Introduction of Analytical Software: What You Can Do Using BioSAXS?

Tsutomu Matsui
SSRL/Dept. of Chemistry
Stanford University

SLAC NATIONAL ACCELERATOR LABORATORY

SSRL

Structural Molecular Biology

Software suite for BioSAXS

ATSAS: A program suite for small-angle scattering data analysis from biological macromolecules., developed by EMBL Hamburg, Dr. Dmitri Svergun group.

FoXS and IMP: Computation and modeling programs developed by UCSF, Dr. Andrej Sali group.

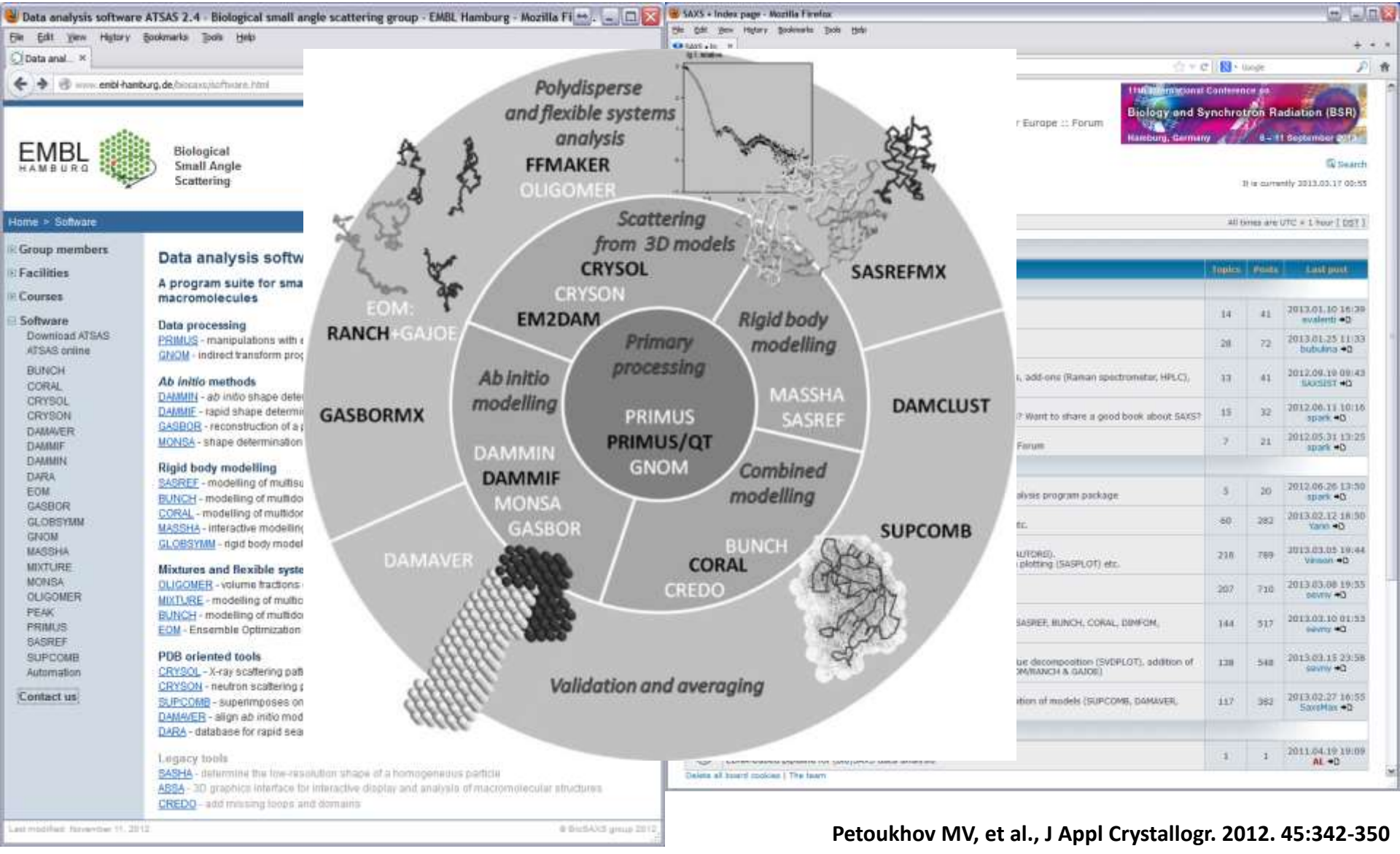
SAXSTB: SAXS tool box developed by LBL, Dr. Peter Zwart group.

Martin: Coarse Grain Force Field for Biomolecular Simulations developed by University of Groningen, Dr. Siewert-jan Marrink group.

and so on.

ATSAS and SaxieR

A program suite for SAXS data analysis from biological macromolecules developed by Dmiytri Svergun Group at EMBL



Petoukhov MV, et al., J Appl Crystallogr. 2012. 45:342-350

Programs at UCSF

- IMP, the Integrative Modeling Platform
  - ➔Comprehensive structural analysis by hybrid method approach.
- FoXS, Fast SAXS Profile Computation with Debye Formula.
- FoXS Dock, Macromolecular Docking with SAXS Profile.
- BILBOMD (LBL)
- SAXS Merge

…

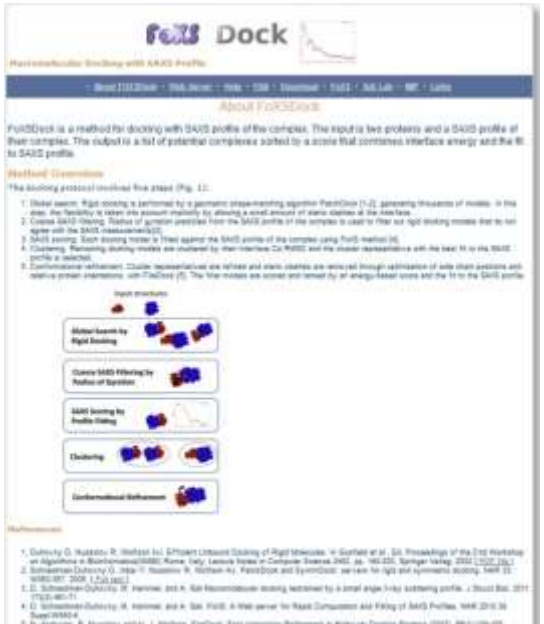
- UCSF Chimera, Small-Angle X-Ray Profile tool
  Modeller interface

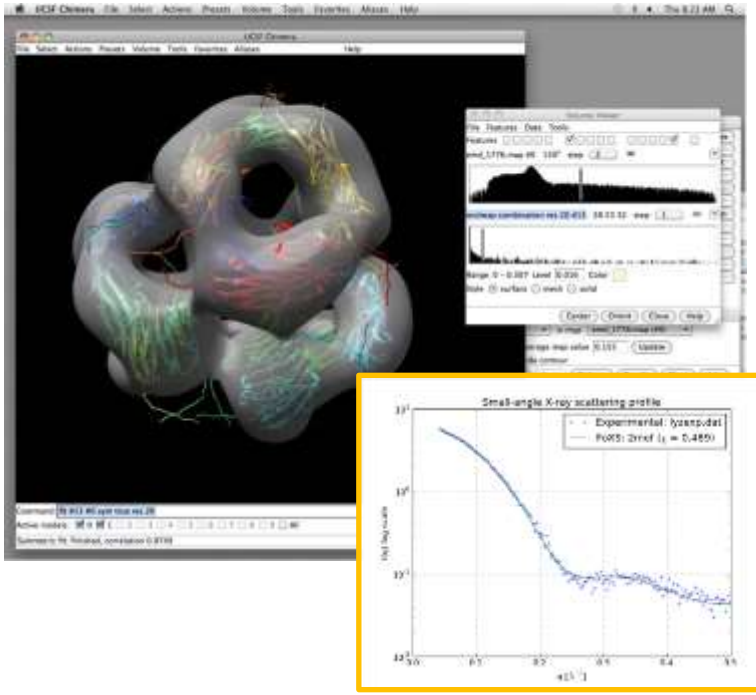
Prof. Andrej Sali

Dr. Dina Schneidman
(now Hebrew University)

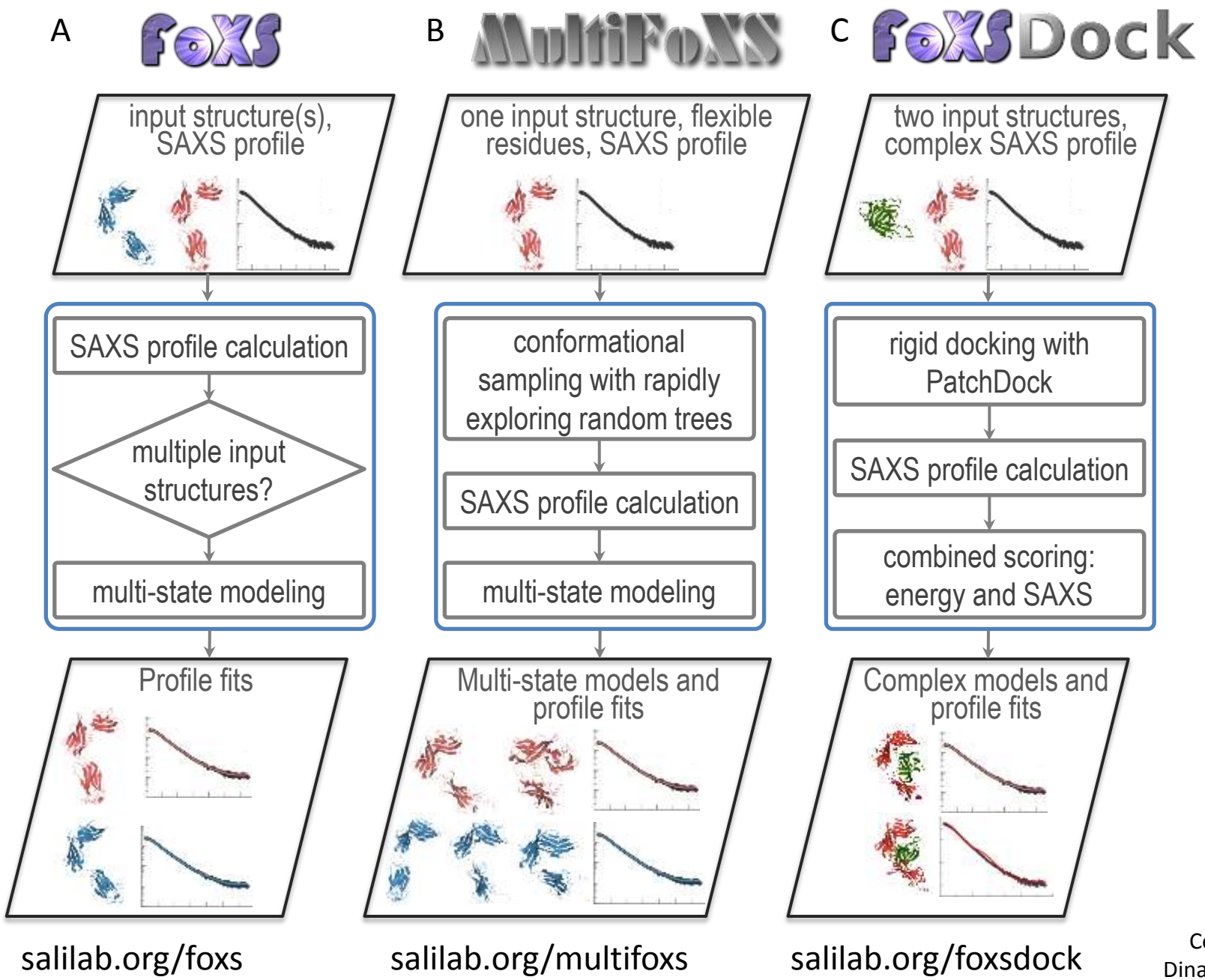
Dr. Seung Joong Kim

Dr. Daniel Saltzberg





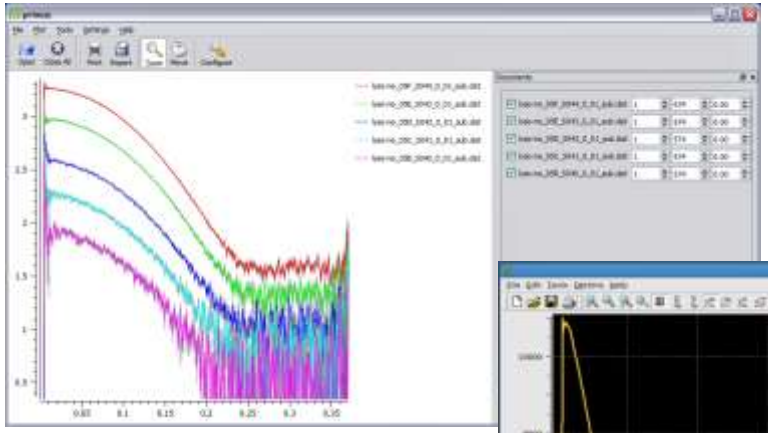
Sali Lab SAXS servers (UCSF)



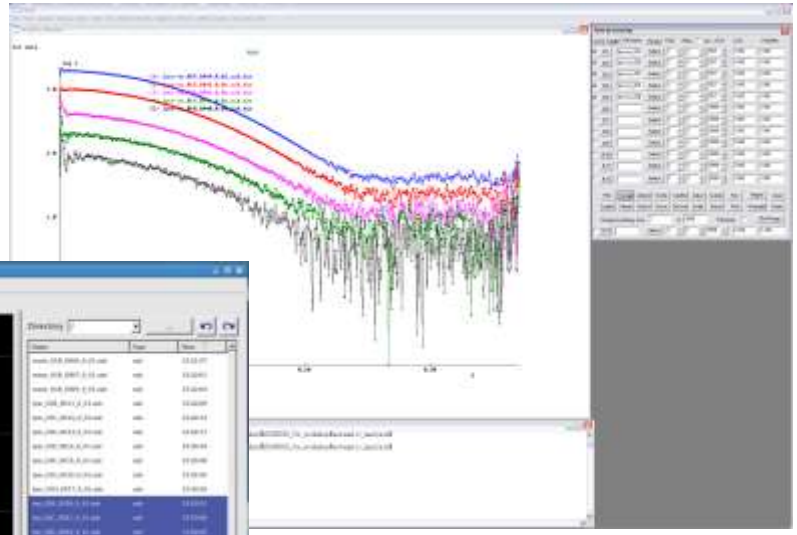
Courtesy of Dr. Dina Schneidman

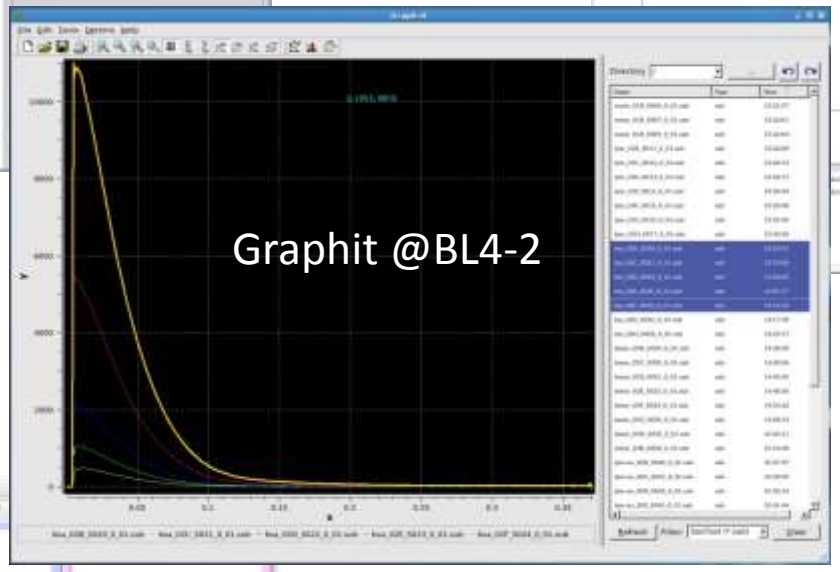
Data viewer for initial analysis/inspection

Primusqt (ATSAS)



Primus (ATSAS)





ScÅtter

http://www.bioisis.net

- MS Excel
- Other math software

Typical data format (txt file):

1st column: q

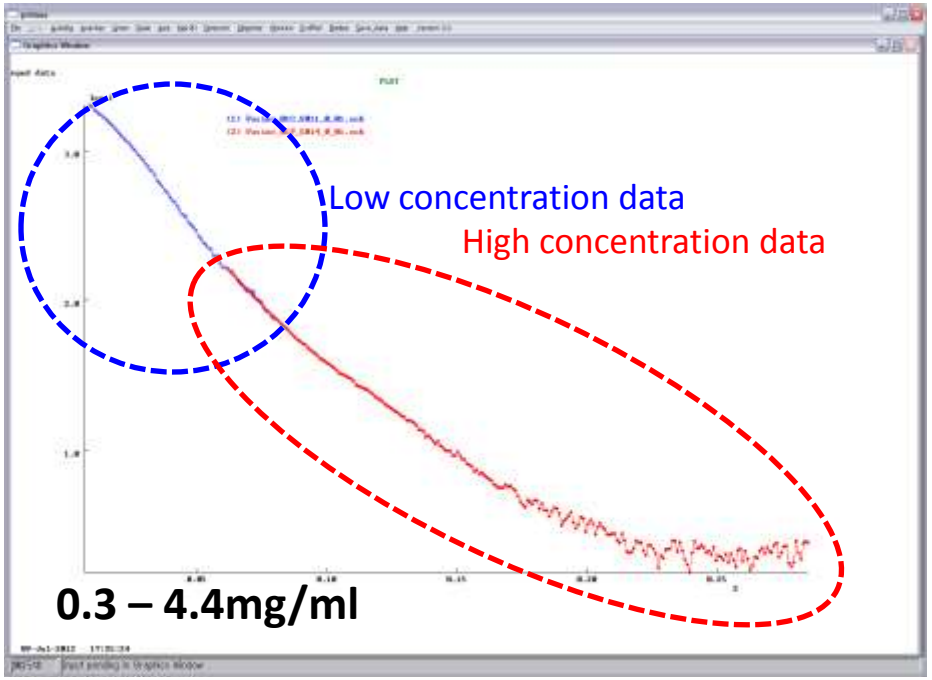
2nd column: Intensity

3rd column: sigma of intensity

Merging

Zero concentration extrapolation from a concentration series

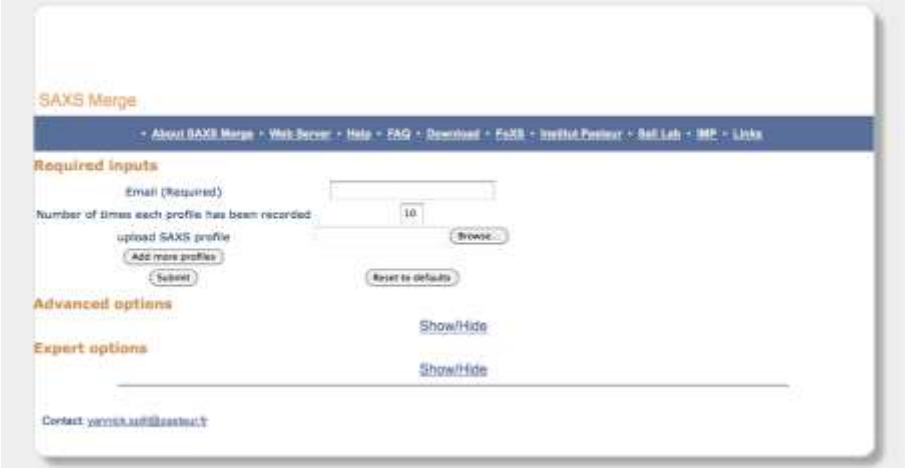
PRIMUS



Need to measure low concentration sample as low as possible.

Confirm if concentration dependence is indistinguishable at low conc.

SAXS Merge



Almerge

- merge two SAS data sets together or to extrapolate several SAS data sets to infinite dilution (zero concentration)

Mw estimation

Standard curve

$Mw_{exp} / I(0)_{exp} \approx Mw_{standard}/I(0)_{standard}$

Porod volume*

Mw ≈ Porod volume *1.2/2

SAXS MoW*

JAVA applet (http://www.if.sc.usp.br/~saxs/)

Water scattering (available in Graphit)

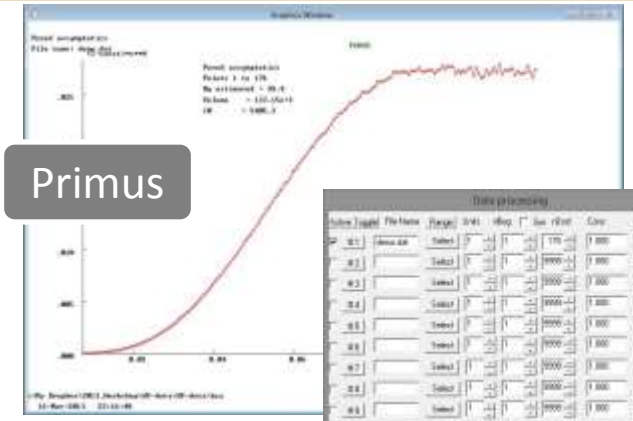
Orthaber *et al*., J Appl Cryst (2000) 33, 218

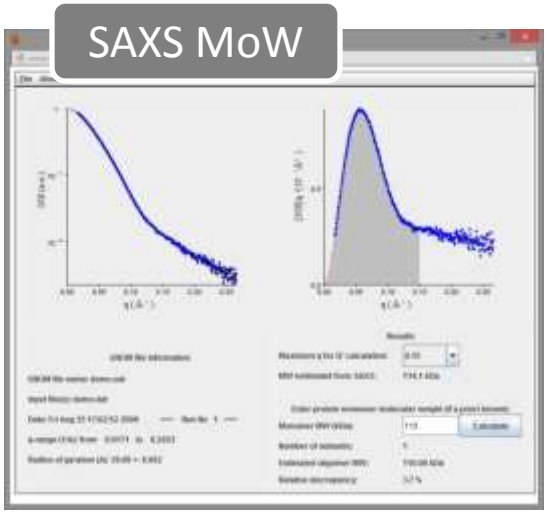
DATVC/DATMOW (ATSAS)*

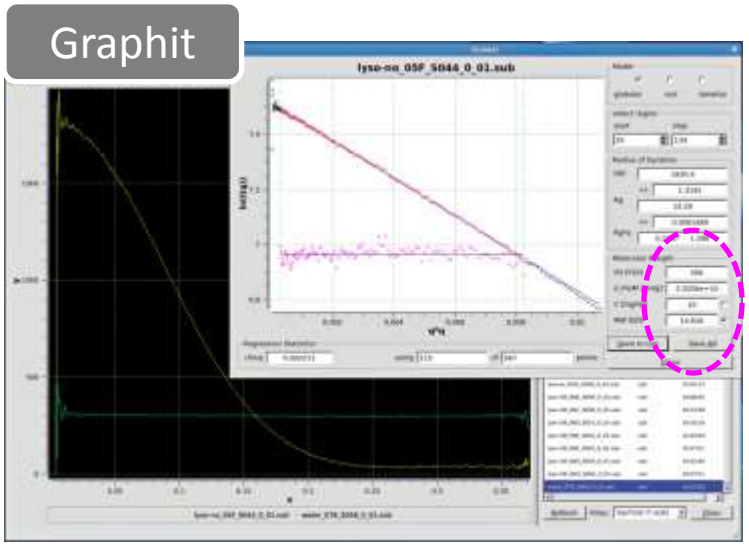
Volume of dummy atoms*

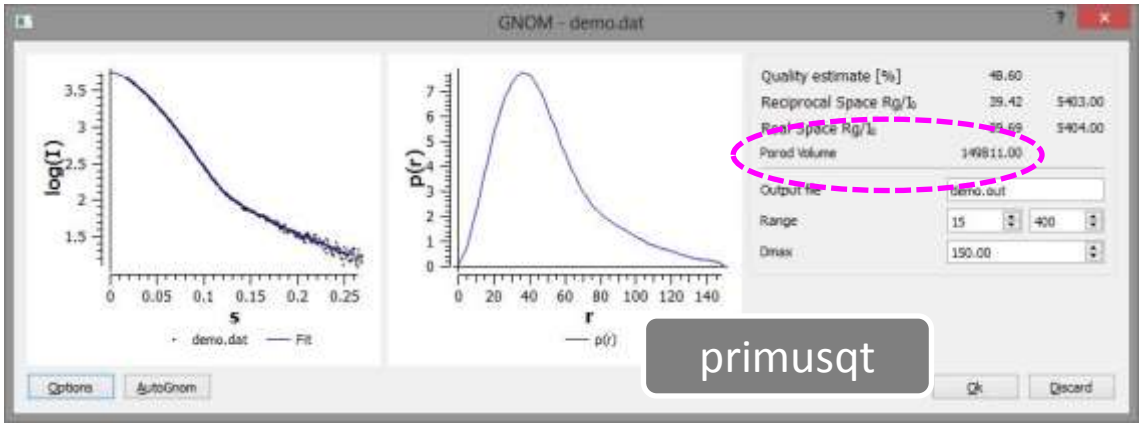
Mw ≈ (Dammif or Dammin volume)/2

*Concentration (scale) independent





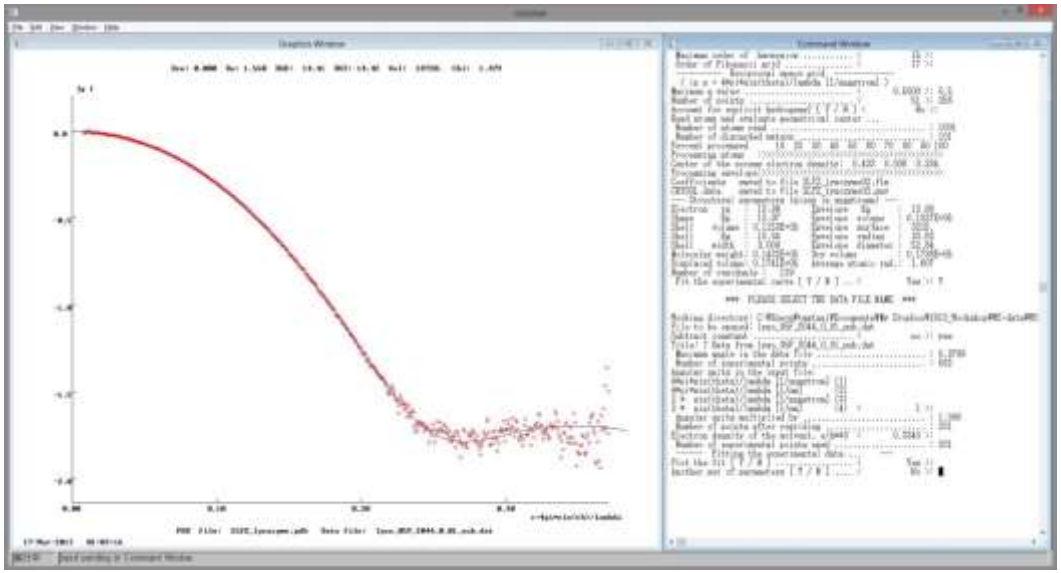




Fitting with known structure

Input file: Scattering data
PDB file

CRYSOL



A variety of programs are available.

FoXS

AquaSAXS

Fast-SAXS

HyPred

e.t.c.

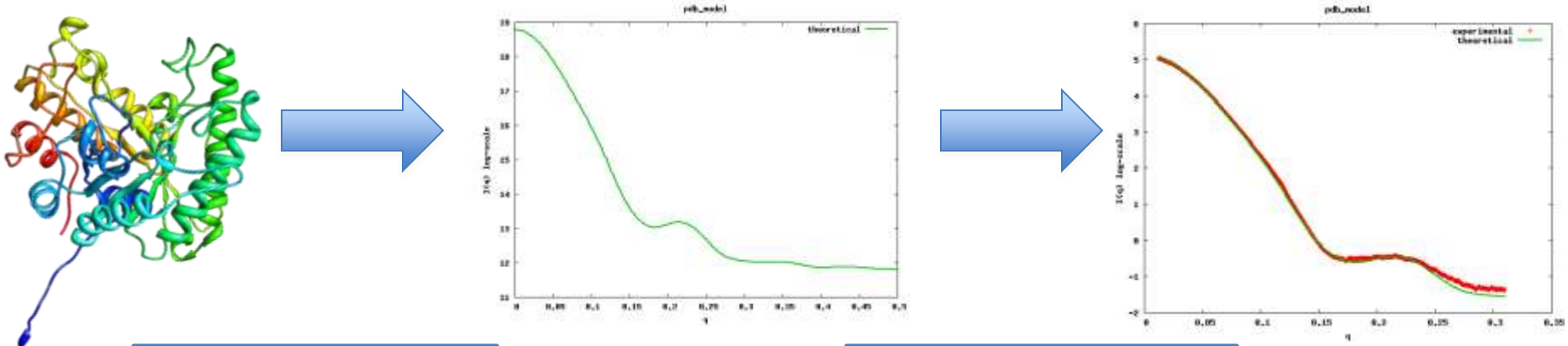
Curve fitting is evaluated by χ or χ^2 value:

$$\chi^2 = \frac{1}{N-1} \sum_{k=1}^{N} \left[\frac{I_{exp}(q_k) - c\, I_{model}(q_k)}{\sigma_{exp}(q_k)} \right]^2$$

Fast open-source X-ray Scattering

FoXS

A rapid method for computing a SAXS profile of a given structure and for matching of the computed and experimental profiles



Compute theoretical profile

Fit experimental profile

Debye formula (1915)

$$I_m(q) = \sum_i \sum_j f_i(q) f_j(q) \frac{\sin(q d_{ij})}{q d_{ij}}$$

$$c^2 = \frac{1}{N-1} \sum_i \left[\frac{I_{\exp}(q_i) - c I_m(q_i)}{s(q_i)} \right]^2$$

Schneidman-Duhovny D, Hammel M, Sali A. NAR 2010

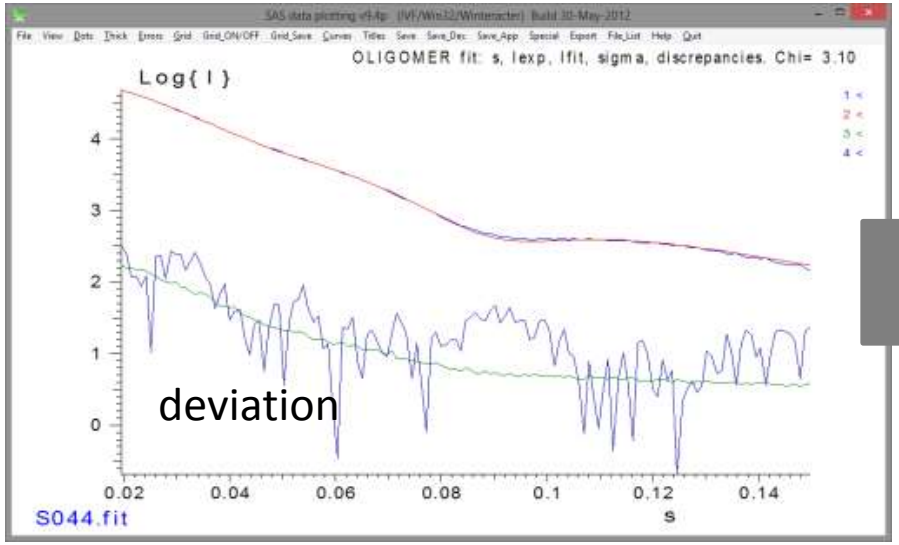
Schneidman-Duhovny D, Hammel M, Tainer J, Sali A. Biophys J 2013

Courtesy of Dr. Dina Schneidman

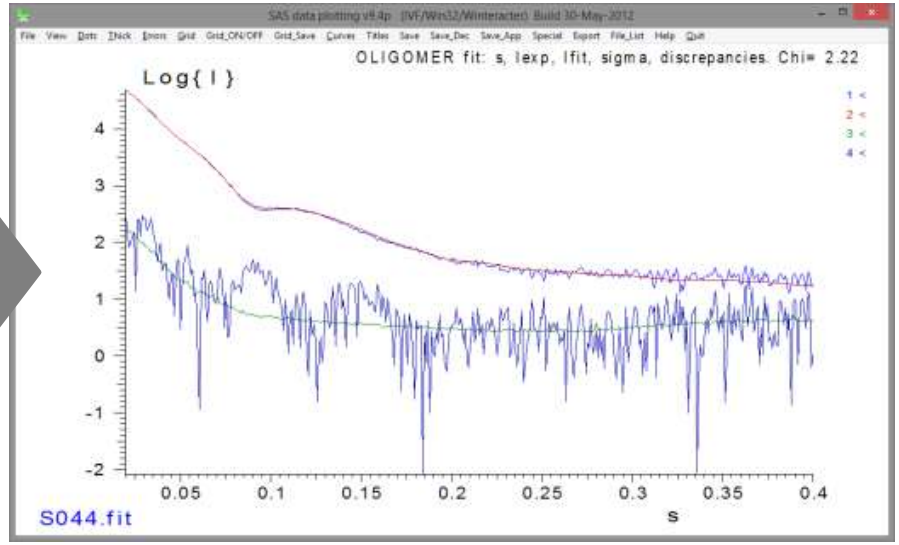
χ and χ^2

$$\chi^2 = \frac{1}{N-1} \sum_{k=1}^{N} \left[\frac{I_{exp}(q_k) - c\, I_{model}(q_k)}{\sigma_{exp}(q_k)} \right]^2$$

q_{max}=0.15



q_{max}=0.4



- Can not compare χ^2 or χ^2 values between different data sets.
- Need to inspect fitting by your eye.

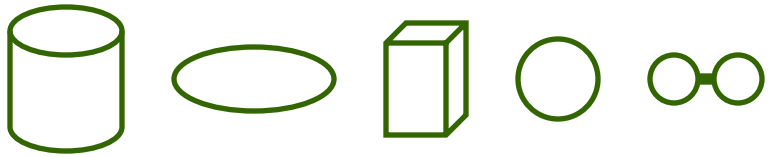
Fitting with simple geometric model

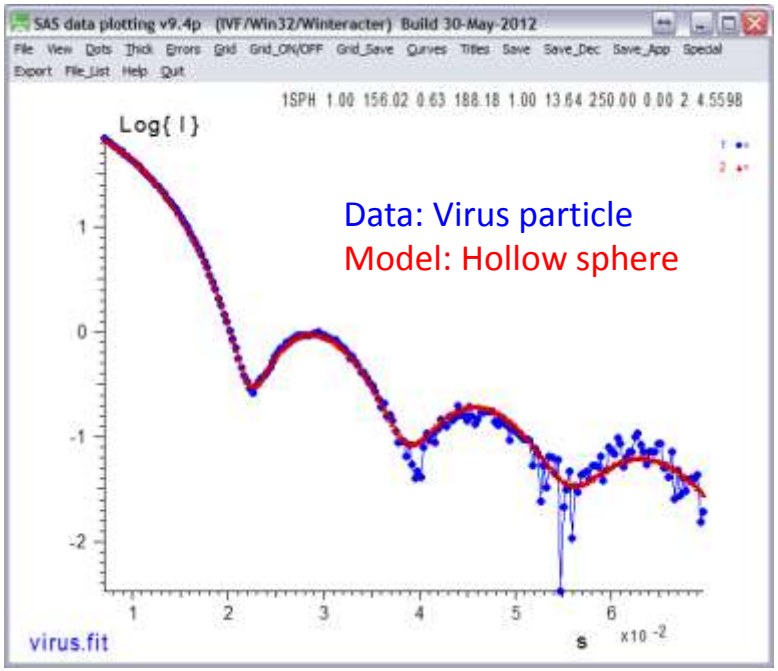
Input file: Scattering data

BODIES

Approximation using:

1. ellipsoid (semiaxes a,b,c)
2. ellipsoid of revolution(semiaxes a,a,c)
3. cylinder
4. elliptic cylinder
5. hollow cylinder
6. rectangular prism





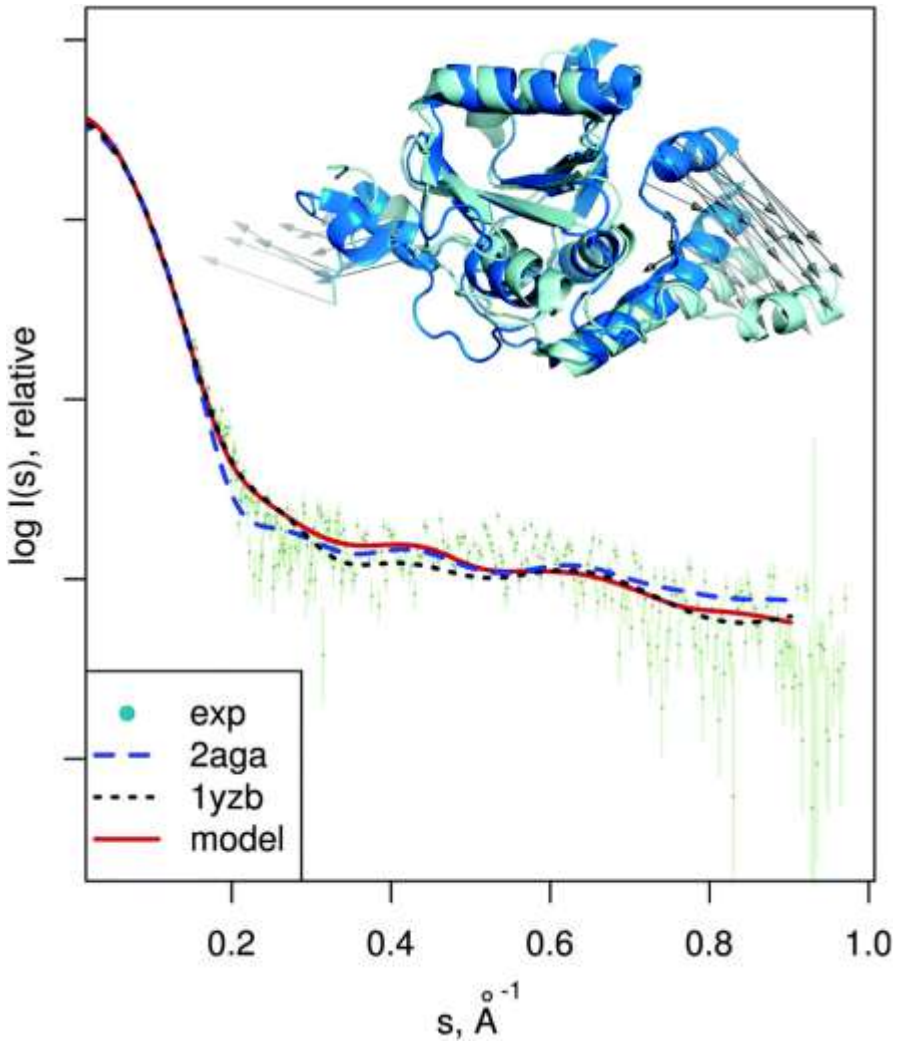
MIXTURE

Single or multicomponent system of:

1. Spheres
2. Cylinders
3. Ellipsoids of rotation
4. Dumbbells

Deciphering conformational transition by normal mode analysis

SREFLEX



Input file: Scattering data
PDB coordinate

Panjkovich A, Svergun DI. Phys Chem Chem Phys. 2016 Feb 17;18(8):5707-19.

Rigid Body Refinement

Input file: Scattering data
PDB files or .alm files
Restraint information (option)
q_{max} limit: 0.25-0.3

SASREF — Global rigid body modeling using known structures.

BUNCH — Rigid body modeling with addition of missing fragments using ab intio modeling approach.

CORAL — Rigid body modeling with addition of missing fragments form random loop library.

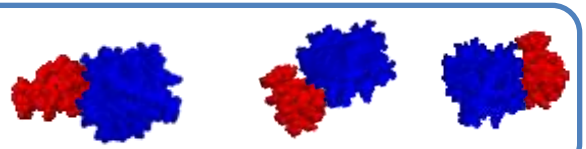
Missing fragments: Disordered region in crystal structure
Expression tag (like His-tag)
Extra residues from expression vector

FoXSDock protein-protein docking with complex SAXS profile

Input structures

Rigid Docking

PATCHDOCK

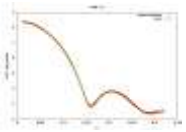


~10^{11} models

~10^{6} models

SAXS-guided Scoring

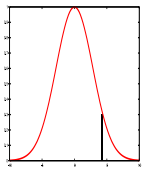
FoXS



SOAP
statistical potential for protein interfaces

~10^{3} models

Composite Scoring



10 models

Schneidman-Duhovny D, Hammel M, Sali A. JSB 2011

Schneidman-Duhovny D et al. Bioinformatics 2012 salilab.org/foxsdock

Courtesy of Dr. Dina Schneidman

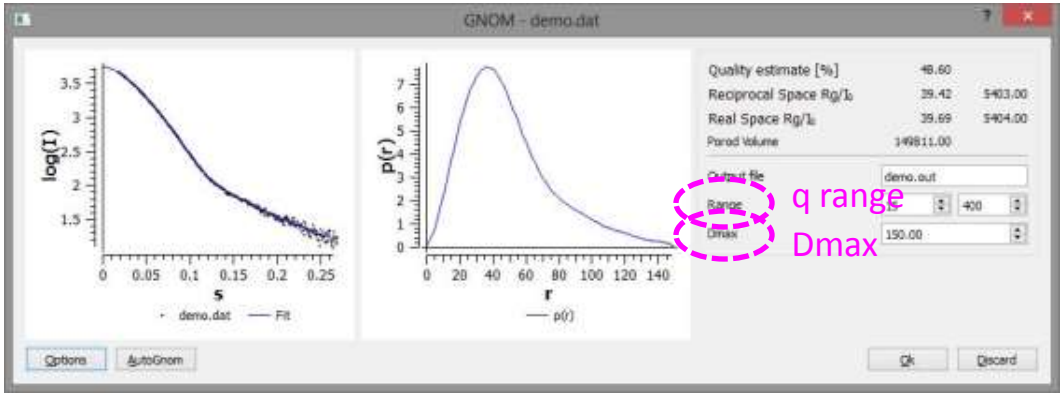
GNOM: Indirect Fourier transform

Indirect Fourier transform (FT) is a solution of ill-posed given by FT of extremely noisy data proposed by Glatter (1977).

Scattering profile (Reciprocal space) ⟶ Distance distribution function (Real space)

Gnom output file is required for some of programs.

GNOM interface in primusqt



GNOM only works for the following system:

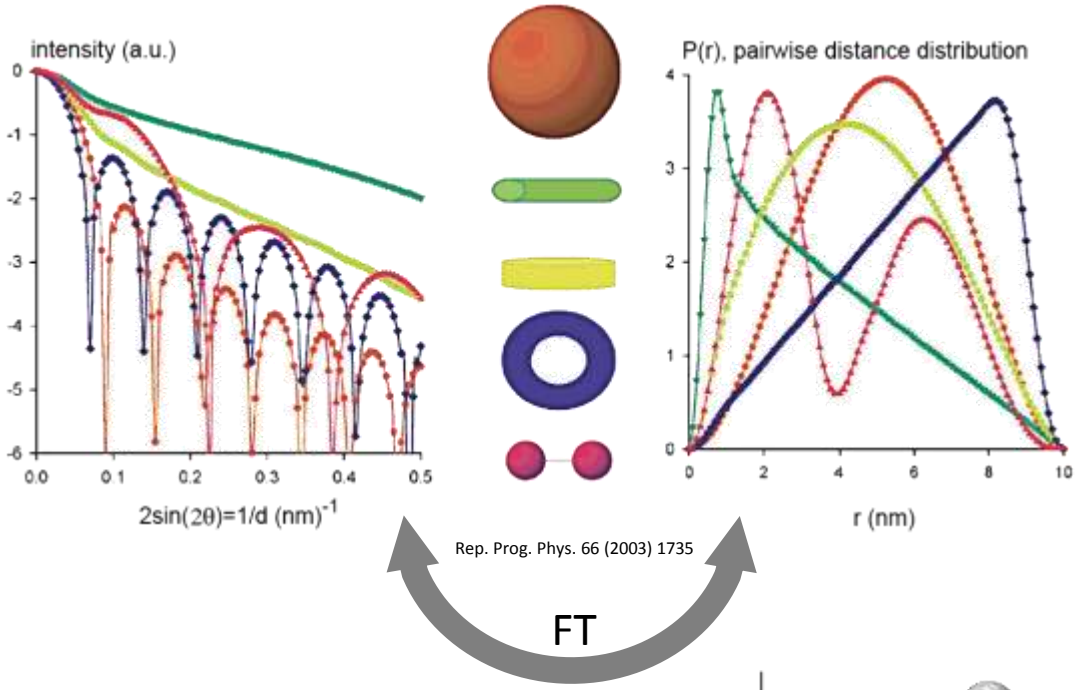
- Monodisperse systems of globular particles;
- Monodisperse systems of elongated particles;
- Monodisperse systems of flattened particles;
- Polydisperse systems of spherical particles;
-Polydisperse systems of particles with an arbitrary form factor.

No standardized way to estimate Dmax.

More detail will be provided in hands-on tutorials. You will play around with P(r).

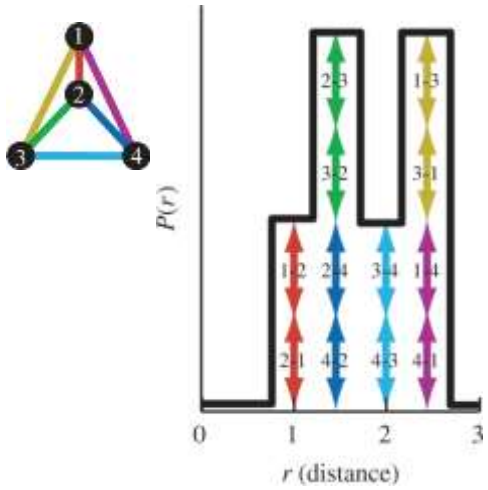
SAXS profile and P(r)*

*Otherwise known as: PDDF (Pair-wise distance distribution function)
Distribution of inter-atomic distance

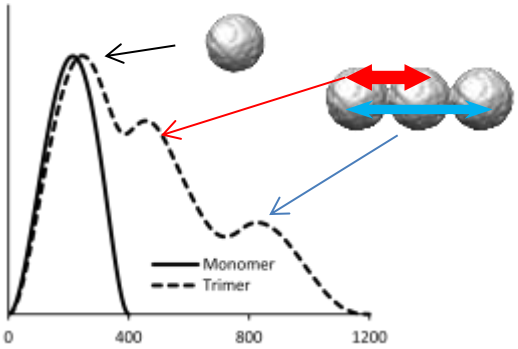


Rep. Prog. Phys. 66 (2003) 1735

P(r) is the SAXS analog of crystallographic Patterson function.



Putnam, et al. 2007 Quarterly Reviews of Biophysics 40, 191



ab initio shape determination (single phase): SAXS envelope

Input file: GNOM output file

q_{max} limit: qRg<7~8 (rule of thumb for DAMMIN & DAMMIF)

No limit for GASBOR

Rg=15Å → q_{max} limit = 8/15 =~0.53
Rg=35Å → q_{max} limit = 8/50 =~0.22
Rg=180Å → q_{max} limit = 8/100 =~0.04

DAMMIN/DAMMIF Low-resolution bead-modeling.

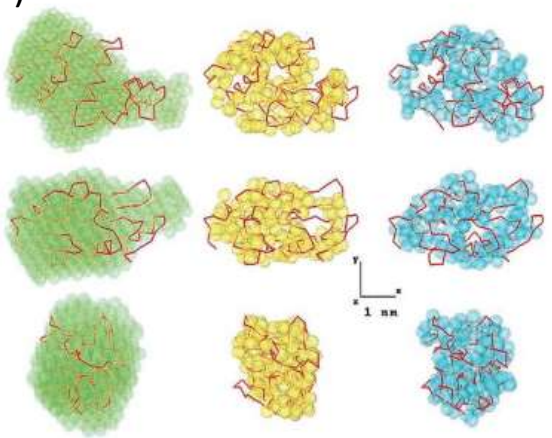
GASBOR High resolution bead-modeling. Ensemble of dummy residues forming a chain-compatible model.

Need to take account of chirality of the model.

Dali_GA The 1st bead model program. (genetic algorithm)

Simulated annealing is employed during modeling.

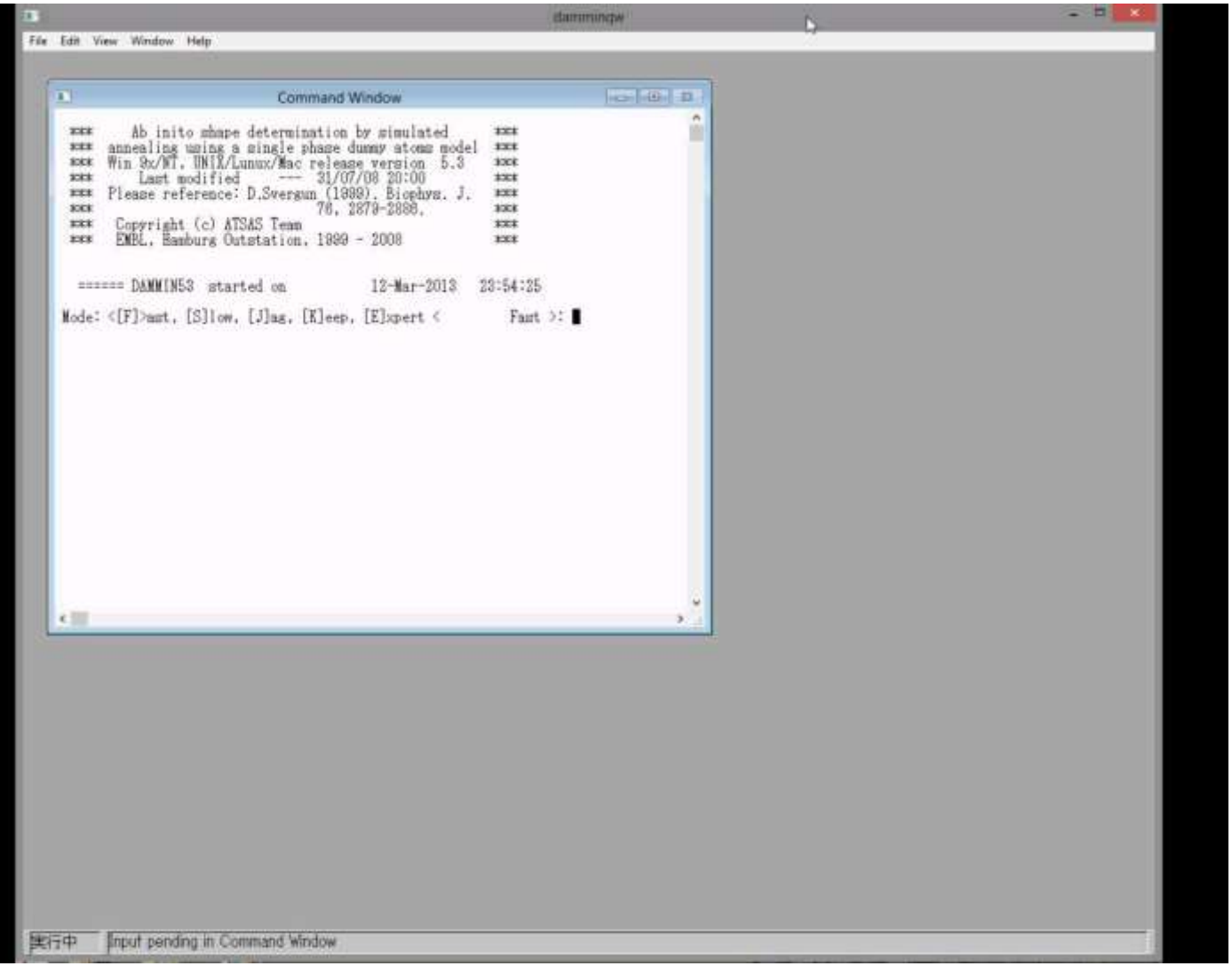
Designed for single phase modeling.
Try MONSA for multiphase modeling (e.g. protein-DNA/RNA complex).



Svergun et al., (2001) Biophys J, 80 2946

Damaver A set of programs to align ab initio models, select the most typical one and build an averaged model.

DAMMIN (GUI) in action



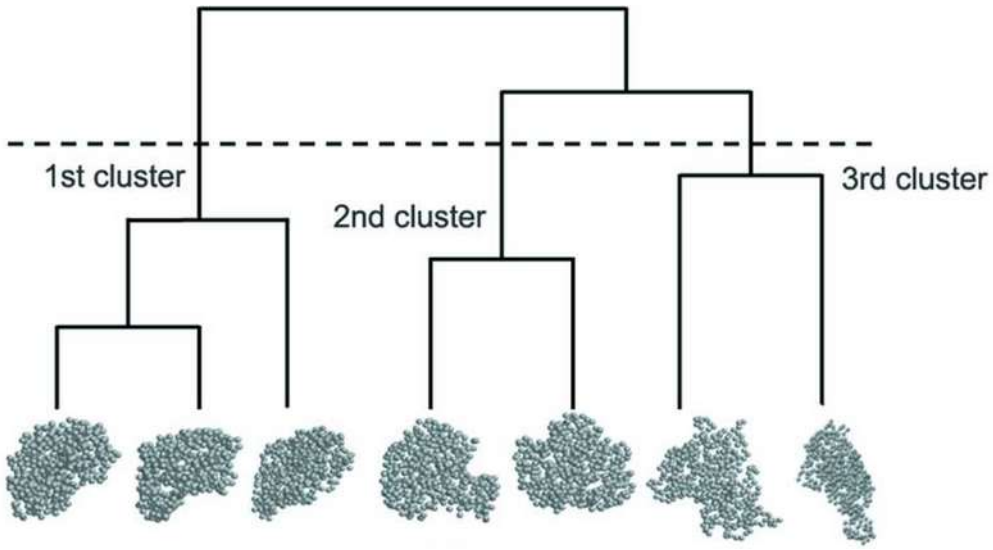
Gnom output file
Symmetry
Aniosometry
(Prolate, Oblate)
Definition of angular unit
…

DAMMIF in action

Ambiguity of SAXS analysis

AMBIMETER Evaluates the ambiguity measure of the given scattering profile

DAMCLUST Clustering the models obtained from ab initio or rigid body modeling.



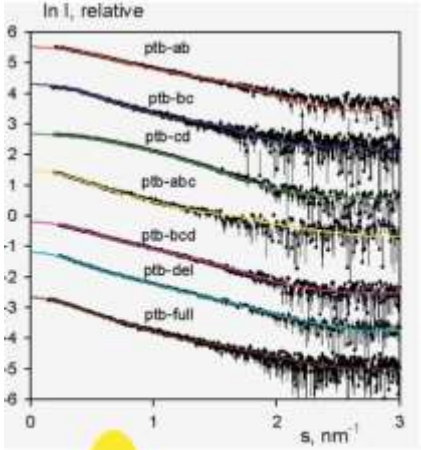
Petoukhov MV, Svergun DI. Acta Crystallogr D Biol Crystallogr. 2015 71:1051-8
Petoukhov MV, et al, J Appl Crystallogr. 2012 45:342-350

ab initio shape determination (multiphase)

Input file: Scattering data (online version)
q_{max} limit: < 0.2 (-0.3)

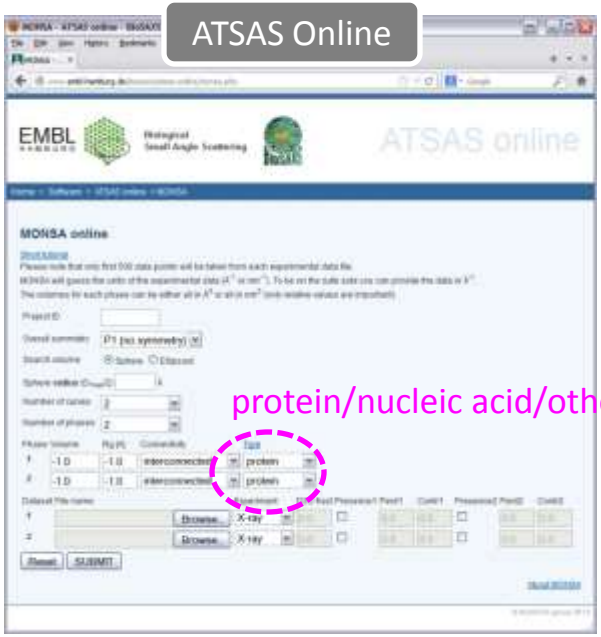
MONSA

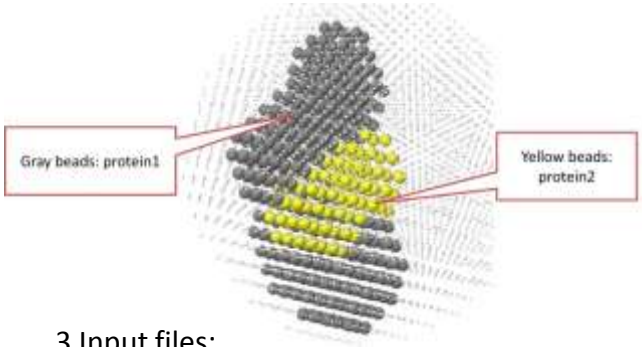
An extended version of DAMMIN for multiphase bead modeling which allows one to fit simultaneously multiple curves (e.g. from X-ray and neutron contrast variation series). Suitable for protein-nucleic acid complexes.



Mapping each component on whole envelope

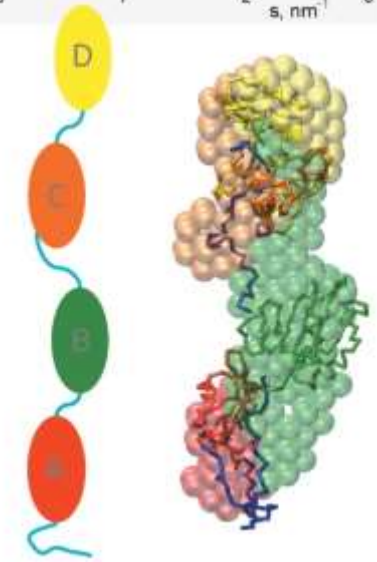
Make sure all conformations are conserved in all curves. No structural change when forming complex.





3 Input files:

protein1, protein2, complex



http://www.embl-hamburg.de/biosaxs/atsas-online/monsa.pdf

Svergun et al. J. Appl. Cryst. 2007, 40, s10

Mixture: Analysis for multicomponent system

MIXTURE Modeling the multicomponent system represented by simple geometrical bodies taking into account interparticle interactions.

OLIGOMER Finding the volume fractions by linear combination of input models.

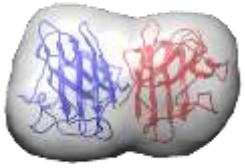
SASREFMX Rigid body modeling taking account of mixture equilibrium.

GASBORMX Ab initio modeling taking account of mixture equilibrium.

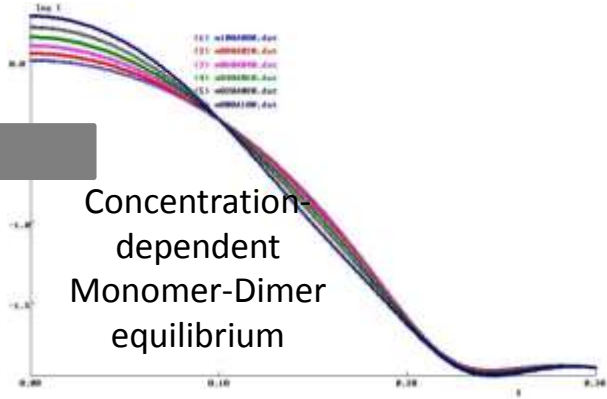
SVDPLOT

Single Value Decomposition

Make sure that transition is NOT intermolecular interactions.



Envelope from mixture
(20% monommer & 80% dimer)



Concentration-dependent Monomer-Dimer equilibrium

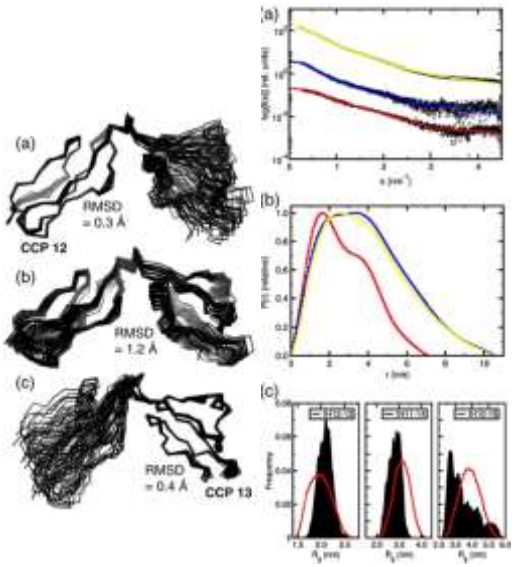
Flexible system

Input file: Scattering data
sequence file
pdb files
constraint, etc

EOM

Ensemble Optimization Method

Fits the averaged theoretical scattering intensity from an ensemble of conformations into the experimental SAXS data.



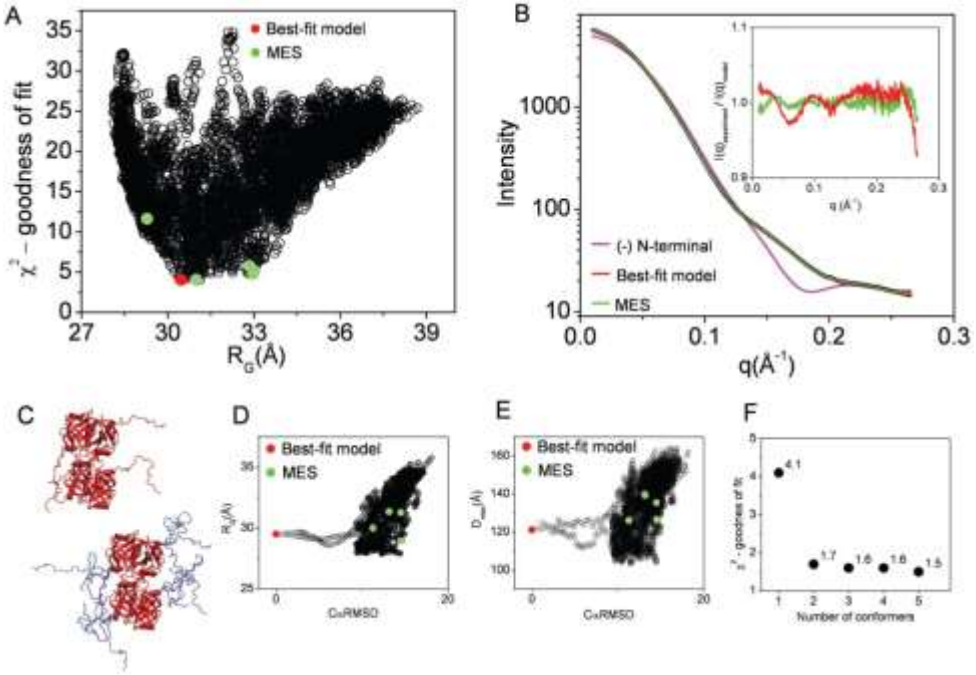
J Mol Biol. 2010 395, 105

BILBOMD + FoXS

Conformational sampling using molecular dynamics (MD) approach.

+

Minimal Ensemble Search (MES)

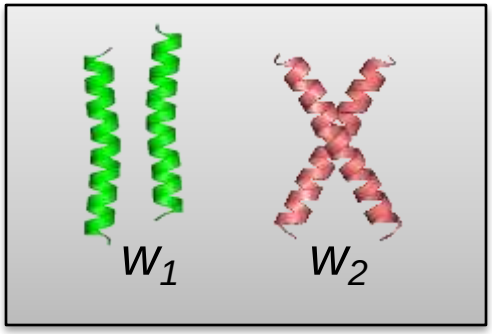


Gen Physiol Biophys. 2009 Jun;28(2):174-89

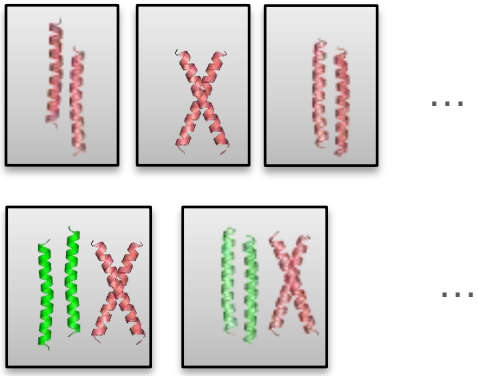
Heterogeneous sample requires a multi-state model

Heterogeneous sample — compositional or conformational heterogeneity in the sample used to generate the data

Multi-state model — a model that specifies two or more co-existing structural states and values for any other parameter

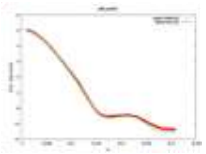


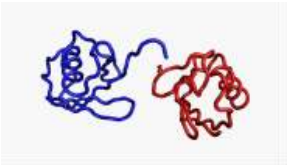
Ensemble of models — an ensemble of (good scoring) single or multi-state models



MultiFoXS

salilab.org/multifoxs

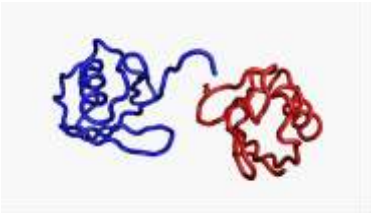




flexible residues:
35 A
36 A
…

Sampling

Generation of conformations with Rapidly exploring Random Trees (RRTs)

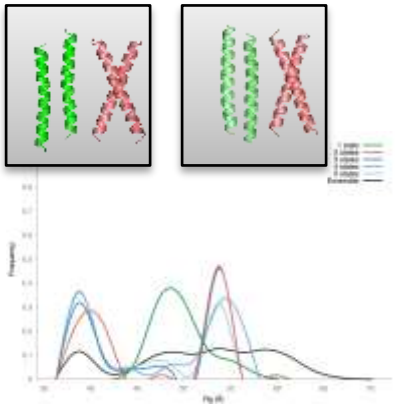


Scoring

Debye formula
salilab.org/foxs

Enumeration

Enumeration of multi-state models that fit the data within noise



…

Analysis

Quality of fit to data, Rg variance among top scoring models

SSRL BL4-2

Thomas M. Weiss

Ivan Rajkovic

Ping Liu

All collaborators

All BL4-2 users