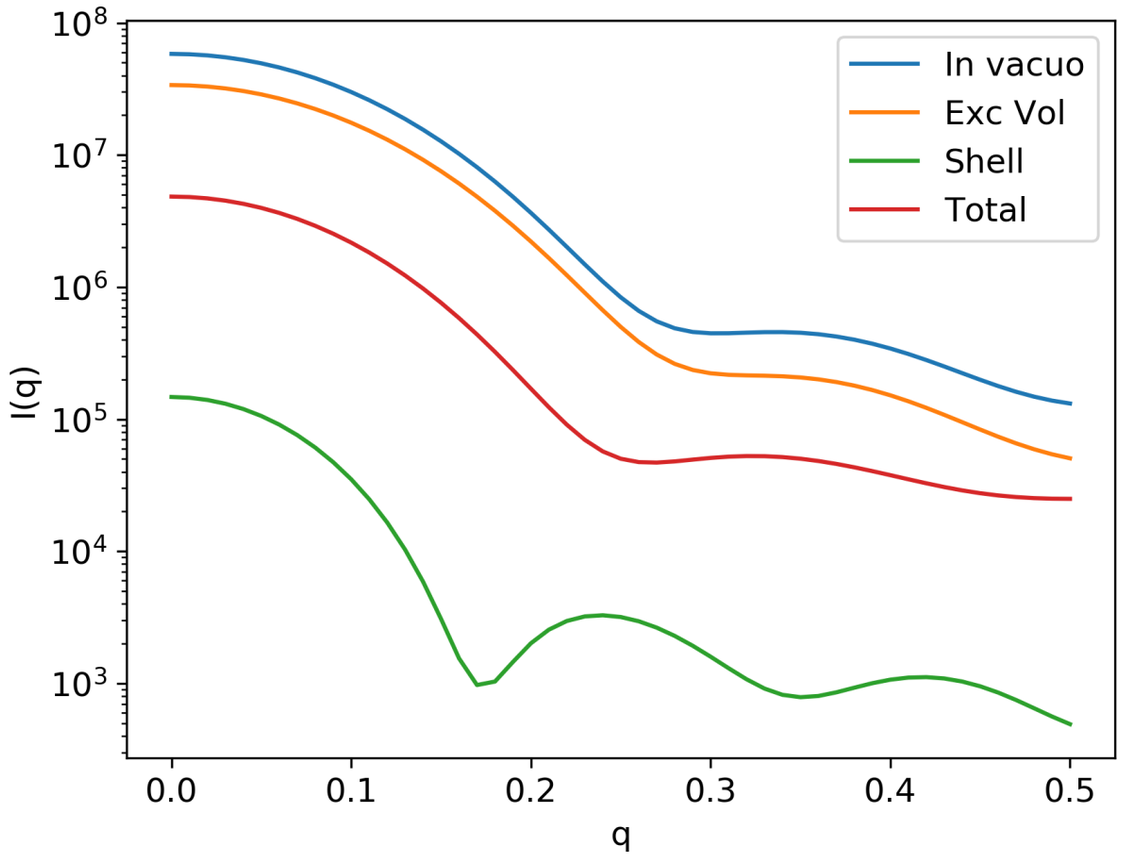
$$I(q) = \sum_i \sum_j f_i(q) f_j(q) \frac{\sin\left(q \cdot r_{ij}\right)}{q \cdot r_{ij}}$$

Calculating SAXS Profiles from Models

- Modeling solvent shell and excluded volume is major hurdle for fitting in reciprocal space
  - solvent contribution varies from molecule to molecule
  - chemical conditions affect scattering terms such as contrast
- Coarse-grained methods enable speed-up with slight accuracy cost
- Most common algorithms used are CRYSOL and FoXS, but many others exist, all giving trade-offs for accuracy, speed, and complexity

Solvent terms in scattering calculations

- Two primary solvent considerations in programs like CRYSOL, FoXS, and many 3D modeling algorithms
- Excluded volume: scattering contribution from the displaced bulk solvent
- Hydration shell: scattering contribution from the solvent that associates more tightly with the particle near the surface

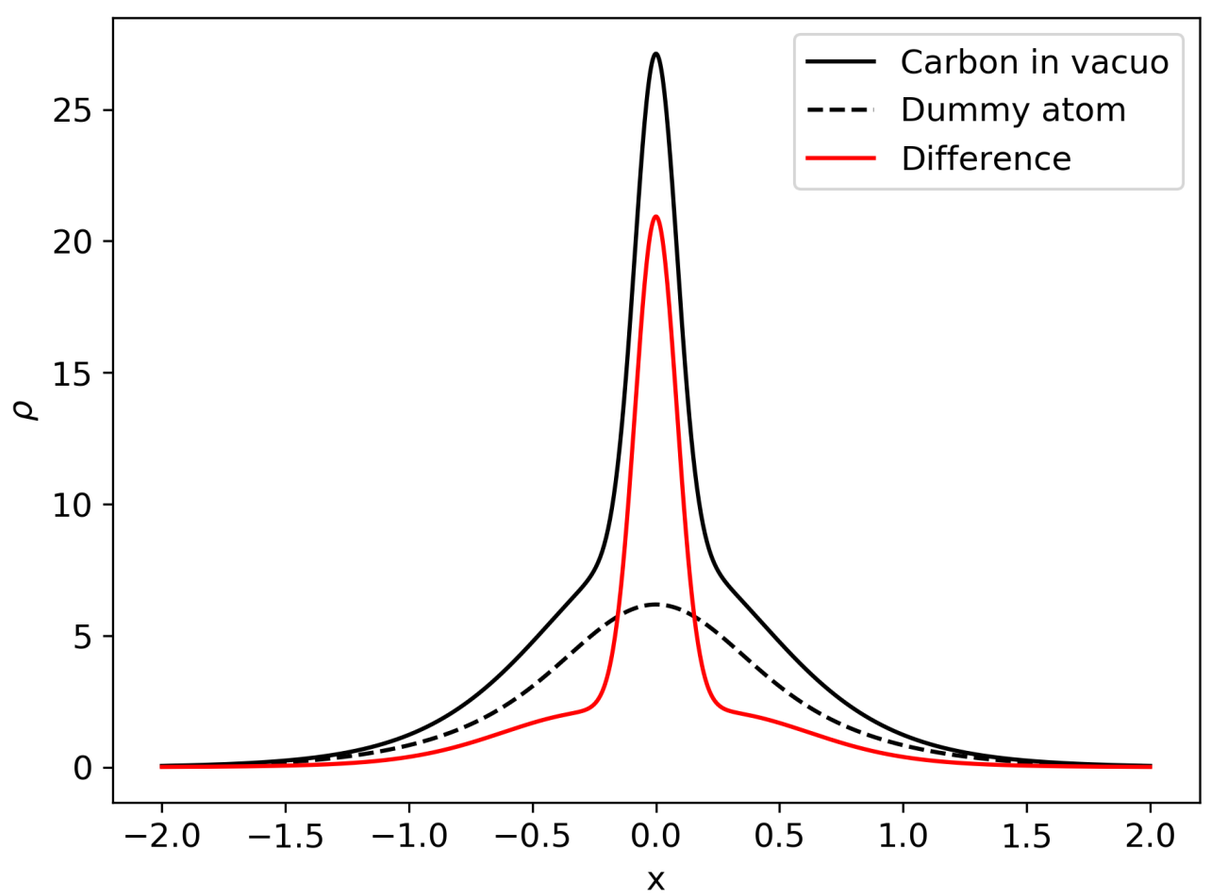


$$I(q) = \left\langle \left| F_{invacuo}(\mathbf{q}) - c_1 F_{exc.vol.}(\mathbf{q}) + c_2 F_{shell}(\mathbf{q}) \right|^2 \right\rangle_{\Omega}$$

Typically, c_1 and c_2 (or variations thereof) are solved for as free fitting parameters during modeling

Excluded volume solvent term

- SAXS measures particle *contrast*, i.e. the excess electron density relative to the bulk solvent
- Excluded volume solvent term accounts for the scattering from the volume of bulk solvent that has been displaced by the particle
- Most often modeled as a dummy atom placed at each atom location that is shaped as a Gaussian sphere whose size is determined by a free fitting parameter



Hydration shell

- Solvent near particle surface tends to associate and reorganize, resulting in ~10% - 15% greater average density in the layer surrounding the particle
- Scattering from this shell often modeled implicitly as a thin (~3Å) layer of excess density (though this approach varies between algorithms)
- The contrast of this shell is typically fitted as a free parameter
- Some algorithms allow for explicit modeling of hydration shell

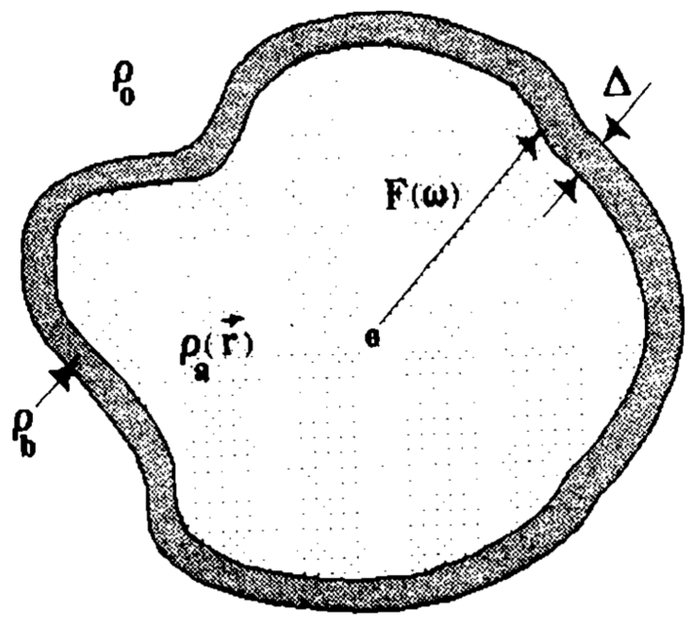


Fig. 1. Schematic representation of a macromolecule in solution. For explanations see text.

Figure 1 from CRYSOL paper:
Svergun, D. et al. (1995) *J. Appl. Cryst.* **28**, 768-773

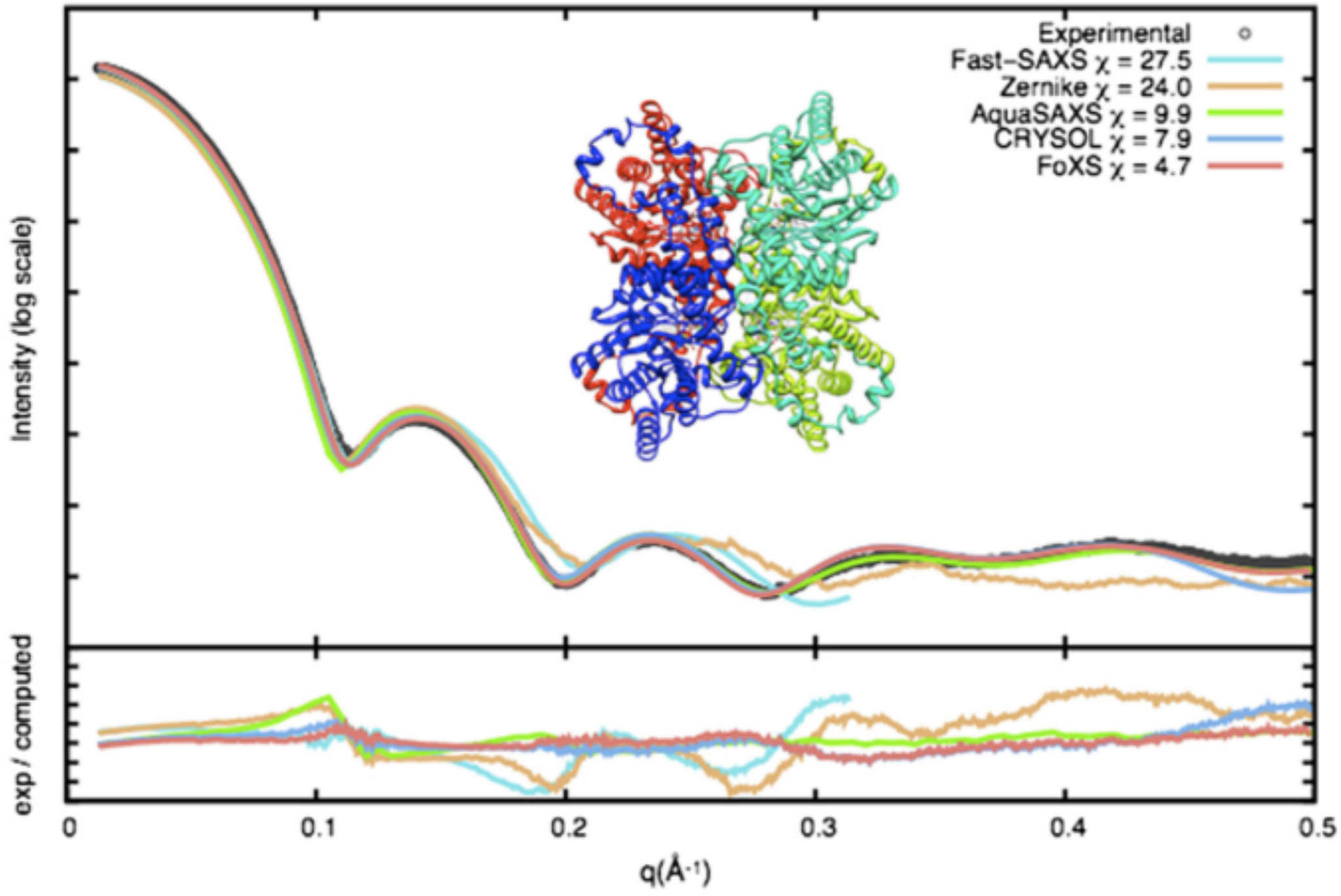
Calculating SAXS Profiles from Models

Table 1 Methods for theoretical profile calculation

Method	Spherical Averaging	Hydration Layer	Representation	Availability
CRYSOL [34]	Multipole expansion	Implicit water layer based on envelope function	Atomic	Server, download http://www.embl-hamburg.de/biosaxs/crysol.html
solX [35]	Debye formula	-	Atomic	
ORNL_SAS [36]	Monte-Carlo sampling	Implicit water layer	Grid representation	Download http://www.ornl.gov/sci/csd/Research_areas/MS_csmb_comp_methods.htm
SoftWAXS [37]	Numerical quadrature	Implicit water layer	Atomic	
Fast-SAXS [38]	Debye formula	Explicit placement of water molecules	Coarse-grained residue level	http://yanglab.case.edu/software.html
Park et al. [39]	Spherical quadrature	Explicit placement of water molecules	Atomic	
Stovgaard et al. [40]	Debye formula	-	Coarse-grained, 1 or 2 points per-residue	
AXES [41]	Numerical quadrature	Explicit placement of water molecules	Atomic	Server http://spin.niddk.nih.gov/bax/nmrserver/saxs1/
FoXS [42]	Debye formula	Implicit water layer based on surface accessibility	Atomic or coarse-grained residue level	Source code, server, download, Chimera http://salilab.org/foxs/
AquaSAXS [43]	Cubature formula	AquaSol solvent density map	Atomic	Server http://lorentz.dynstr.pasteur.fr/aquasaxs/aquasaxs_submission.php
Virtanen et al. [44]	Debye formula or Cube model	HyPred based on MD simulations	Atomic, MD simulation	
Zernike Polynomials [45]	Zernike polynomial expansions	Hydration layer from voxelized representation	Atomic	Source code, server, download http://sastbx.als.lbl.gov/cgi-bin/intensity.html

Schneidman-Duhovny, et.al. (2012) *BMC Structural Biology*

Calculating SAXS Profiles from Models

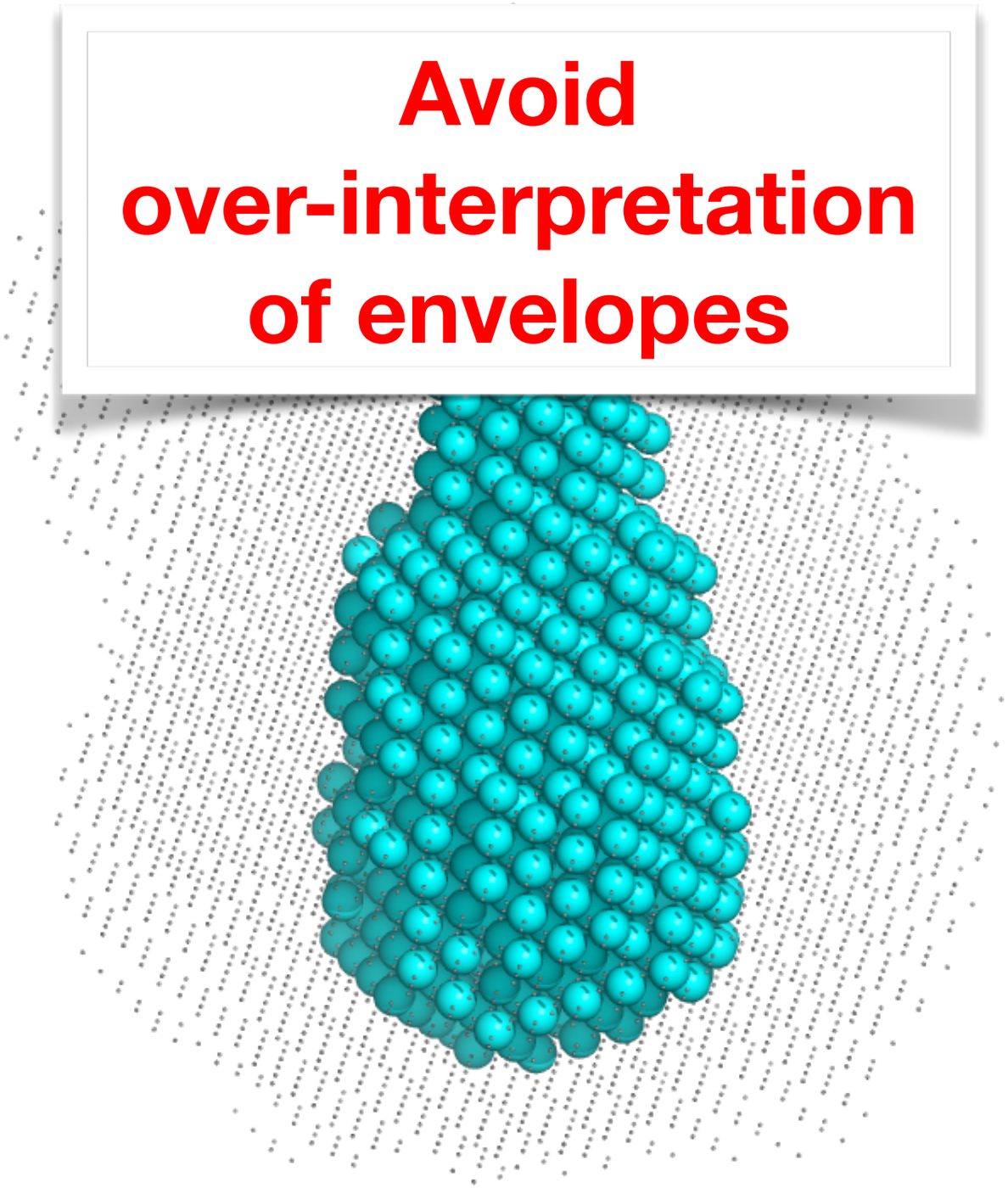


Schneidman-Duhovny, et.al. (2012) *BMC Structural Biology*

Ab initio Envelope Reconstruction

- Several programs exist for *ab initio* envelope reconstructions, most common is DAMMIN/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

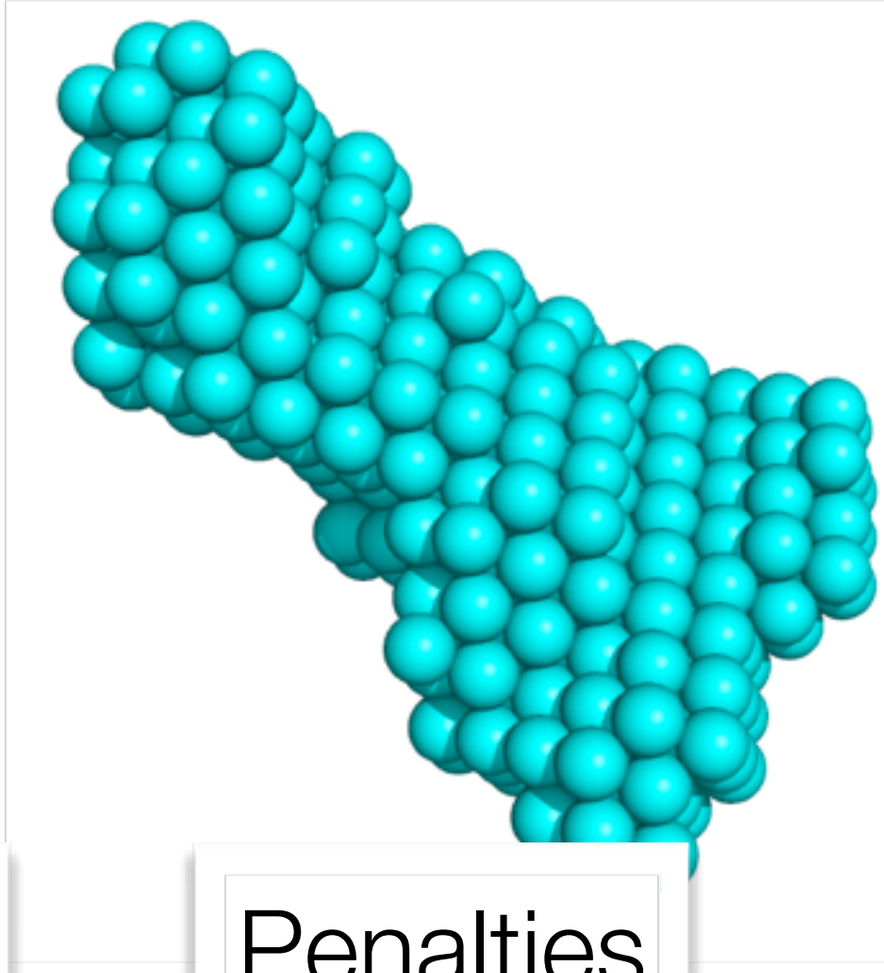
Avoid over-interpretation of envelopes



Envelope Reconstruction

$$\chi^2 = \frac{1}{N}\sum_{i}^{N}\left(\frac{I_{exp}(q_i) - I_{calc}(q_i)}{\sigma_i}\right)^2$$

- 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



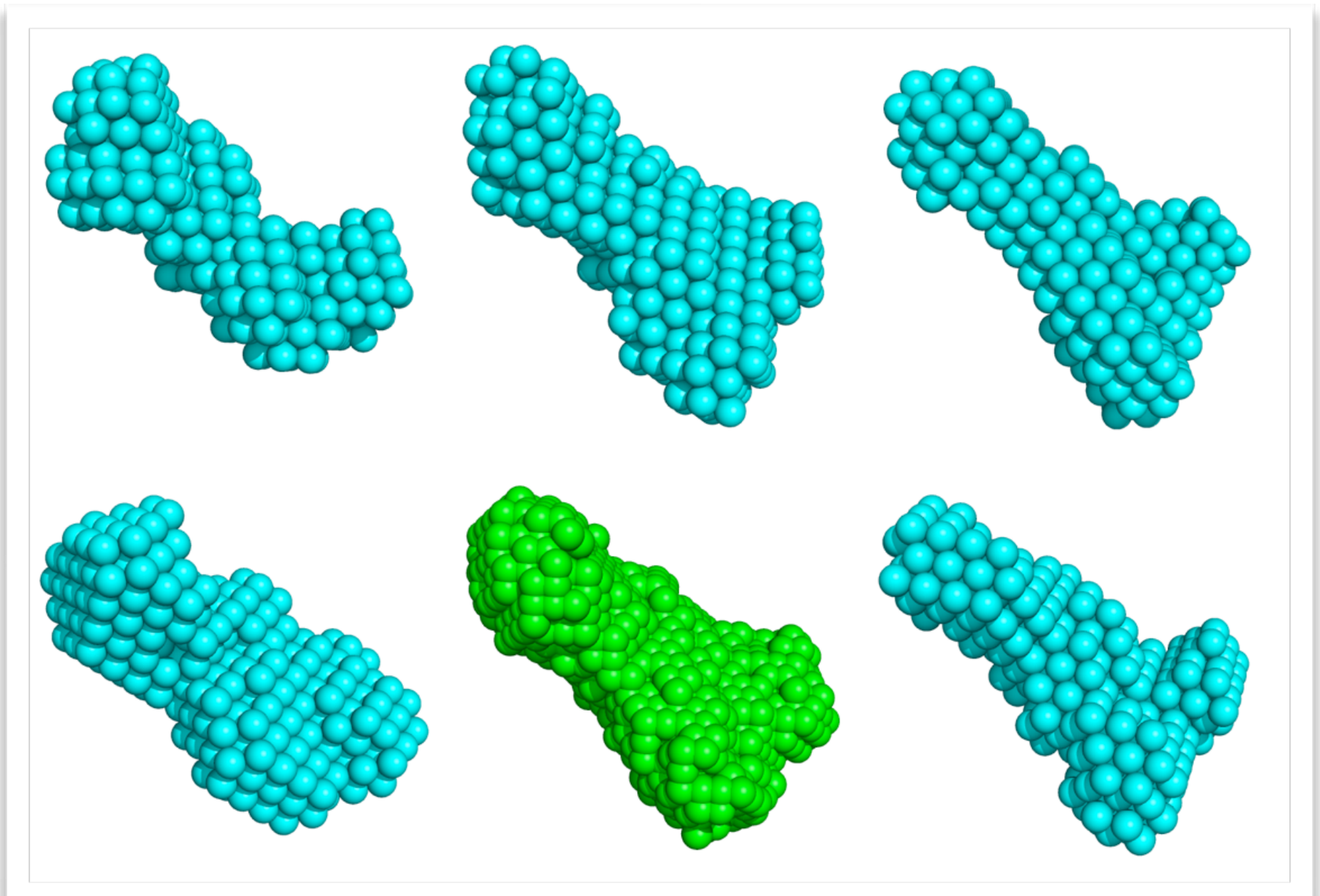
Fit to data

Penalties

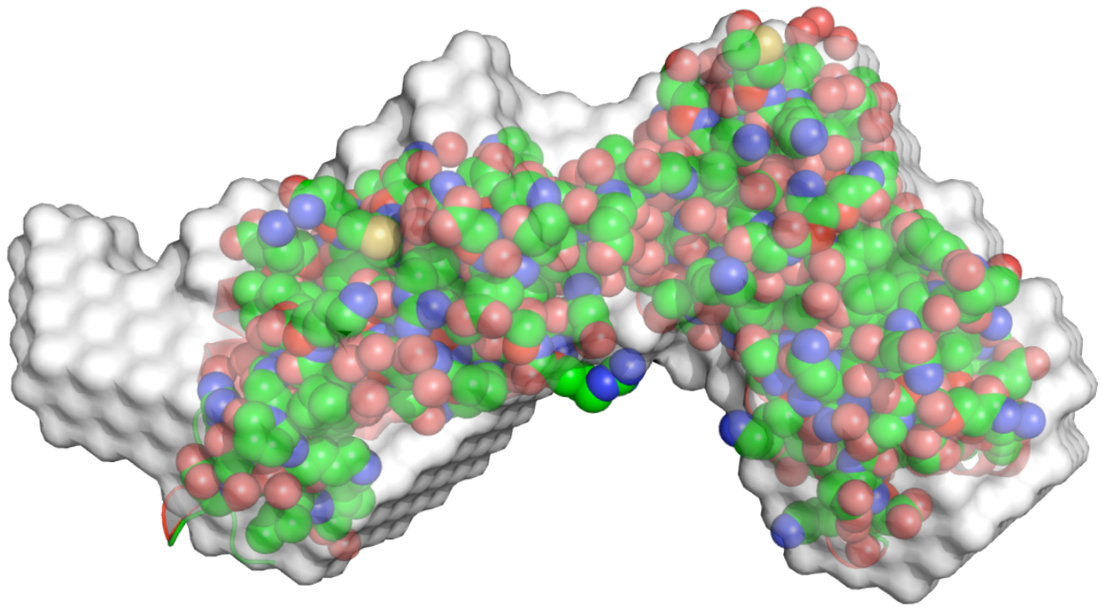
$$Score = \chi^2\,[I_{exp}(s), I_{calc}(s)] + \alpha P(x)$$

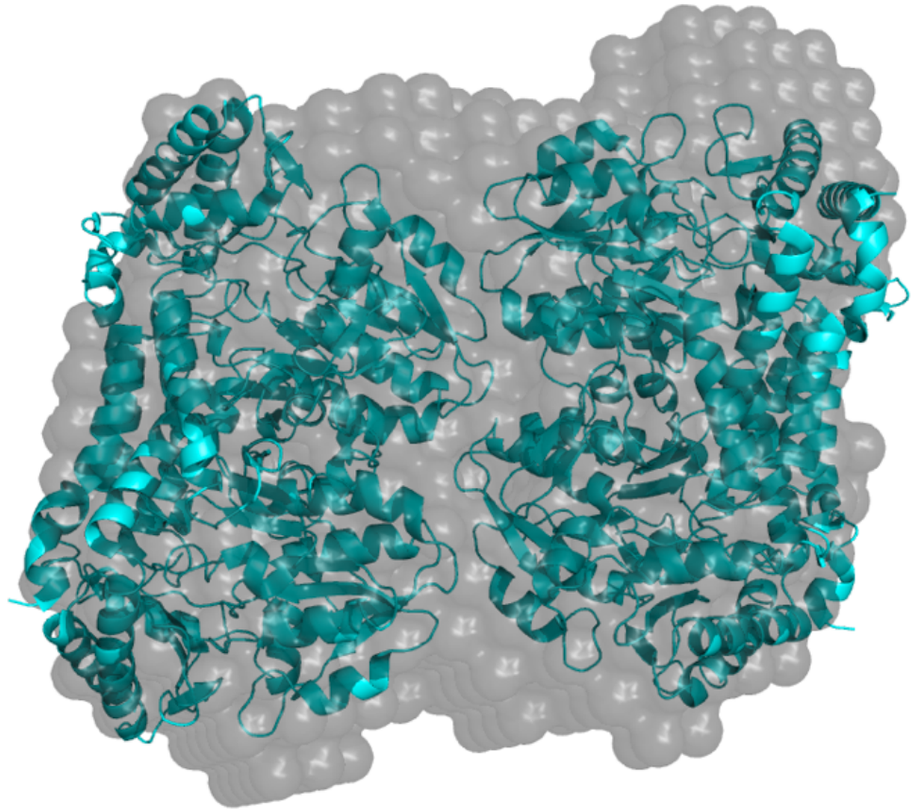
Envelope Reconstruction

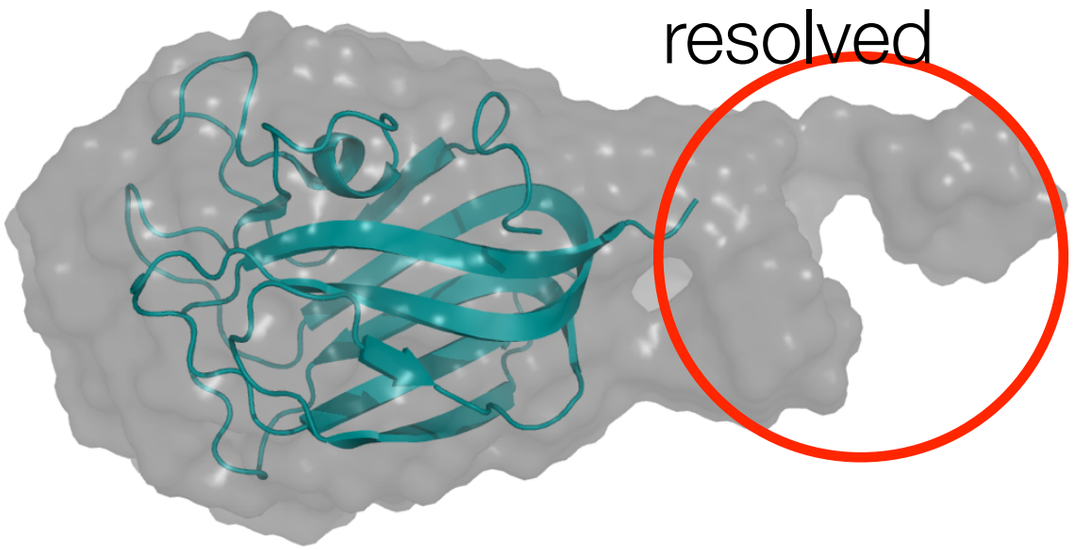
- Averaging with DAMAVER (typically 10-15 bead models) results in a “consensus” model, i.e. where the beads typically appear



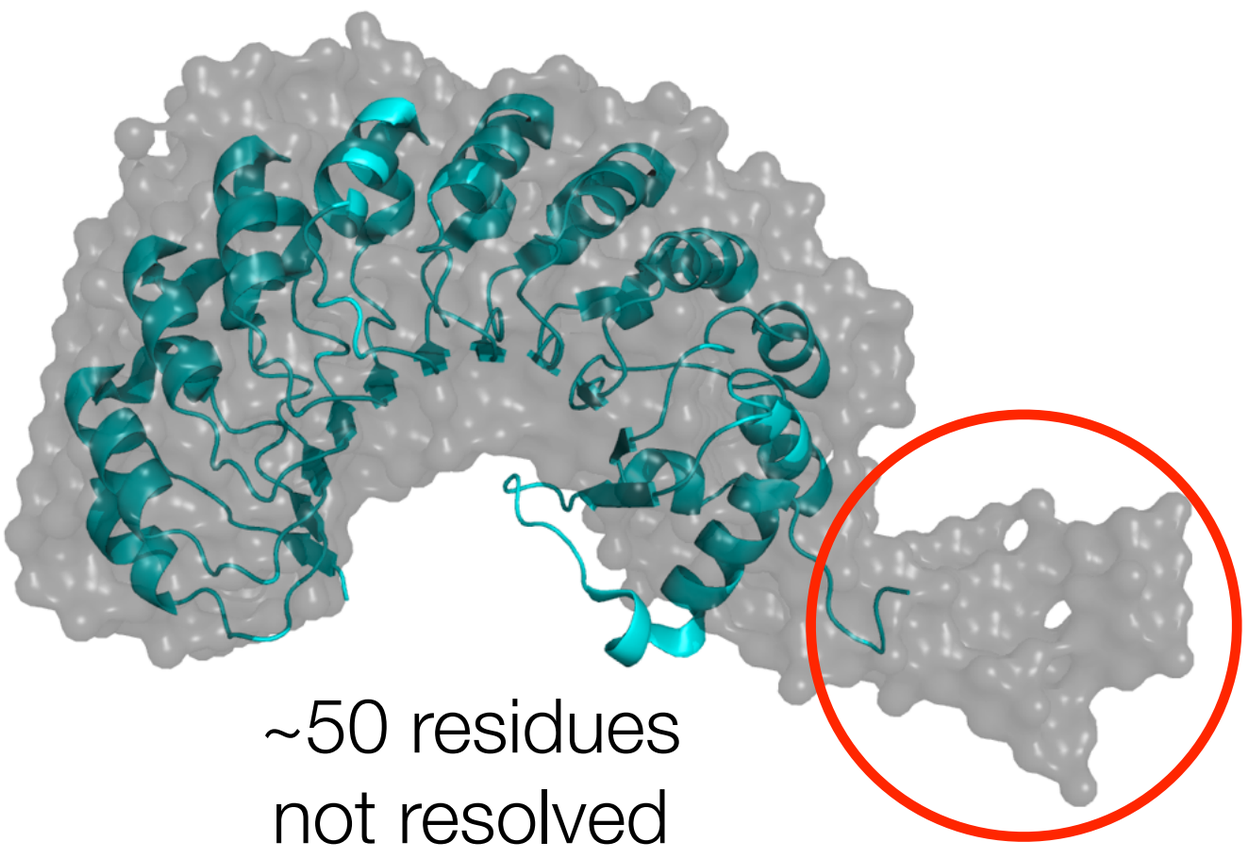
Examples







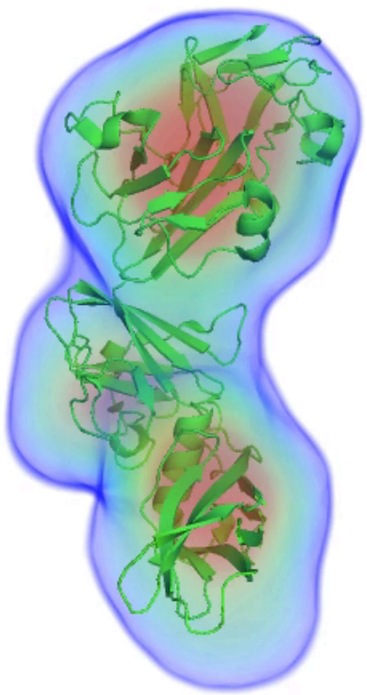
12 residues not resolved

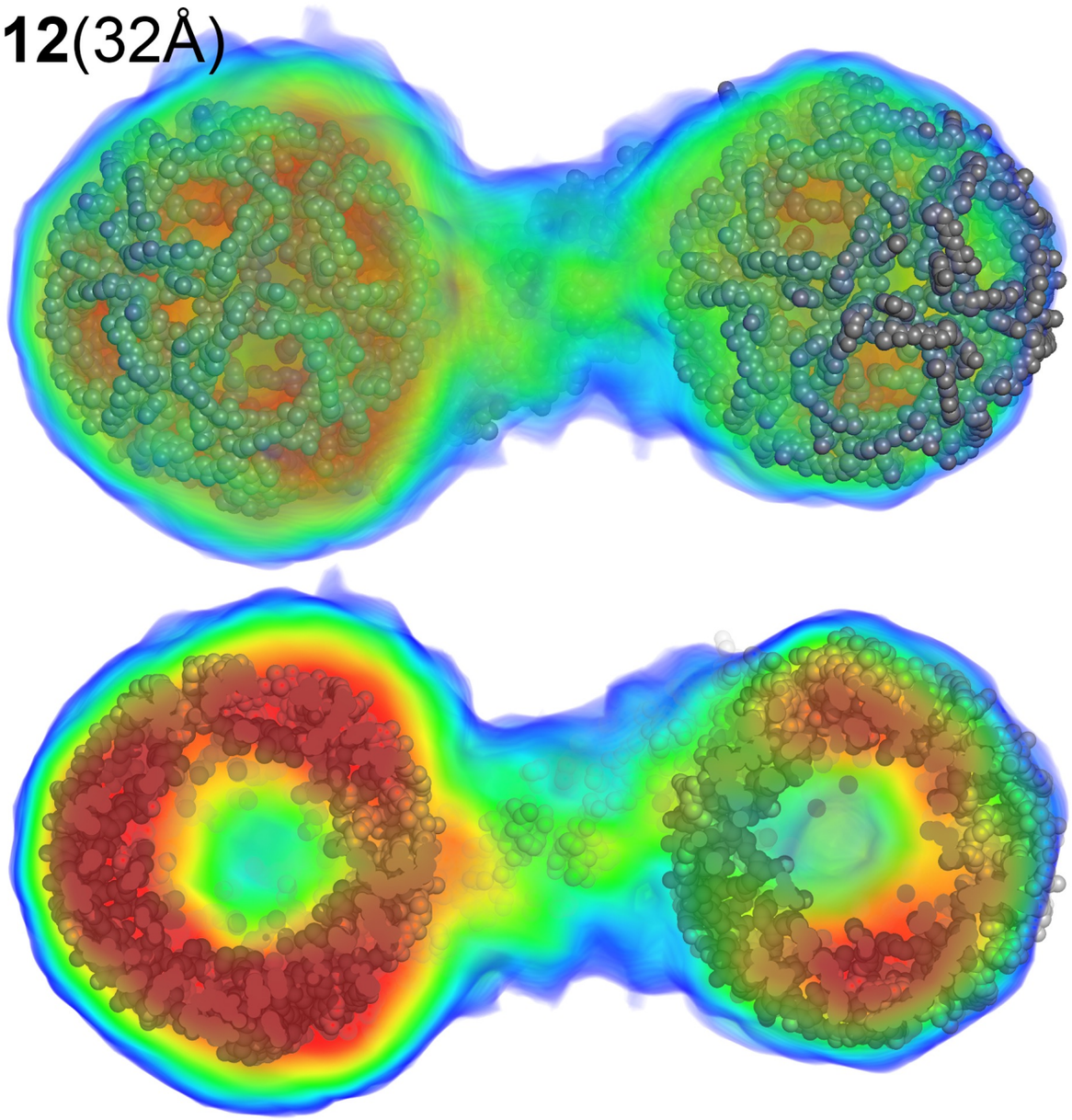


~50 residues not resolved

Ab initio 3D Density Reconstruction (DENSS)

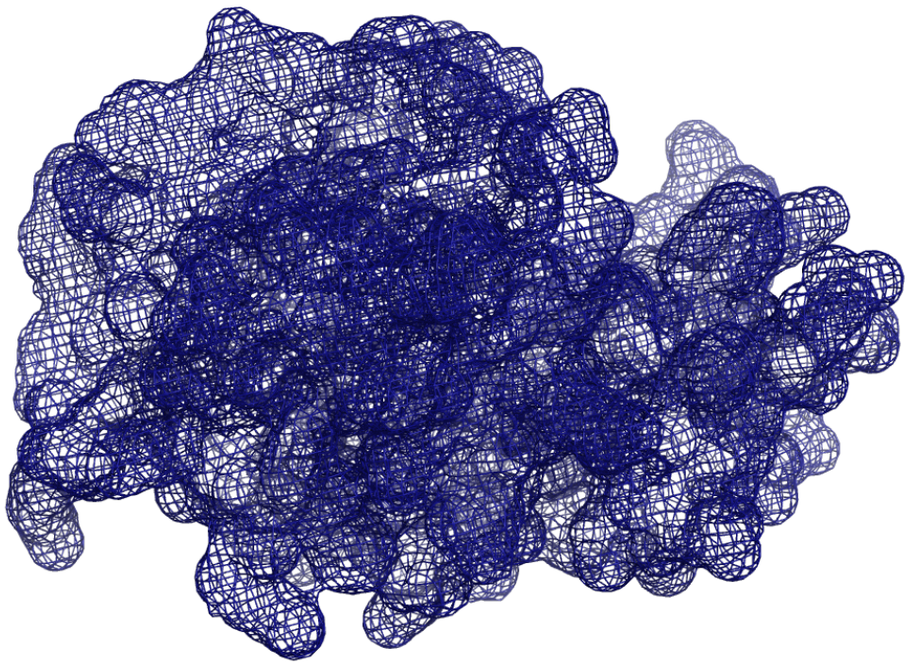
- Fundamentally different approach than bead modeling (solves the inverse scattering problem)
- DENSS calculates *density*
- Can model multiple different particle electron densities (e.g. protein-lipid)





Grant, T.D. (2018) *Nature Methods*

Electron Density

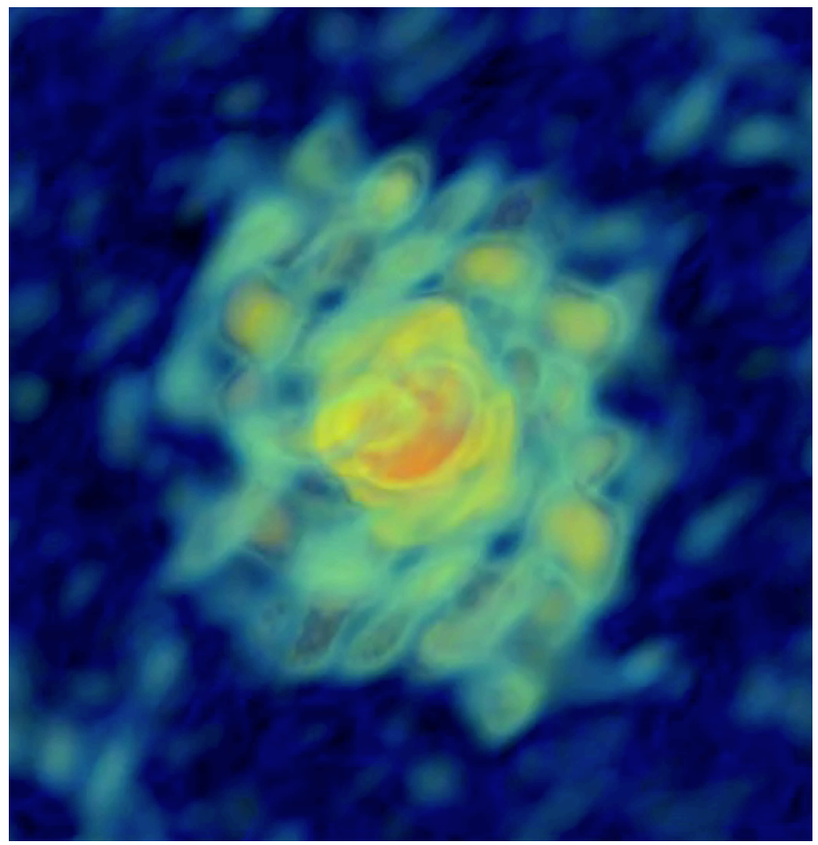


Fourier Transform

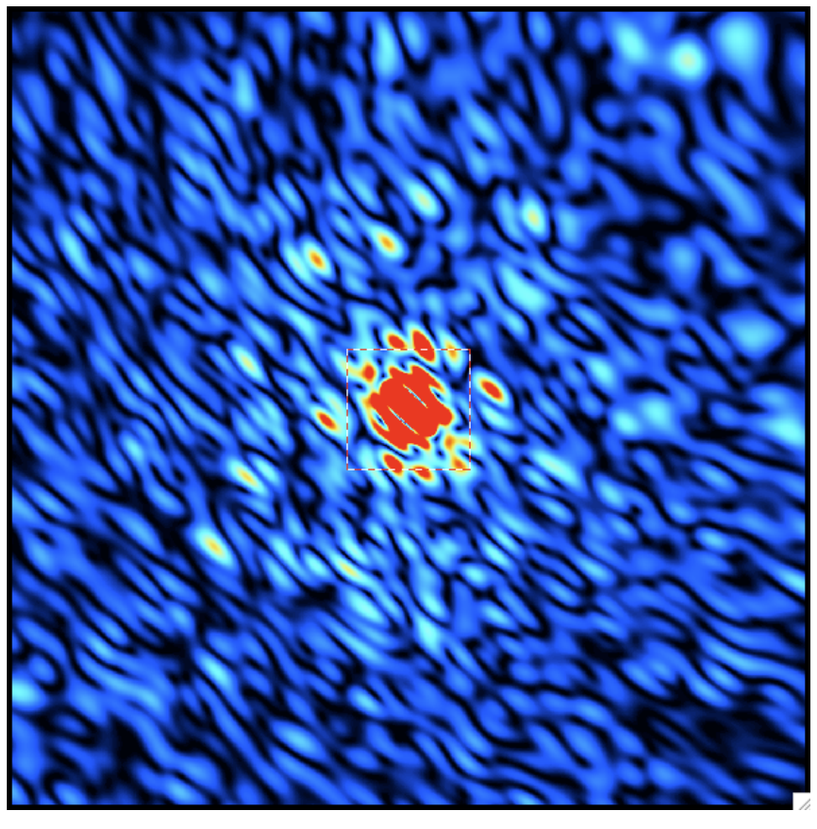
$\mathcal{F}$

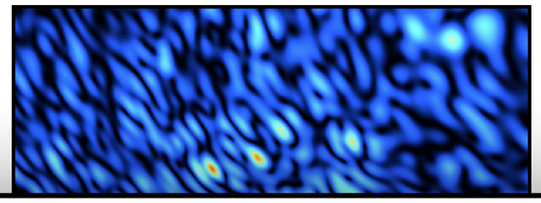
$\mathcal{F}^{-1}$

Molecular Transform



What it would look like on a 2D detector (cross section through origin of molecular transform)



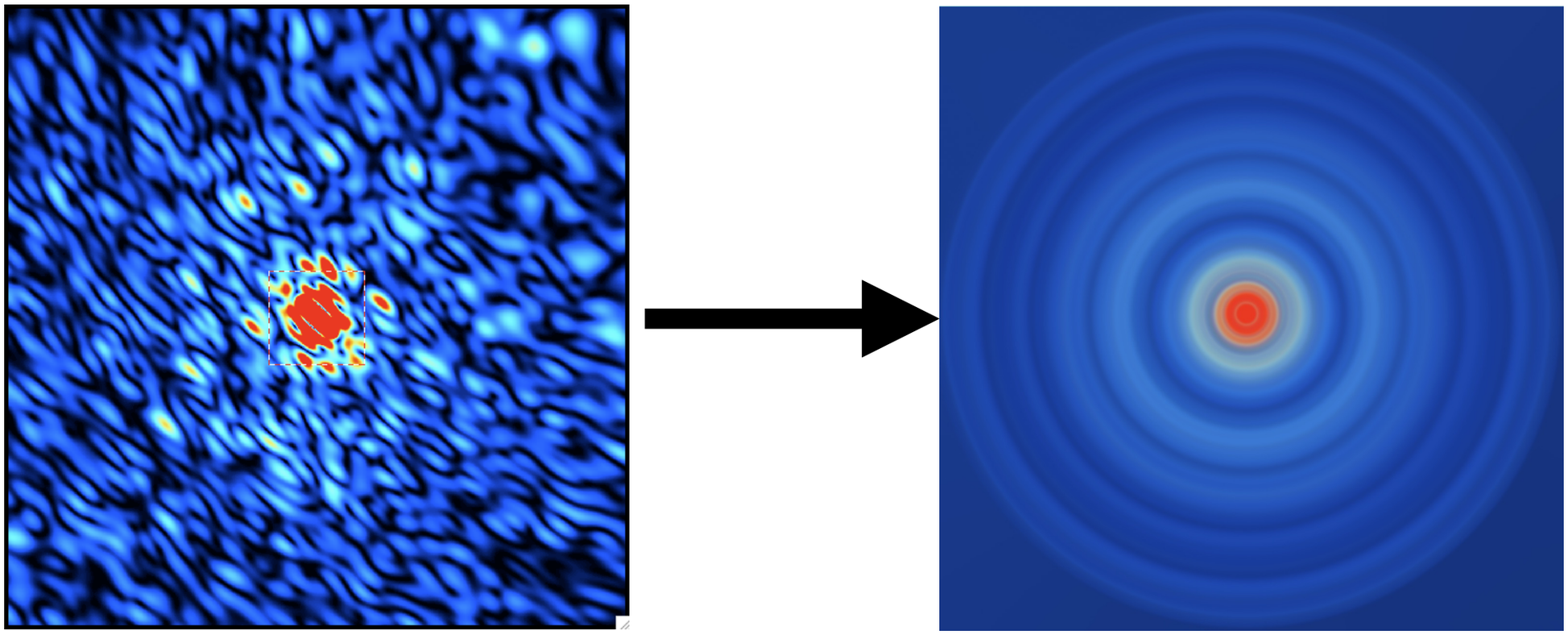


In single molecule imaging,
3D amplitudes known.
Must solve the
“**phase problem**”

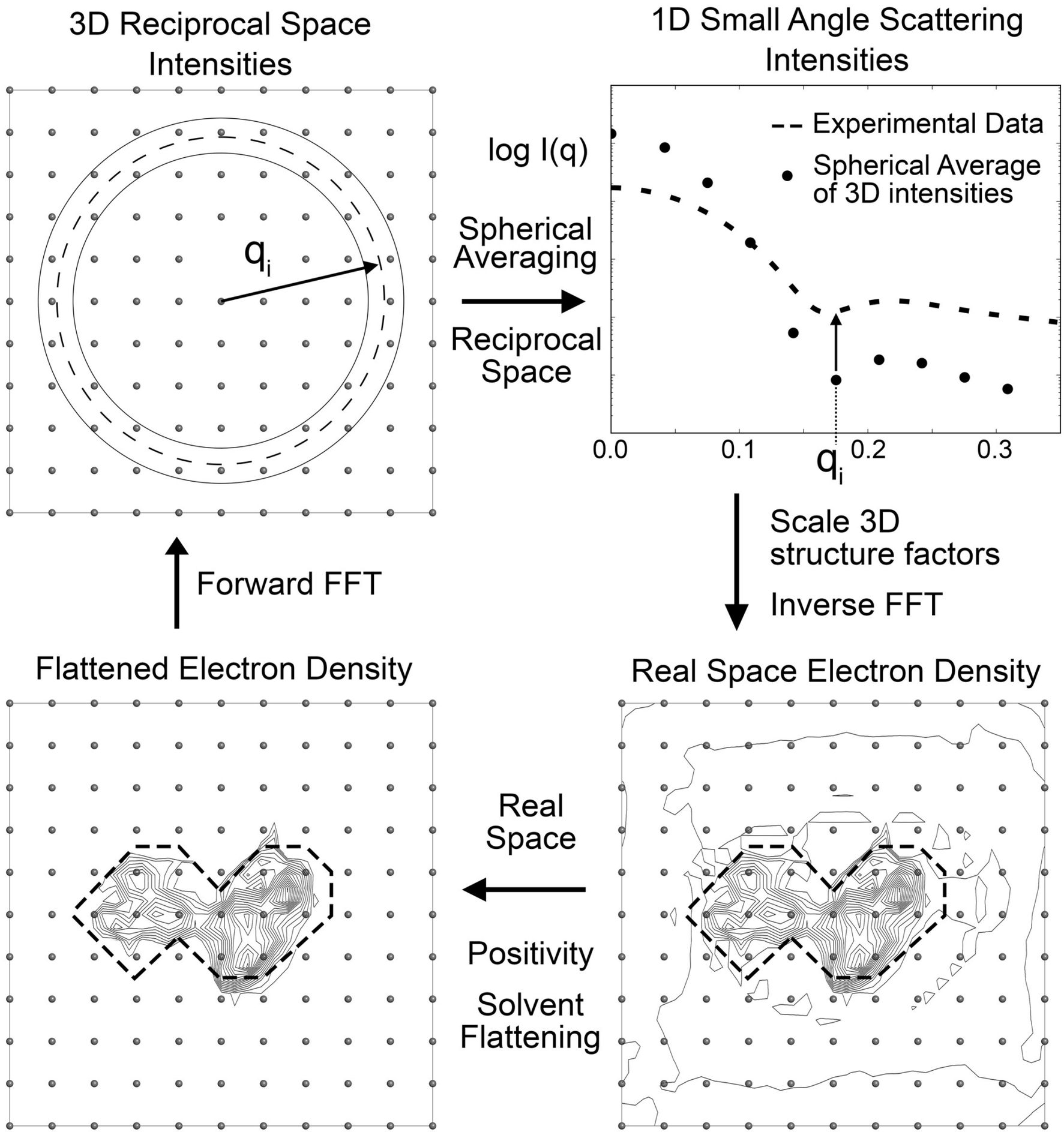
Molecular Transform
(structure factors)

Spherical averaging
from solution of
tumbling molecules

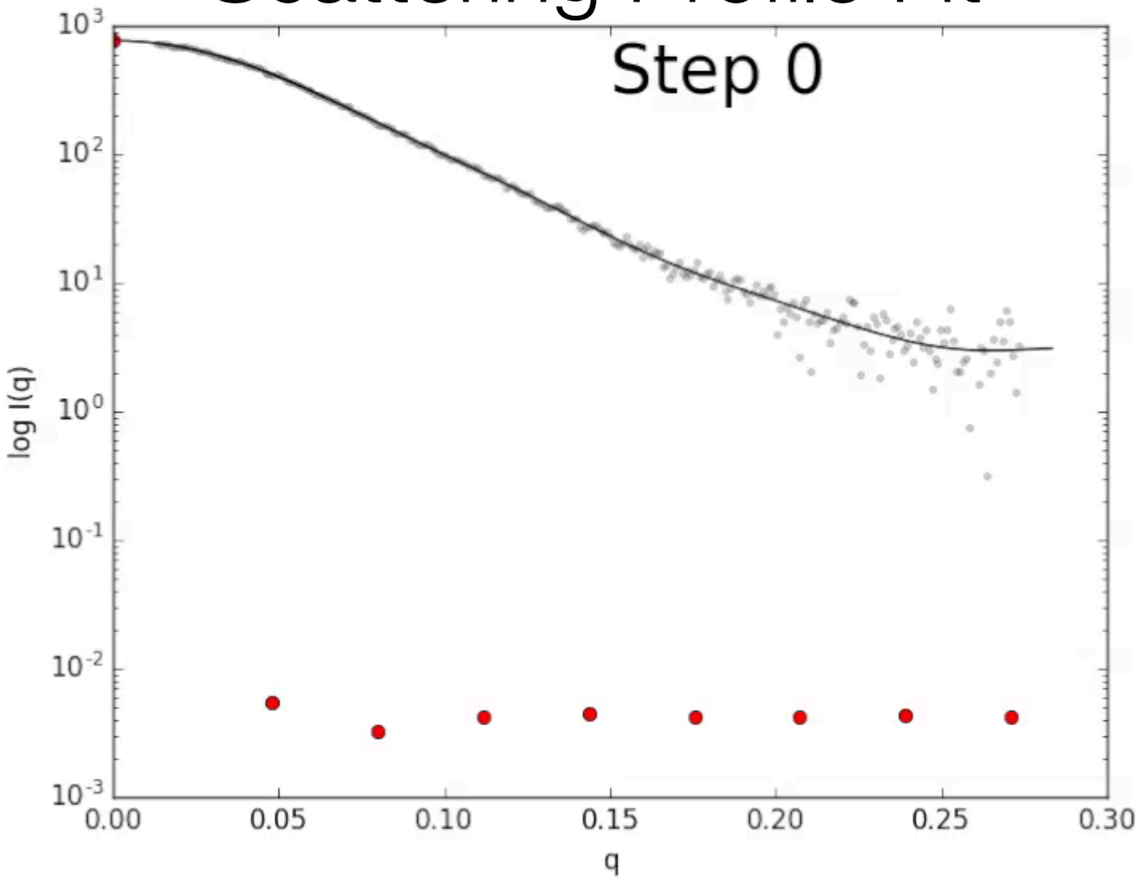
No 3D intensities (only 1D):
Instead of the
“**phase problem**”,
solve the
“**structure factor problem**”



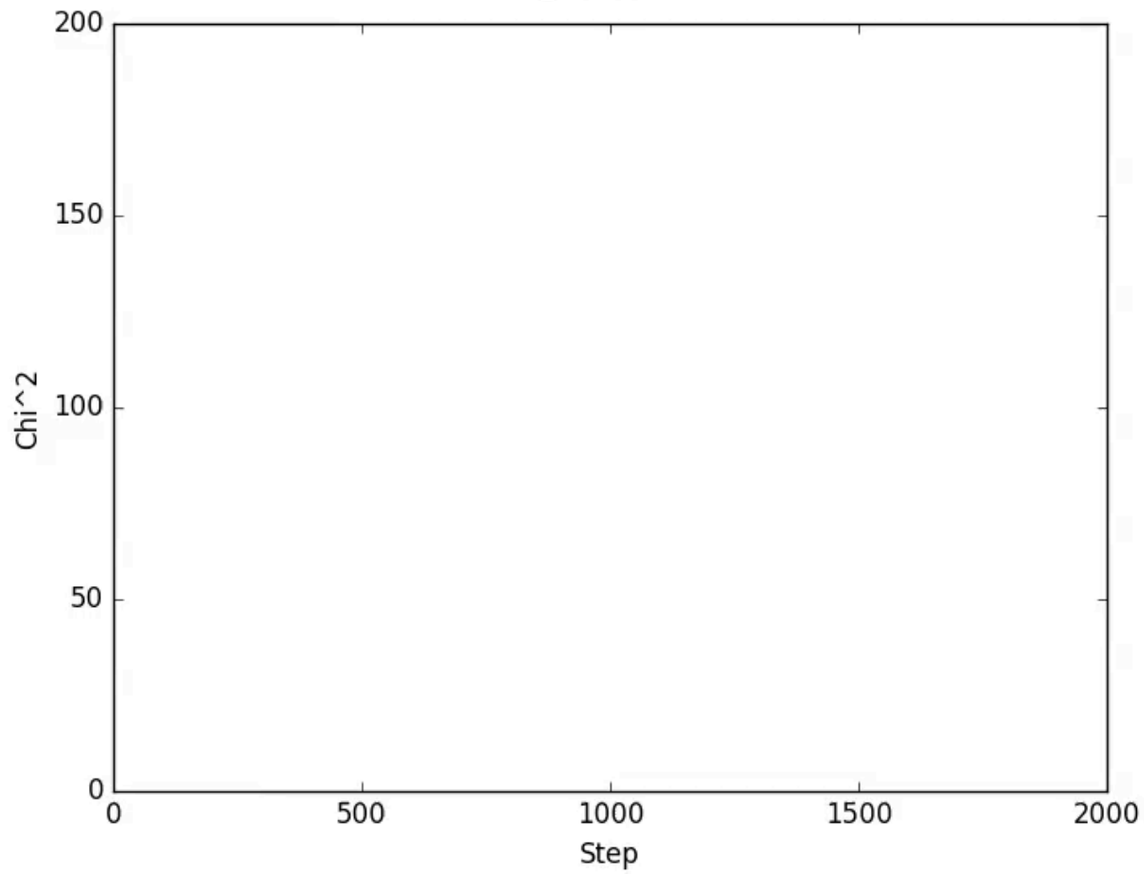
Iterative Structure Factor Retrieval Algorithm



Scattering Profile Fit

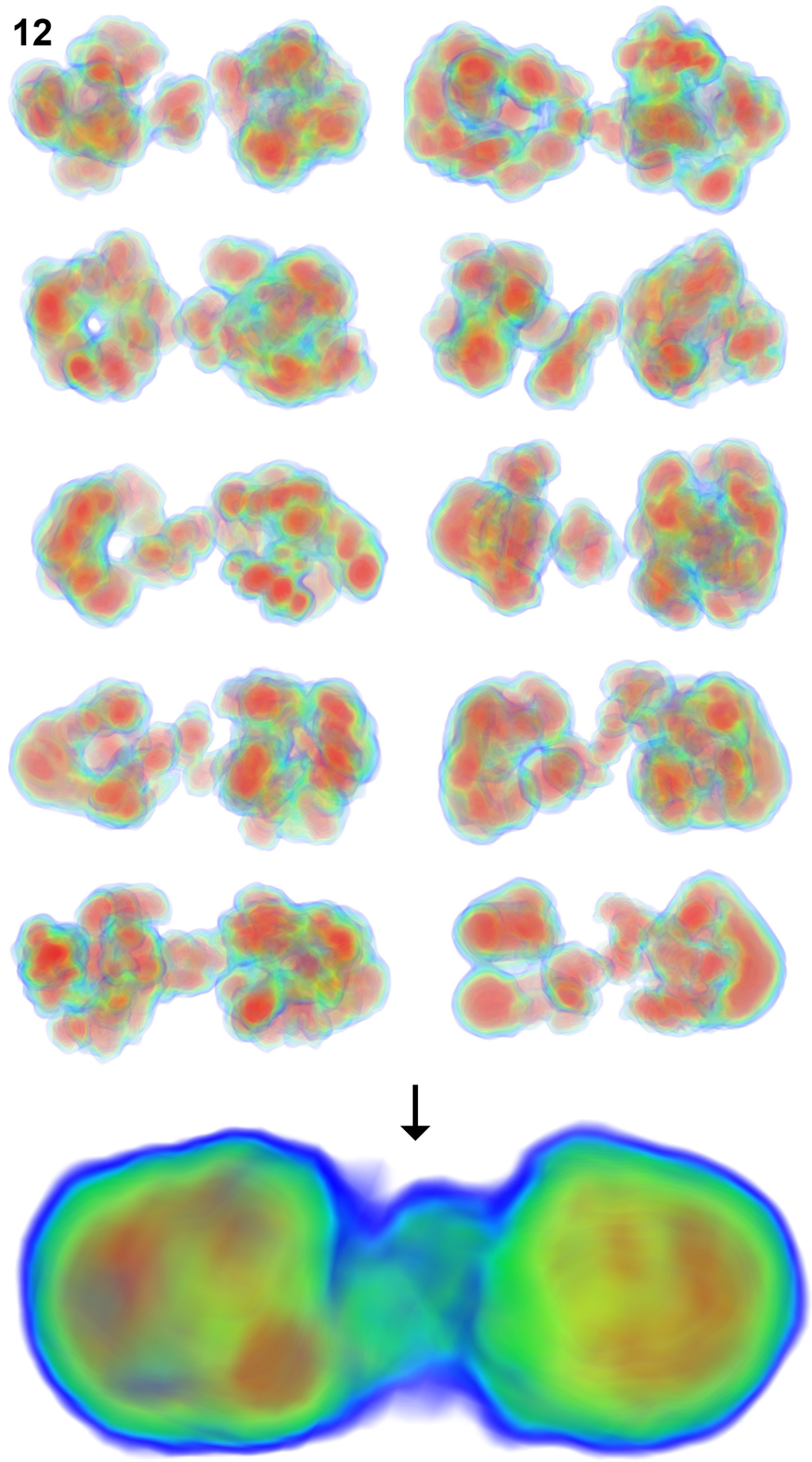


Chi^2



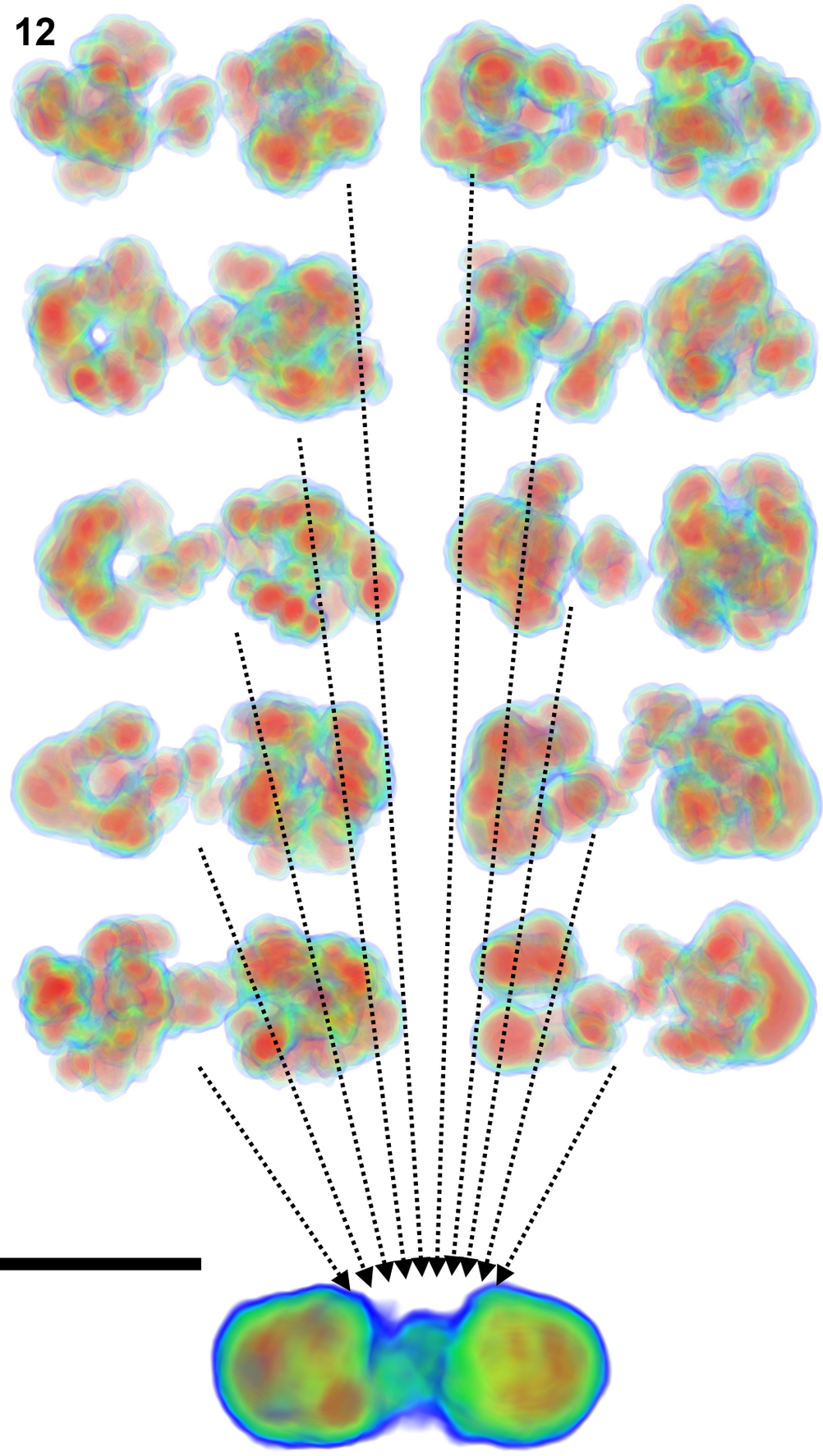
Averaging

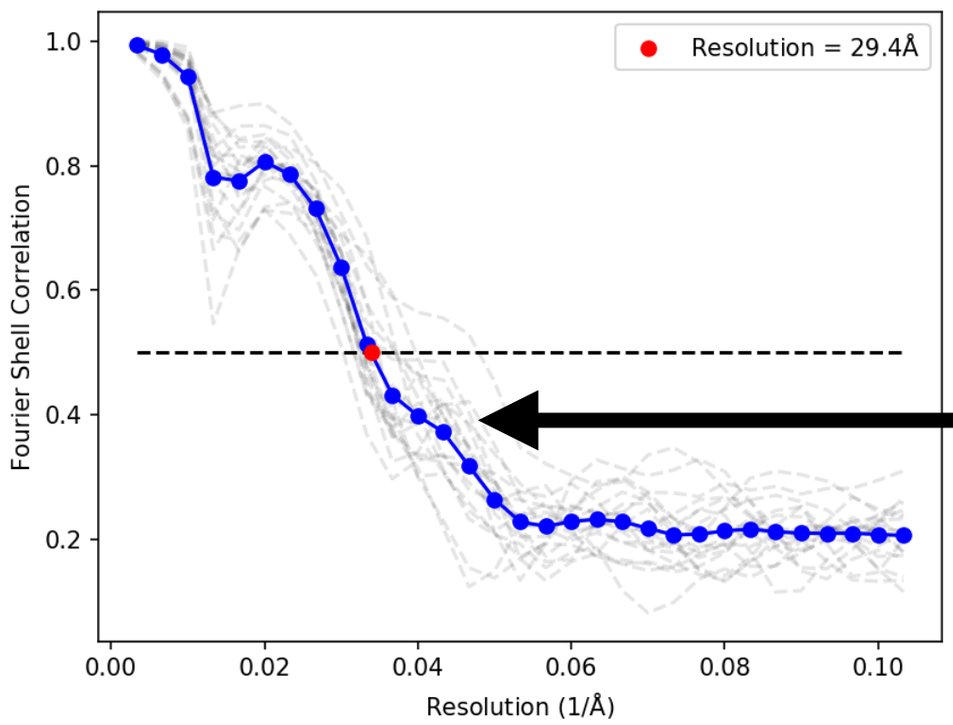
- 3D reconstructions from 1D data → non-unique solutions
- Run multiple reconstructions with different random seeds (typically 20 - 100 times)
- Align, average to produce final density



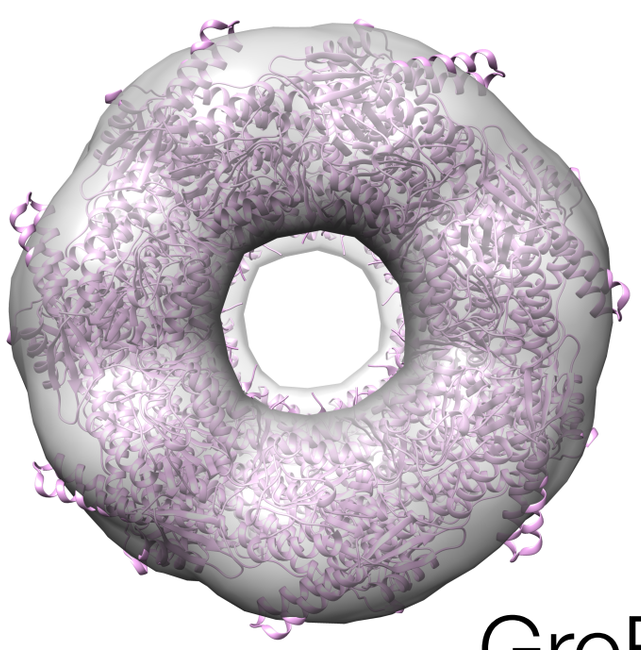
Resolution

- Align all individual reconstructions to some reference (e.g. a single random reconstruction, or a preliminary average)
- Compare via Fourier Shell Correlation

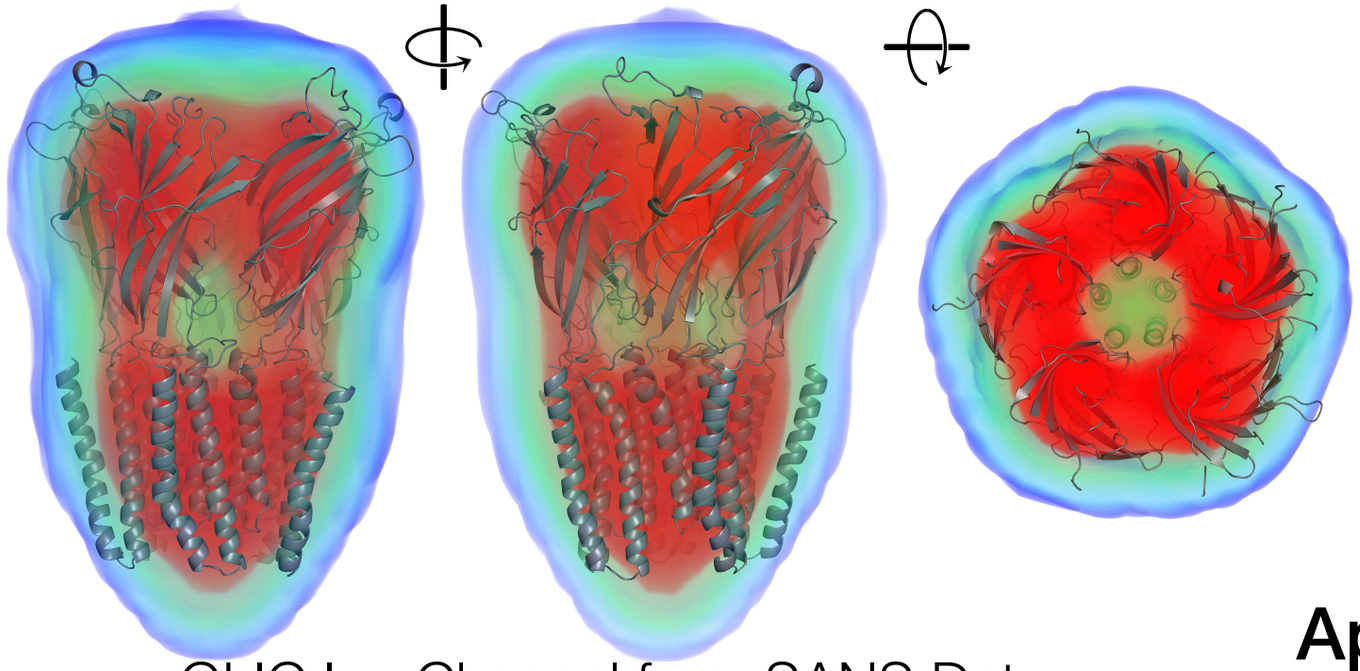




DENSS Examples

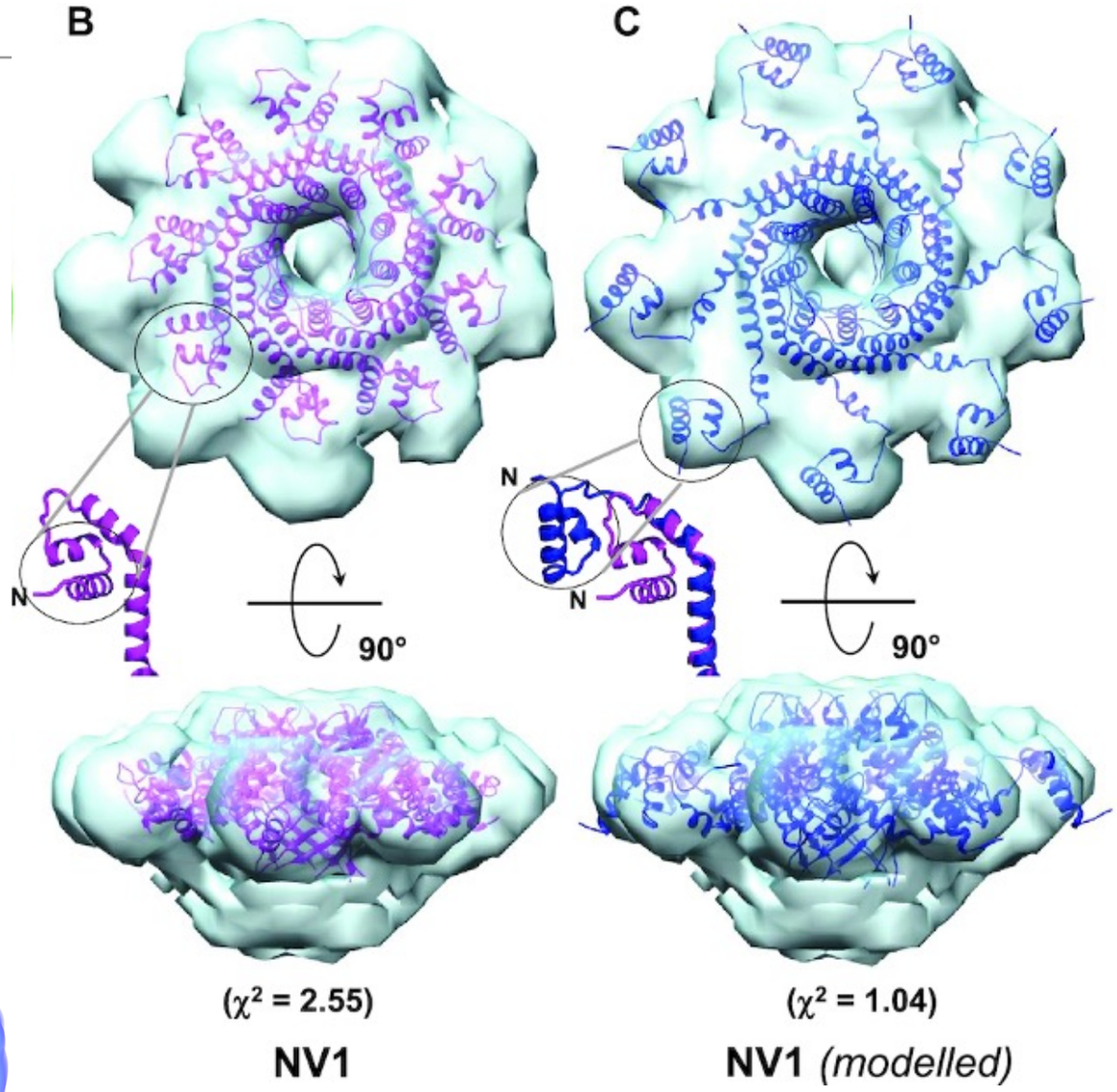


GroEL



GLIC Ion Channel from SANS Data
(SASDL33, Lycksell, et. al. 2021)

Open/Closed conformations of DNA-binding Protein NV1



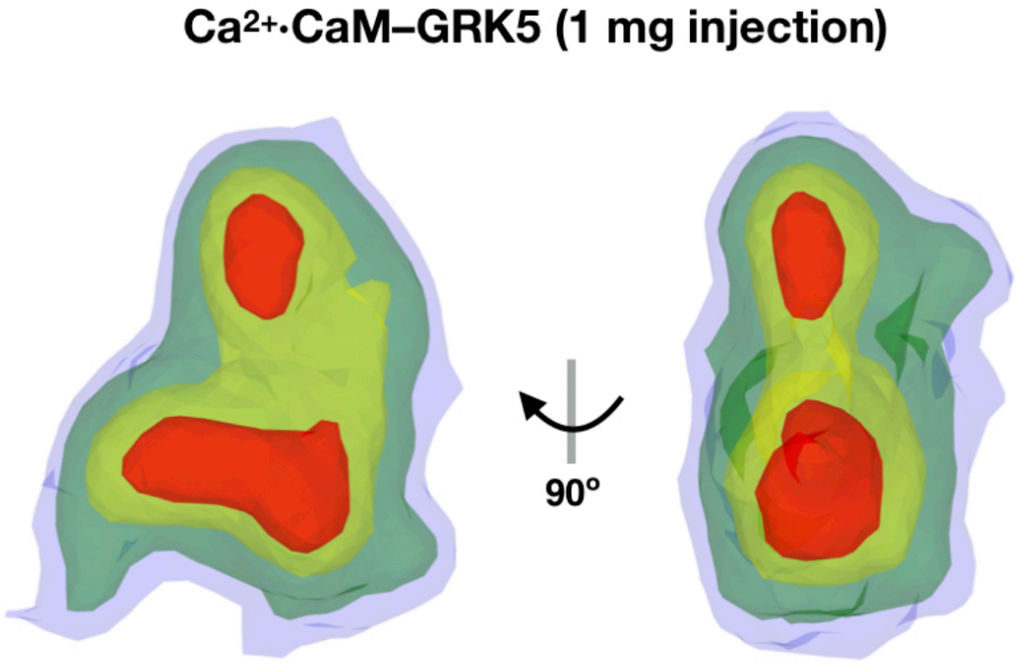
Niazi, et. al. (2020) *Nucleic Acids Research,* **48** (20), pg 11721-36

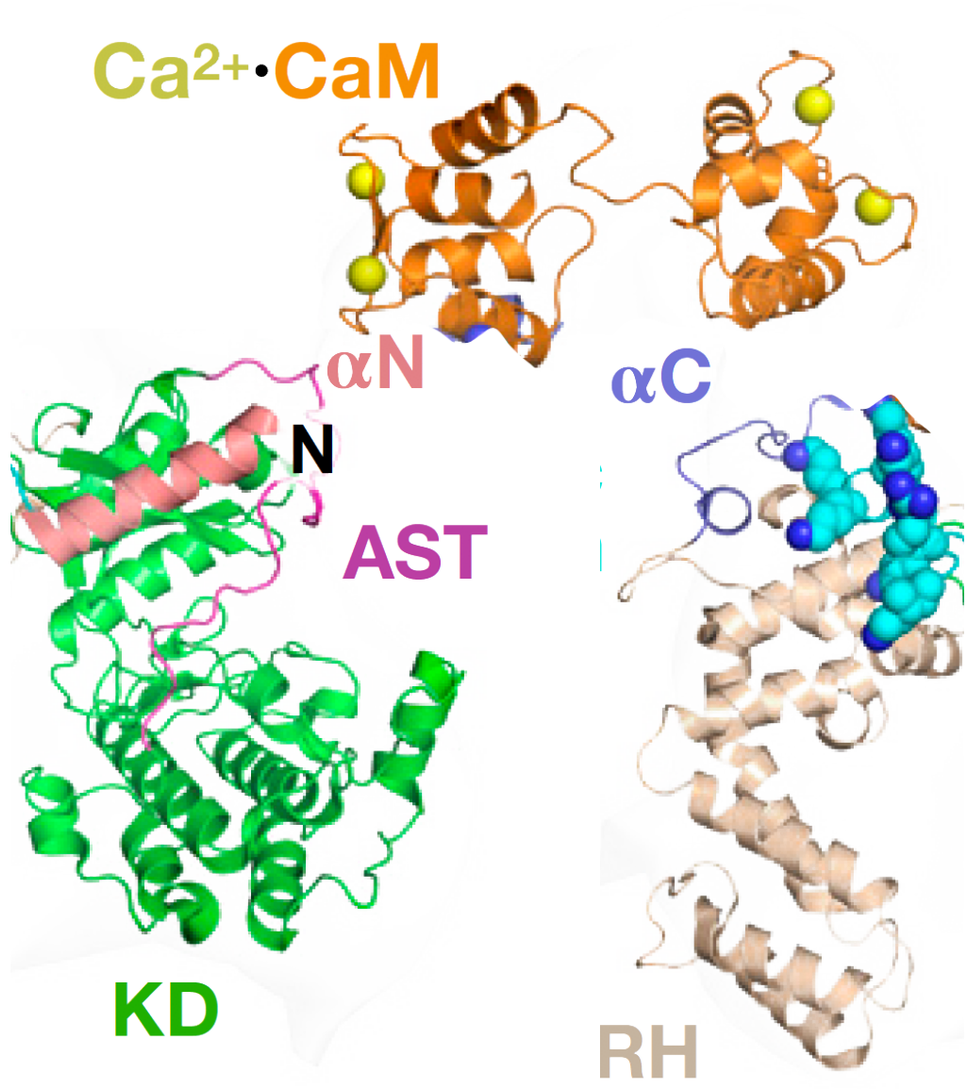
Applying symmetry significantly improves reconstructions

Modeling Density

- DENSS returns MRC formatted electron density map
  - Standard format for cryoEM reconstructions
- Many tools built for cryoEM can also be used for DENSS reconstructions
- Rigid body modeling into density

Bayett, et. al. (2019) *PNAS*



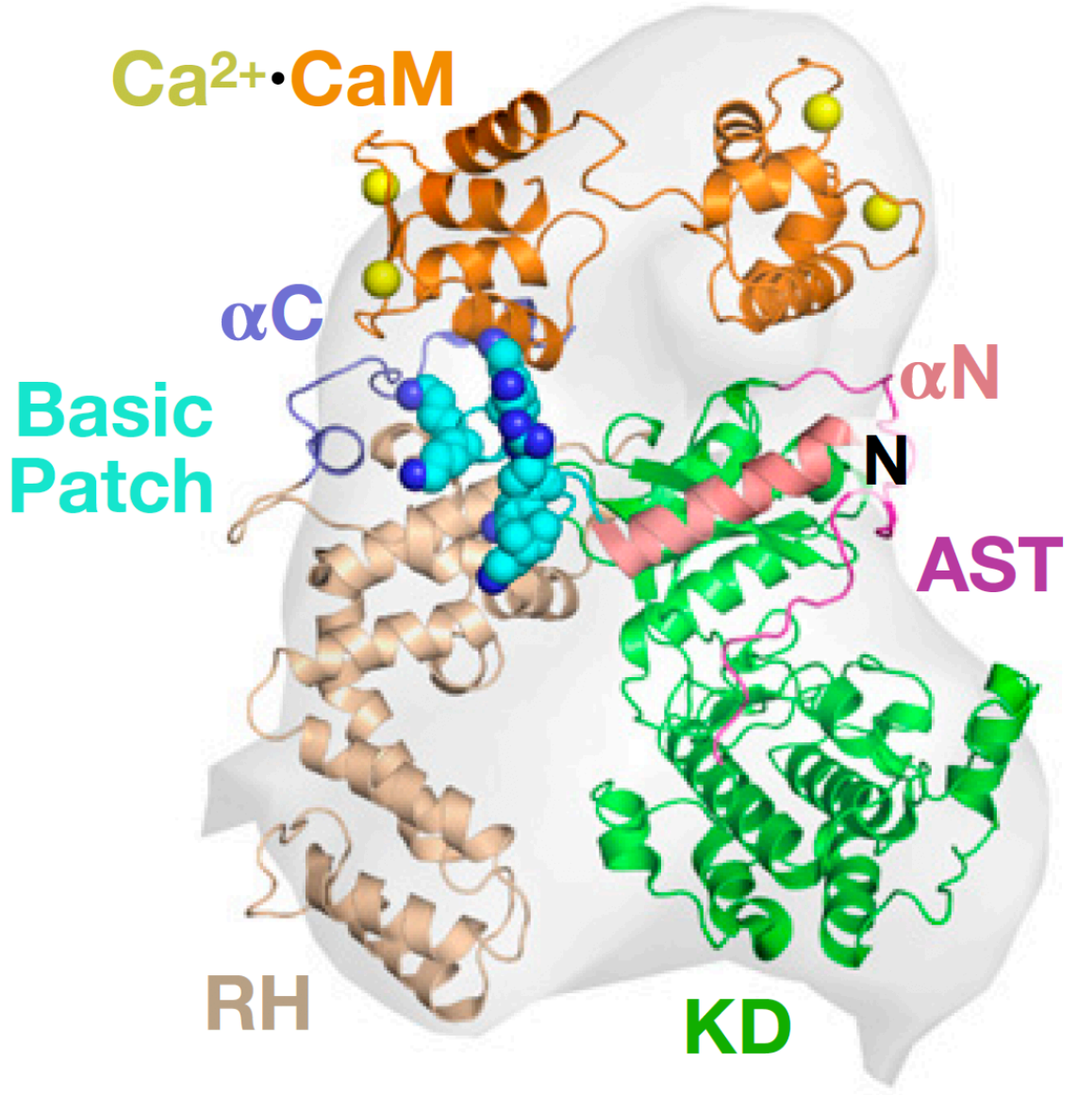


Rigid Body Modeling

- Tools available include Chimera, phenix.dock_in_map, Situs, COOT, and others

- Search density for best fit of each model in turn

- Calculate scattering profile from fitted model, compare to experimental data



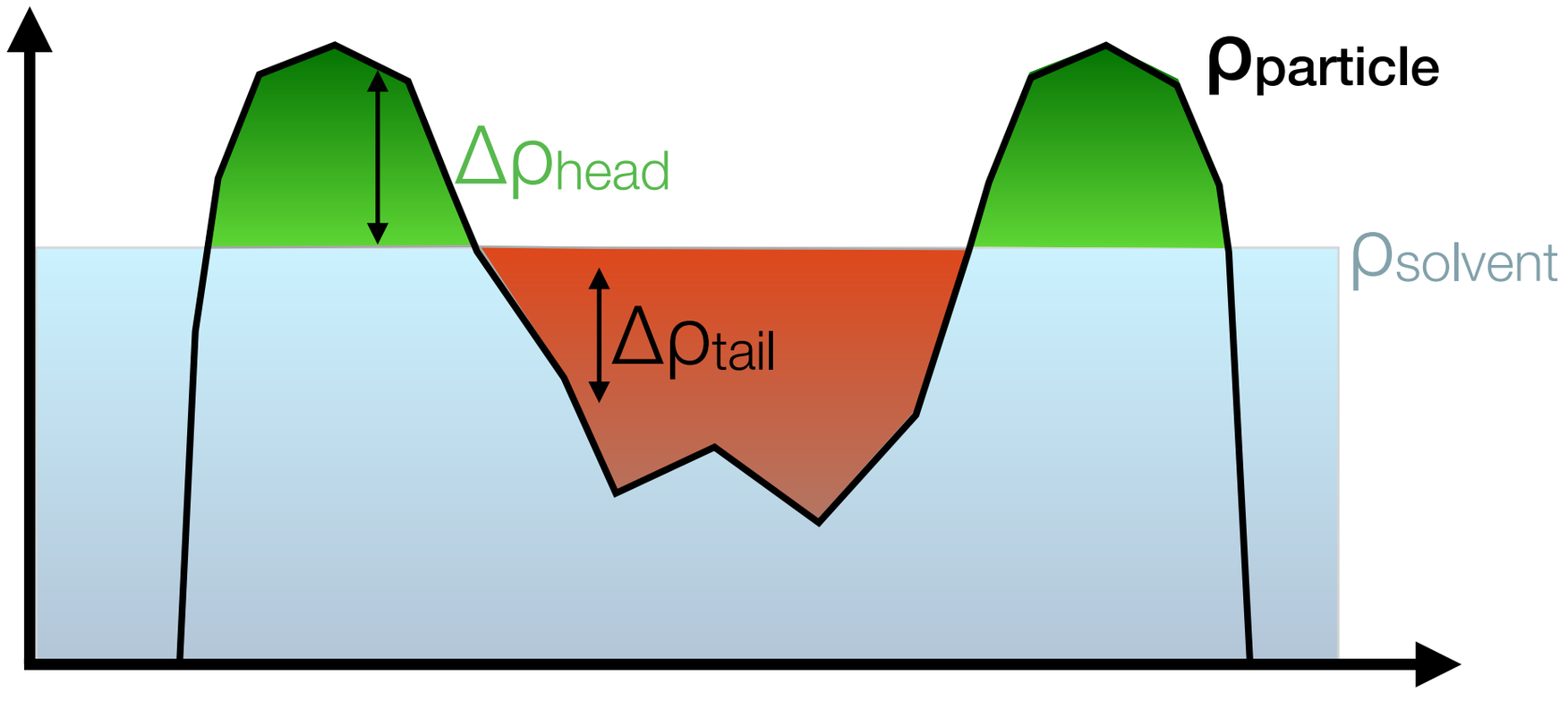
$\chi^2 = 1.07$

Bayett, et. al. (2019) *PNAS*

Using phenix.dock_in_map

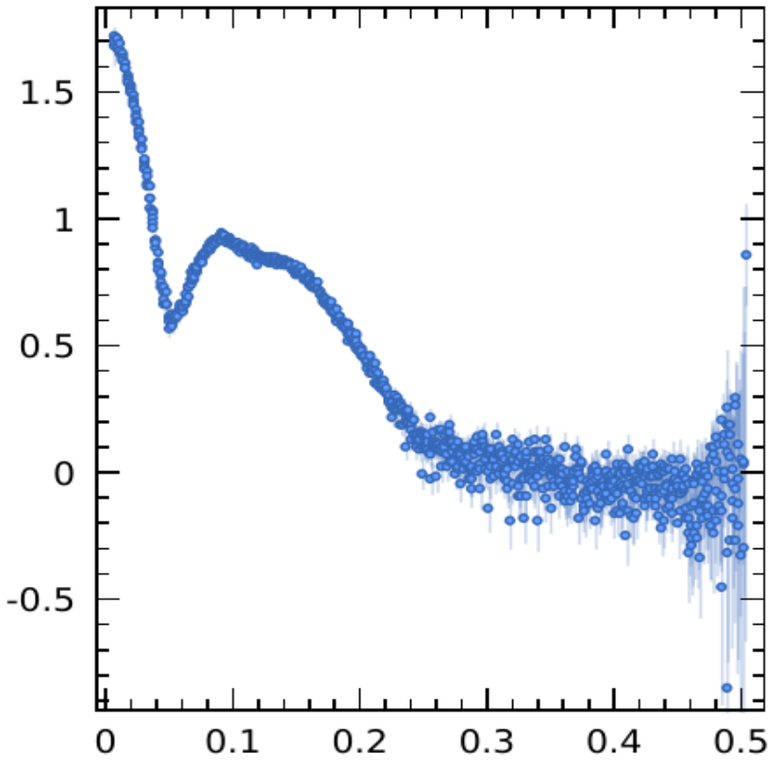
Lipids and Detergents

- Particles with lipids often have negative contrast (lipid tails)
- Micelles or other lipid-containing particles often fail with conventional modeling, since negative contrast can't be modeled
- Remove positivity restraint in DENSS to allow negative contrast

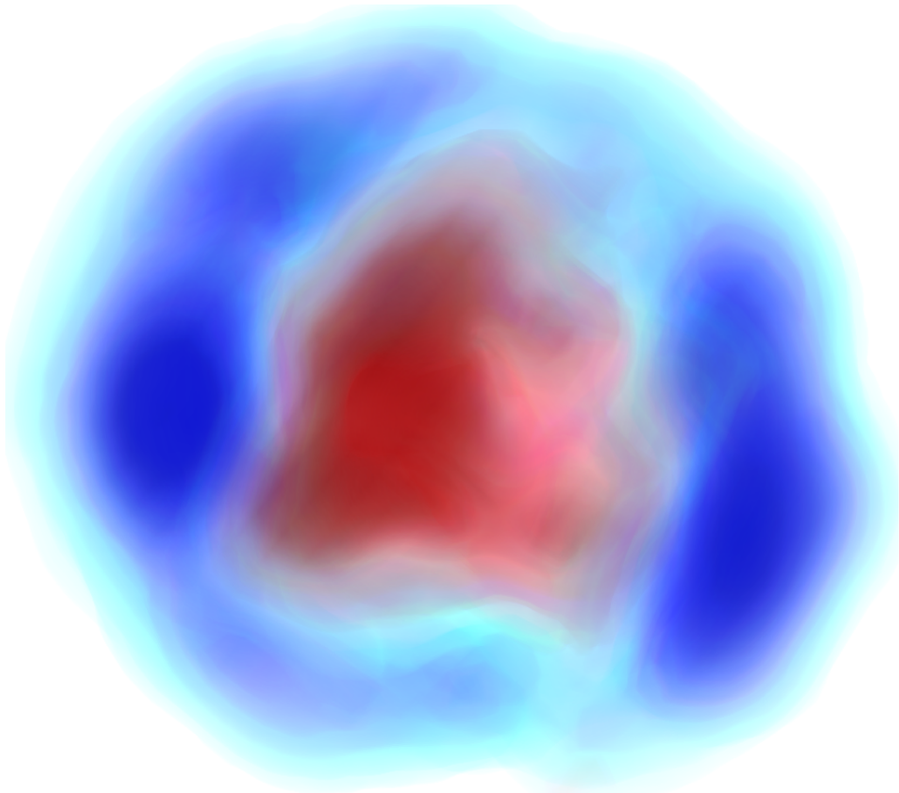


Nanodisc SAXS

- Called “MEMBRANE” mode
- Rely only on shrink-wrap and solvent flattening
- Works for any particle with negative contrast, not just membrane proteins
- Works with SANS data also (out of the box, no modifications)



Top View



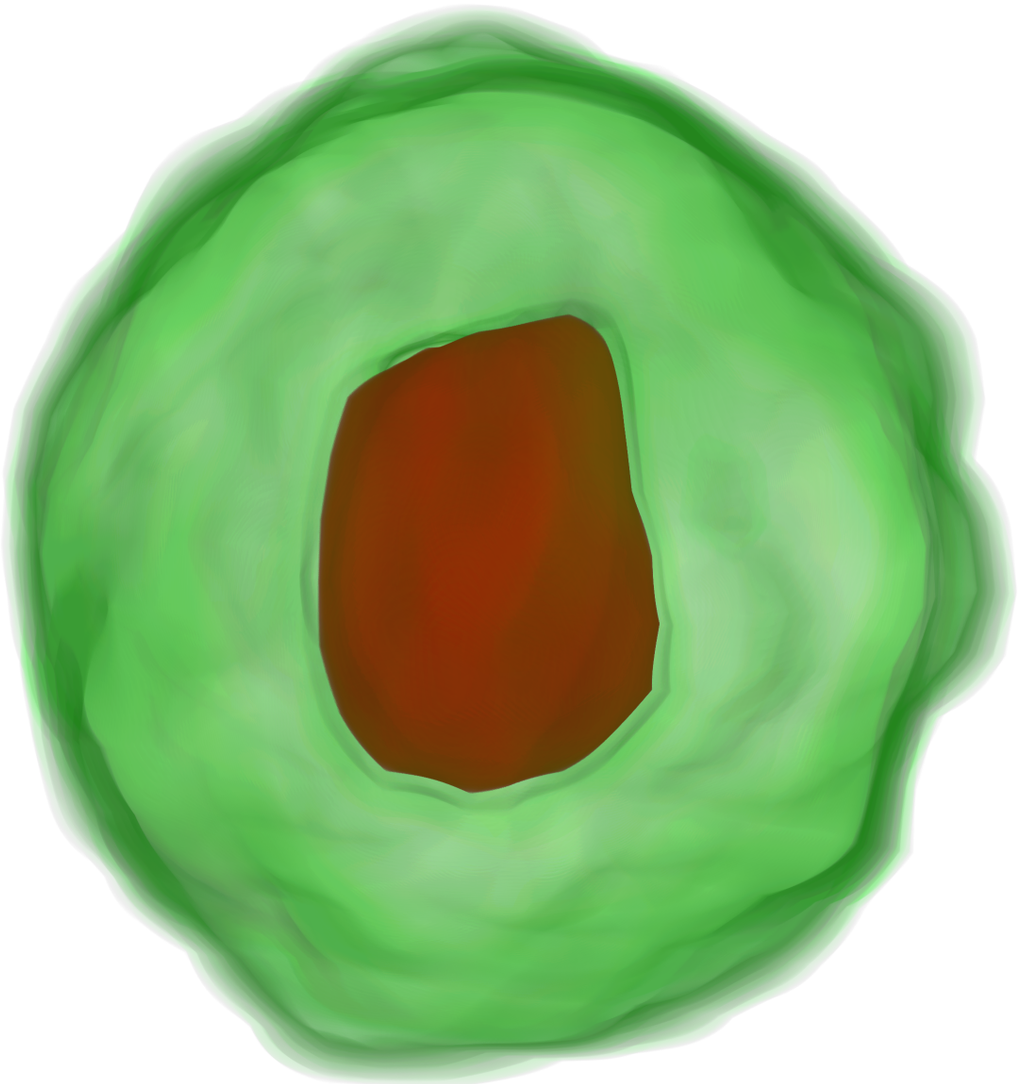
Positive Constrast

Negative Constrast

Nanodisc SAXS data collected in collaboration with M. Shelby (LLNL), Bill Bauer (HWI) at SSRL BL 4-2

DDM Micelle SAXS

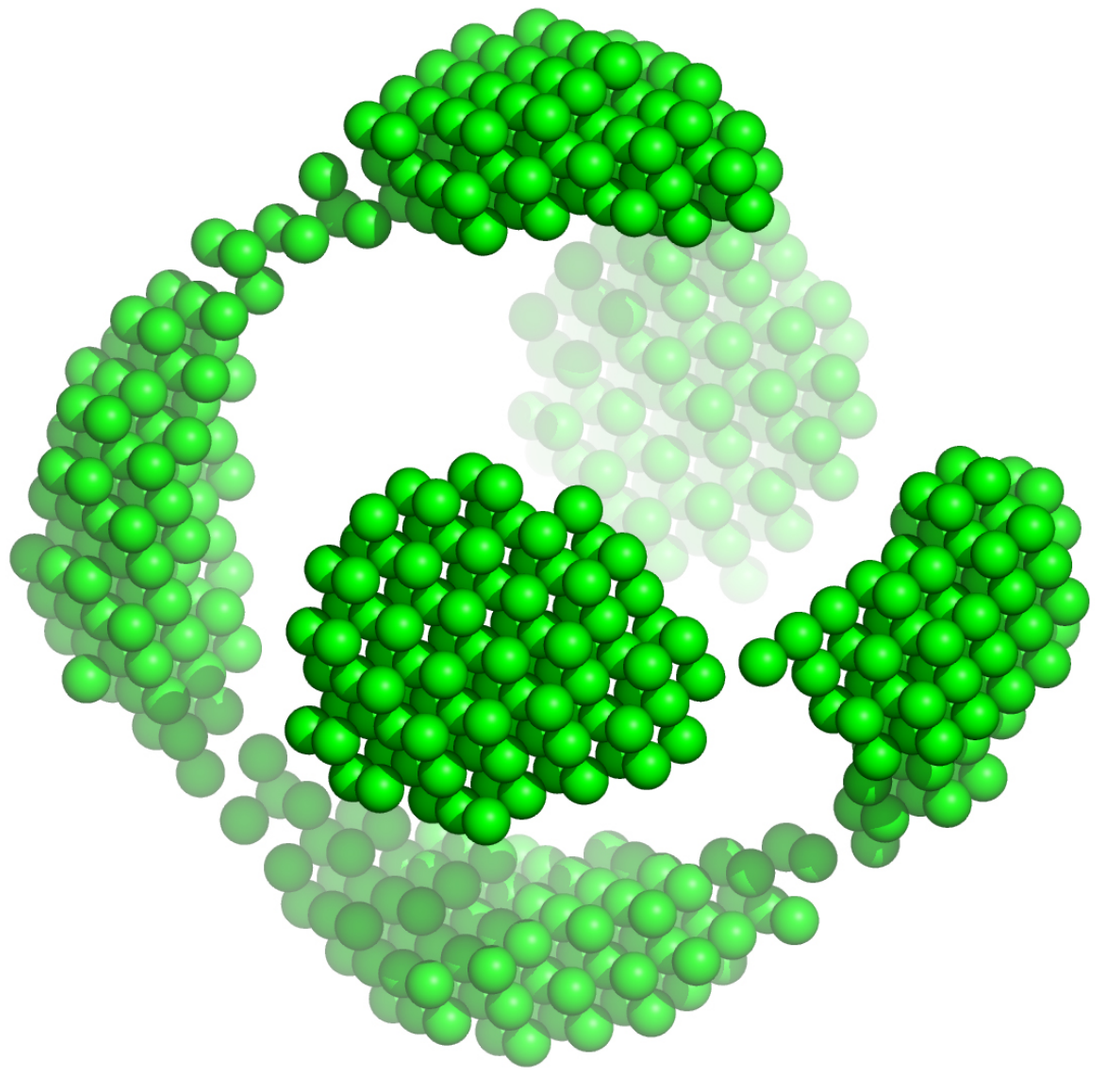
DENSS



Positive Constrast

Negative Constrast

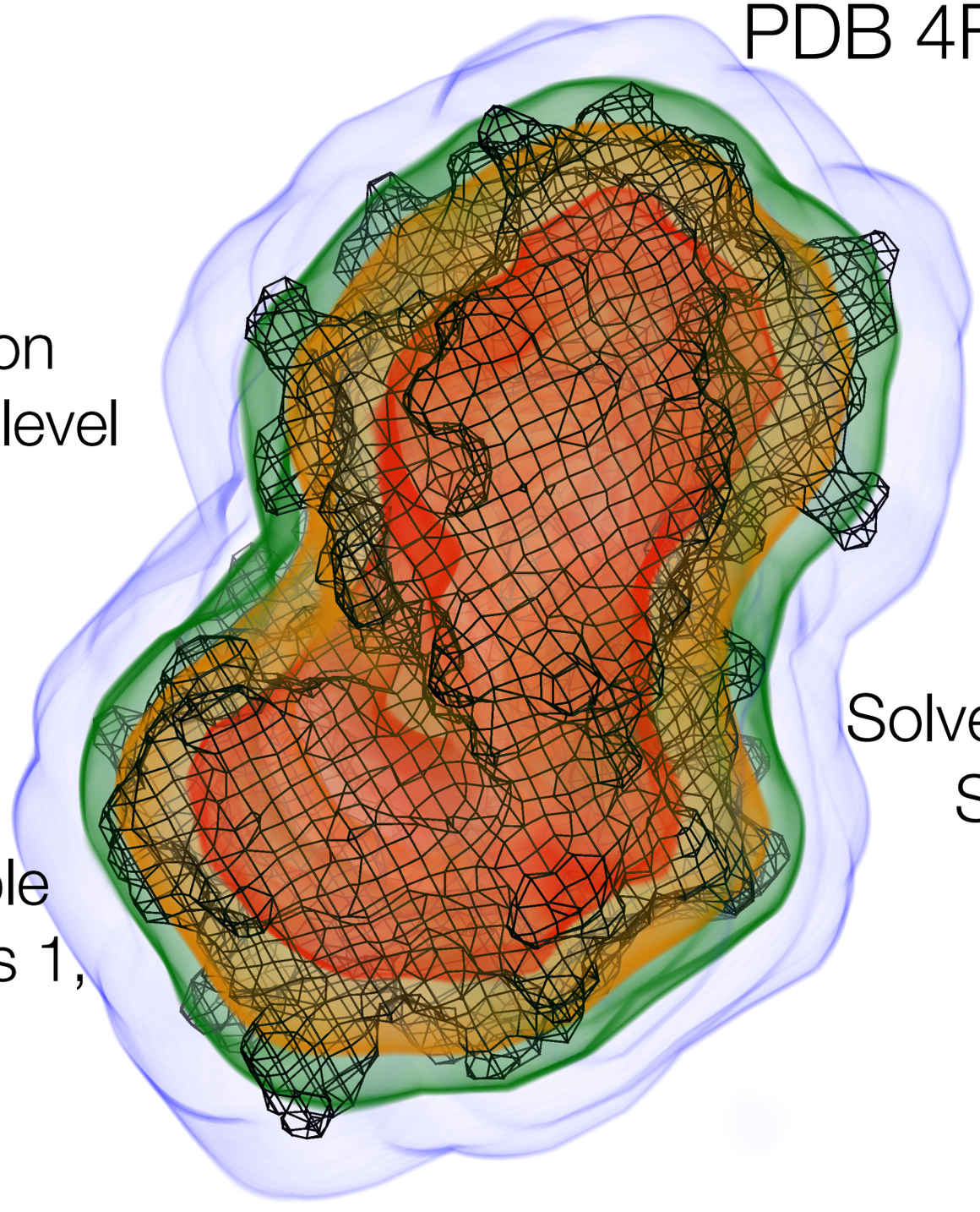
Bead Model



DDM SAXS data courtesy of Frank Gabel.
Gabel, F., et. al. (2019). IUCrJ 6, 521-525.

Choosing isocontours for displaying density

- Bead models have a defined bead radius
- For density, surface representation based on sigma/RMSD contour level
- For default DENSS parameters, volume typically correct at ~8 rmsd
- Often shown as multiple contour levels, such as 1, 5, 10, 20 σ



PDB 4FE9 at 30Å

1 rmsd

5 rmsd

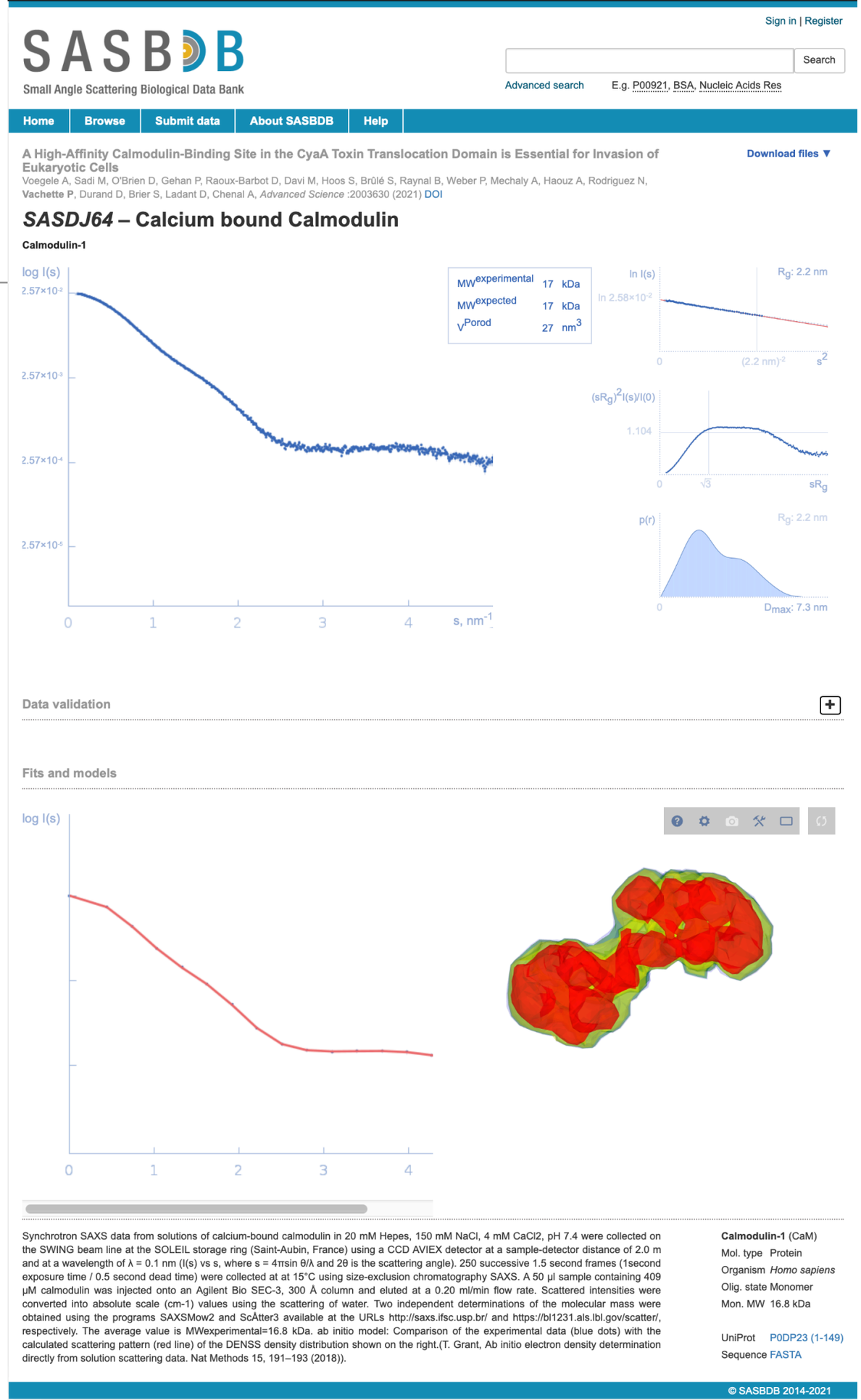
10 rmsd

20 rmsd

Solvent Accessible Surface (mesh)

Archival of 3D Reconstructions

- SASBDB offers deposition of SAS data and models
- Bead models and atomic models are acceptable
- DENSS MRC files are now accepted in the SASBDB, with several depositions already



Acknowledgments

Jacobs School of Medicine and Biomedical Sciences
Department of Structural Biology
University at Buffalo

University at Buffalo
Center for Computational Research

BioXFEL

HAUPTMAN-WOODWARD
MEDICAL RESEARCH INSTITUTE

National Institutes of Health
R01GM133998

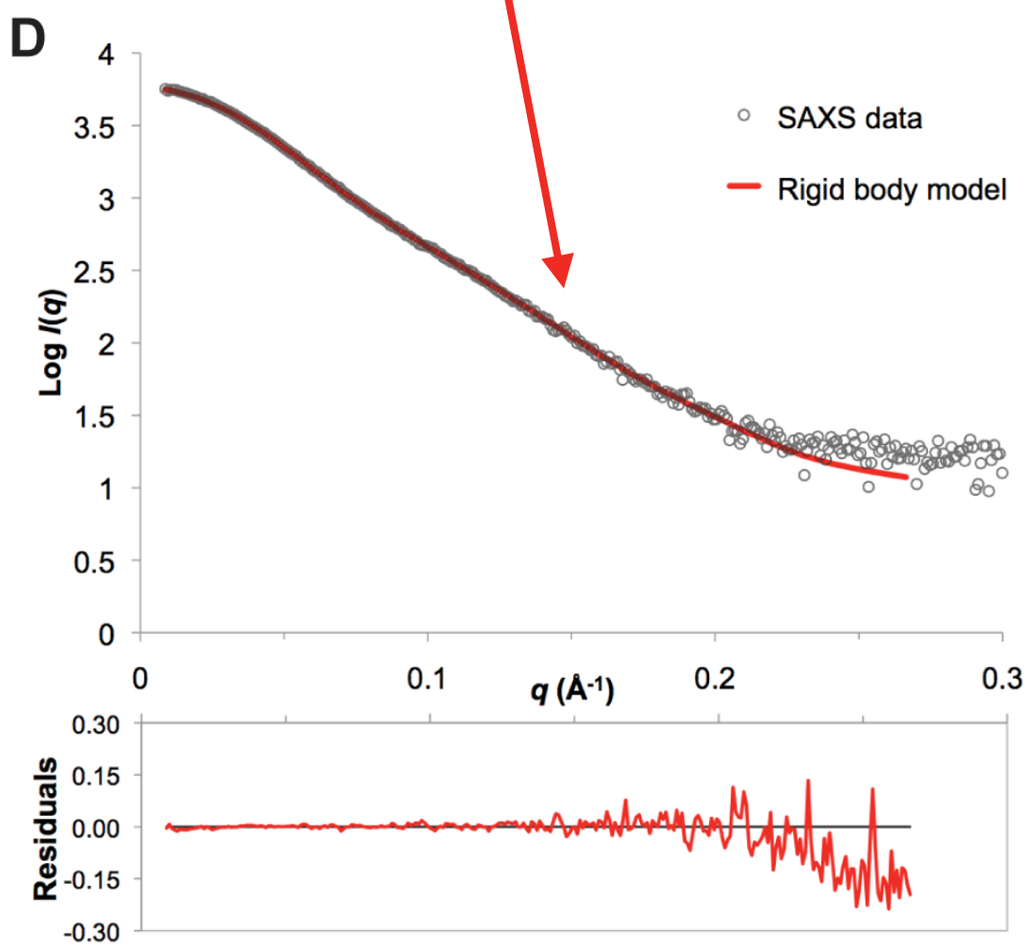
1231306

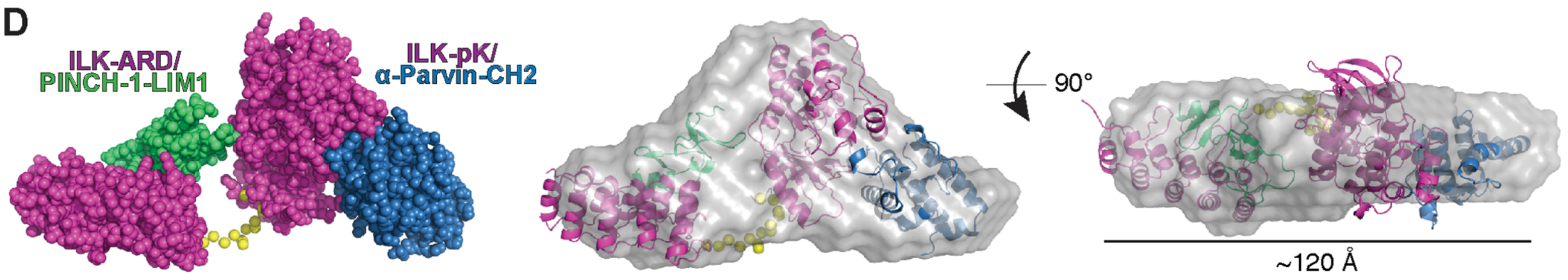
Esther Gallmeier
Nhan Nguyen
Stephen Moore
Richie Singh
Bill Bauer
Andrew Bruno
Joe Chen
Jesse Hopkins
Rick Kirian
Ed Lattman
Megan Shelby
Eddie Snell
John Spence

Rigid Body Modeling

$$\chi^2 = \frac{1}{N}\sum_{i}^{N}\left(\frac{I_{exp}(q_i) - I_{calc}(q_i)}{\sigma_i}\right)^2$$

- Start with known high resolution structures
- Search six-dimensional configuration space (3 translational, 3 rotational) to determine the relative orientation of molecules
- Minimize χ^2 between calculated scattering profile of model and experimental SAXS data
- Resulting 3D model should fit well to *ab initio* envelope determined without knowledge of structures

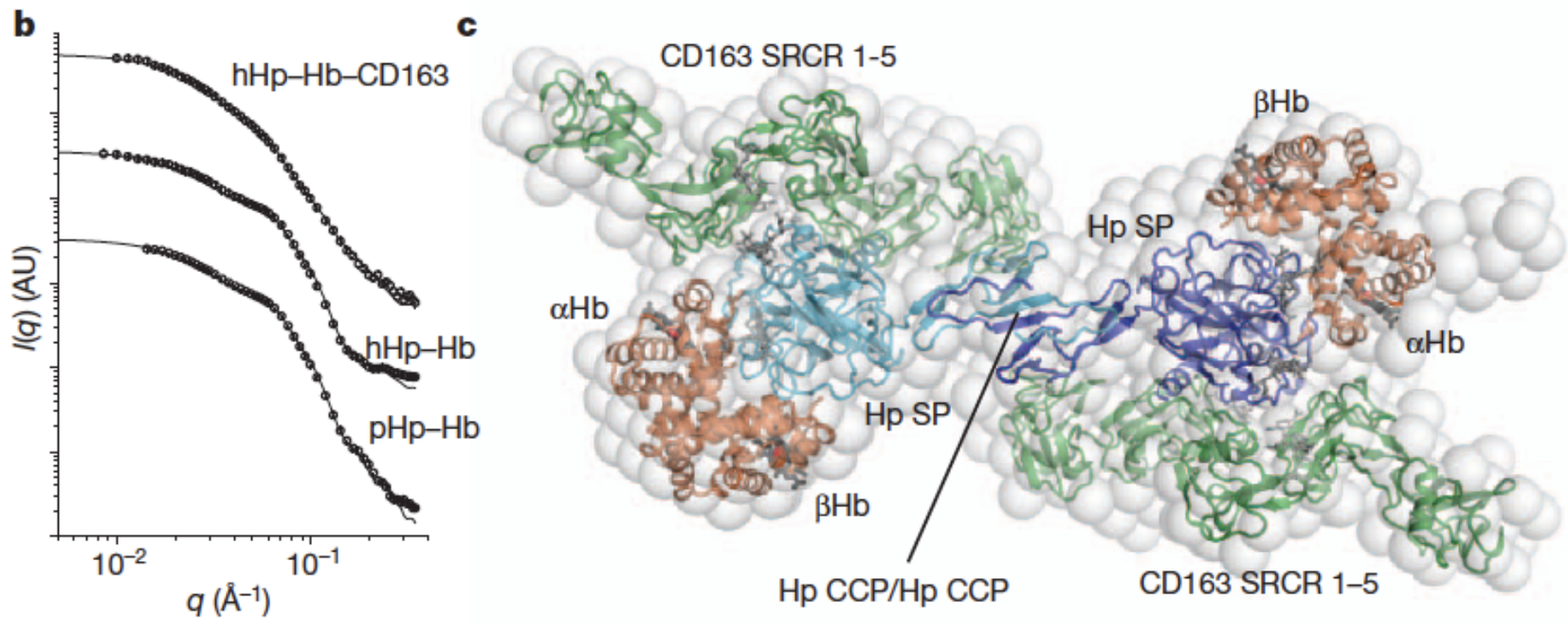




Steigler, et.al. (2013) *PLoS One*

Rigid Body Modeling

- SASREF from ATSAS can search for multiple subunits, domains, complexes
- Can utilize multiple scattering profiles
- Some programs (e.g. BUNCH, CORAL) can fill in missing linkers/regions where structure is unknown
- Can use contact distances as restraints if known *a priori*



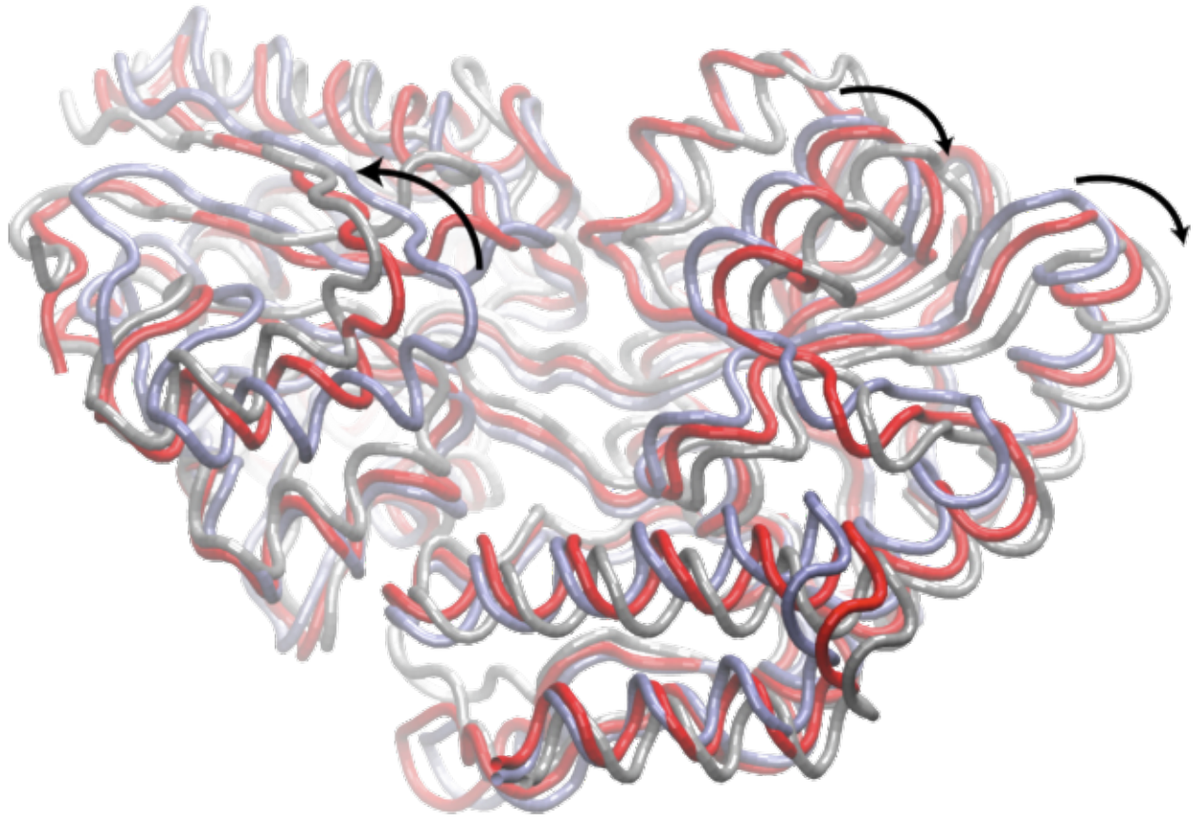
Anderson, et.al. (2012) *Nature*

Flexible Fitting

- Crystal contacts or chemical additives can trap molecules in conformations
- SAXS data can provide low resolution info on molecular conformations in solution
- Modeling based on rigid bodies that partition models into rigid domains may be unable to capture flexible motions
- Flexible fitting attempts to determine new conformations from existing high resolution models that fit SAXS data
- Often done using low-resolution cryo-EM envelopes, now able to be done with SAXS data, e.g. using DENSS maps directly, or bead models can be converted to map format

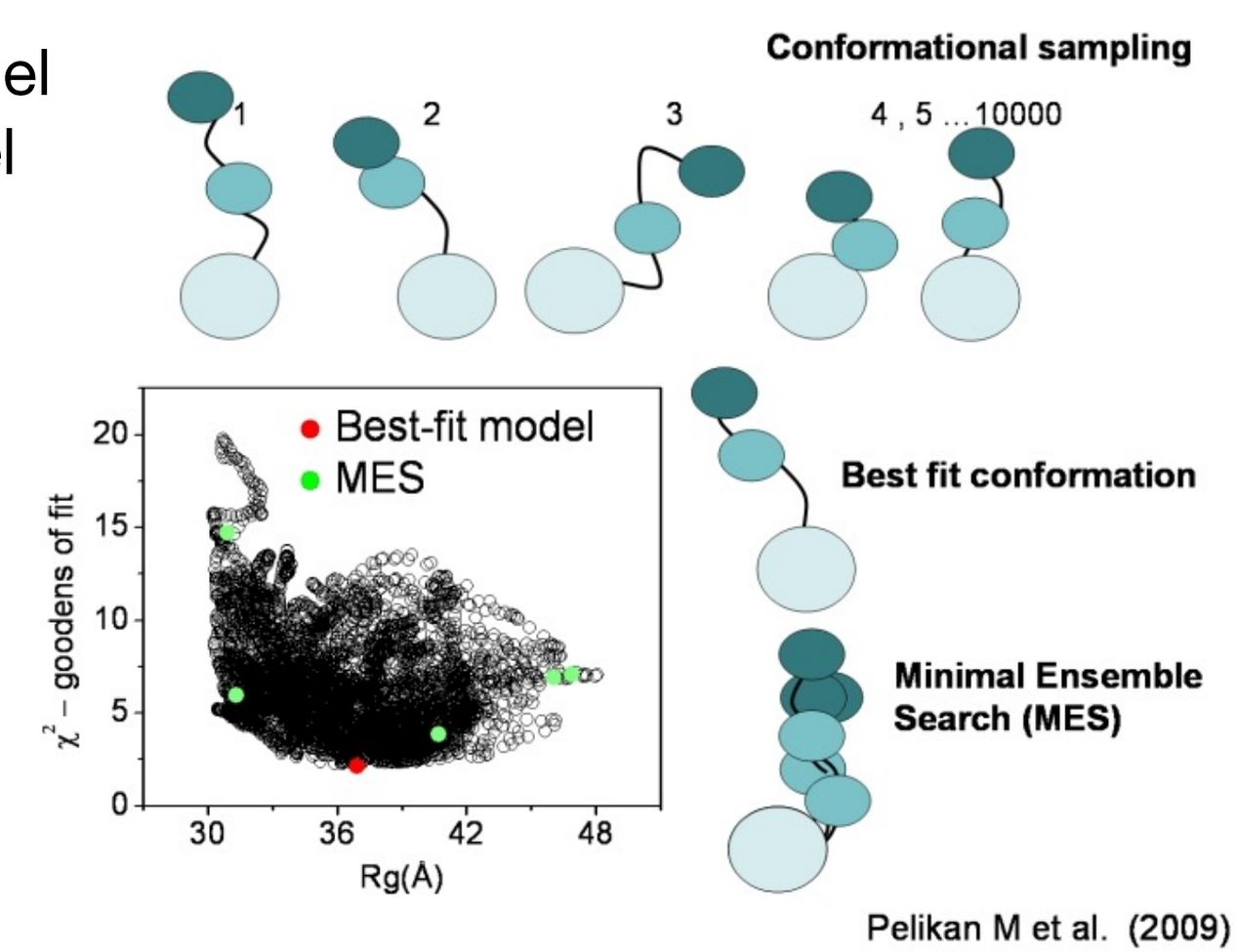
Flexible Fitting

- Normal mode analysis a common way to generate conformations of macromolecules that are biologically realistic
- Fit conformations to raw scattering profile, pair distribution function, or 3D molecular envelope (similar to cryo-EM)
- Two main methods:
  - sample many conformations and filter using SAXS data
  - drive conformations directly by SAXS data
- Steered molecular dynamics also possible



Ensemble Modeling

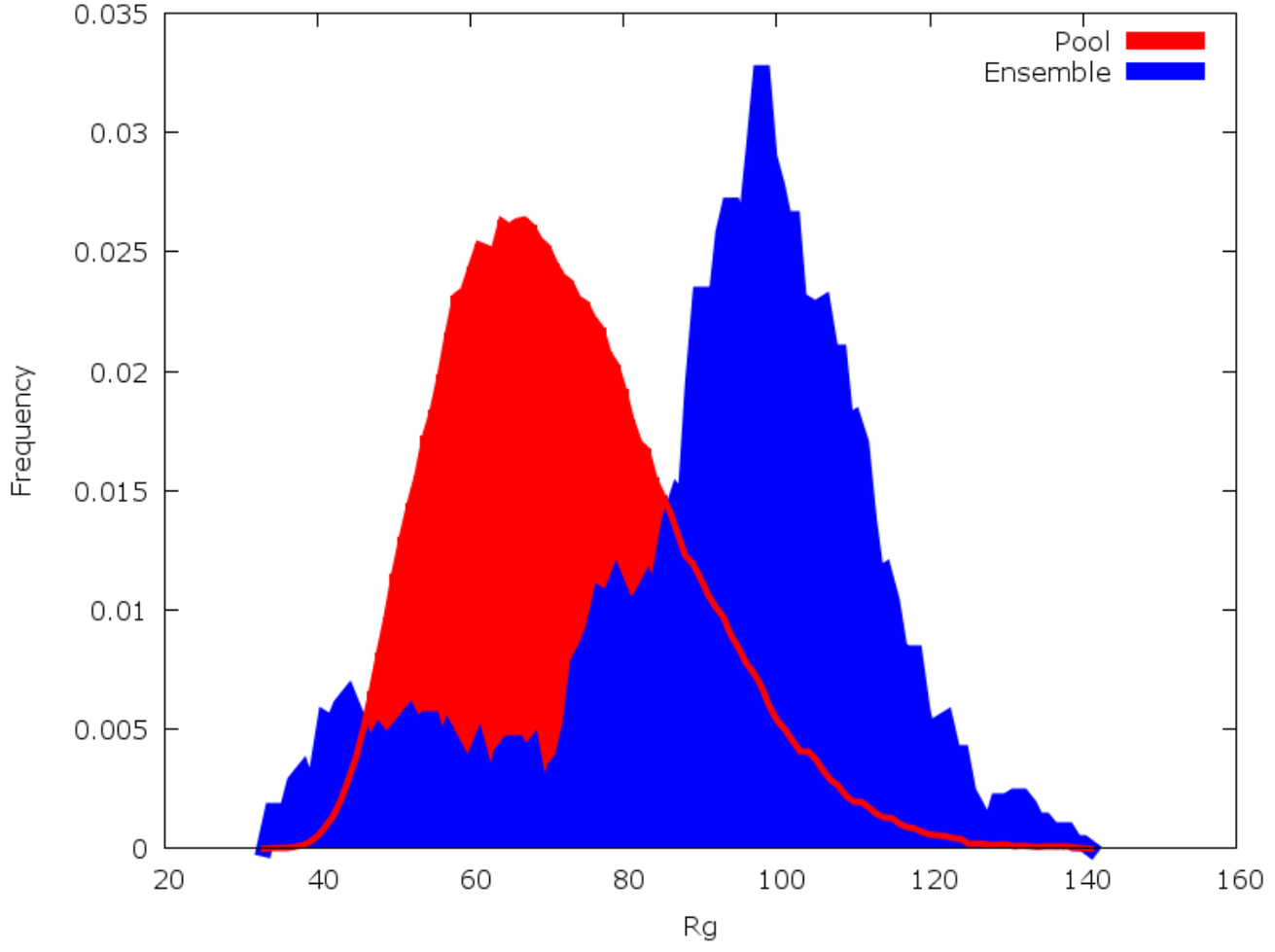
- Single rigid models may not adequately represent the average scattering of all molecules in the illuminated solution volume
- Protein dynamics resulting in an equilibrium of various conformational states contribute to scattering profile
- Ensemble modeling attempts to model SAXS data with more than one model by averaging SAXS patterns of all models in ensemble to match experimental data
- SAXS one of the few and most powerful methods to describe large scale conformational dynamics



Pelikan M et al. (2009)

How to Evaluate Ensemble Modeling Results

- Do not believe structural models from final ensemble are real, they are used as a tool only to describe the gross structural properties of the ensemble
- Compare Rg distribution of ensemble to Rg distribution of random pool of thousands of conformers
  - ensemble distribution as wide or wider than random pool suggests large degree of flexibility
  - peak position describes extended or compact ensemble
  - multiple peaks suggests distinct conformations

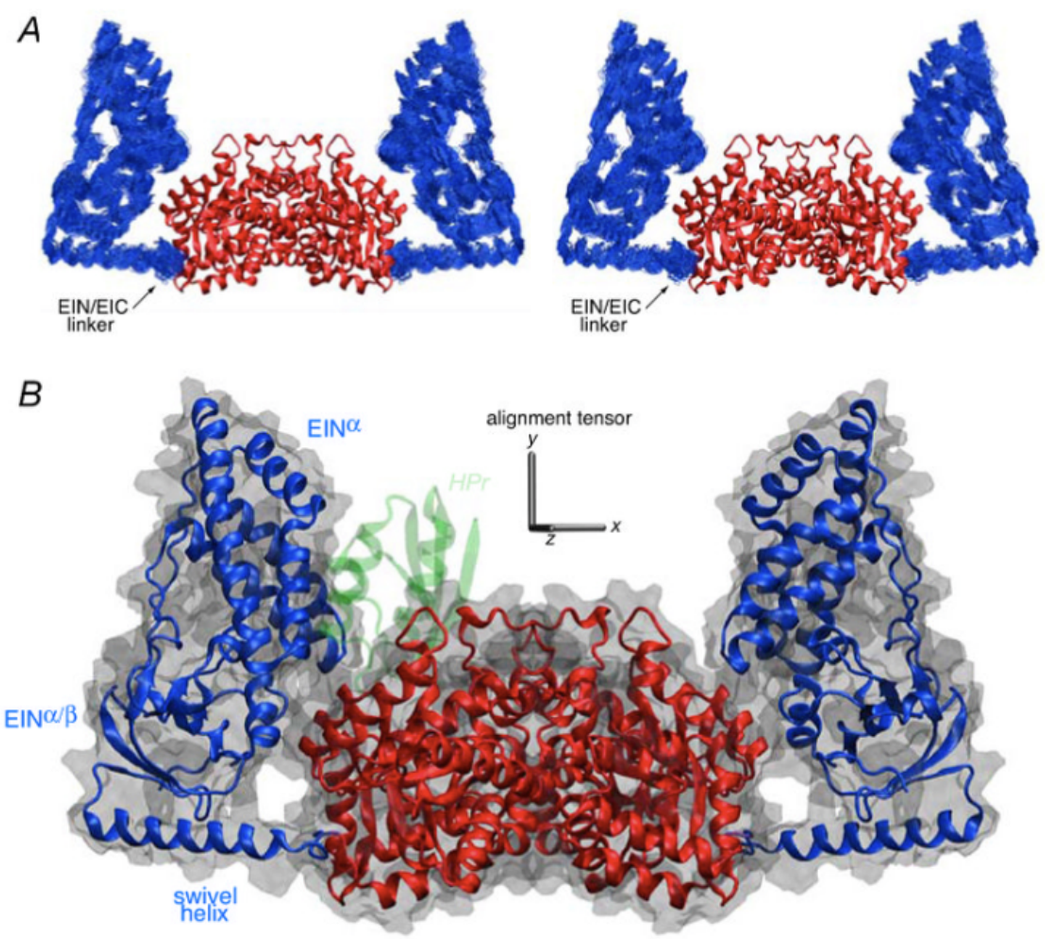


Hybrid Modeling

- SAXS is most commonly and most powerfully used as a complementary component to other structural data

- NMR data describes short range, high-resolution distances, but cannot access long-range information

- Hybrid methods involve combining several different types of experimental data for joint structural refinement, including: SAXS, SANS, WAXS, NMR spectra, residual dipolar couplings (RDCs), X-ray structures, etc.



Schweiters, et.al. (2010) *JACS*