ILLINOIS INSTITUTE OF TECHNOLOGY

CSRRI
CENTER FOR SYNCHROTRON RADIATION RESEARCH & INSTRUMENTATION

BioCAT

Application of ATSAS software package on biological macromolecules

Weifeng Shang

IIT, Center for Synchrotron Research and Instrumentation

Sector 18/BioCAT@APS

Outline

1. What is ATSAS
2. Buffer subtraction and Guinier analysis (PRIMUS)
3. Pair distance distribution function (GNOM)
4. *ab initio* dummy atom modeling (DAMMIN/DAMMIF, MONSA, GASBOR)
5. SAXS profile calculated from atomic models (CRYSOL)
6. Rigid body modeling (SASREF)
7. Combination of ab initio and rigid body modeling (BUNCH)
8. Ensemble Optimization Method (EOM) for intrinsically unfolded proteins
9. Mixtures (OLIGOMER)

What is ATSAS

EMBL HAMBURG

Biological
Small Angle Scattering

BioSAXS

http://www.embl-hamburg.de/biosaxs/software.html

Home > Software

- Group members
- Facilities
- Courses
- Software
  - Download ATSAS
  - ATSAS online
  - BUNCH
  - CORAL
  - CRYSOL
  - CRYSON
  - DAMAVER
  - DAMMIF
  - DAMMIN
  - DARA
  - DATtools
  - EOM
  - GASBOR
  - GLOBSYMM
  - GNOM
  - MASSHA
  - MIXTURE
  - MONSA
  - OLIGOMER
  - PEAK
  - PRIMUS
  - SASREF
  - SUPCOMB
  - Manuals
- Contact us

Data analysis software ATSAS 2.5.1

A program suite for small-angle scattering data analysis from biological macromolecules

Data processing

PRIMUS - manipulations with experimental 1D SAS data
GNOM - indirect transform program that evaluates the particle distance distribution function p(r)
Data manipulation and analysis tools - AUTORG, ALMERGE, DATGNOM, DATPOROD etc.

***Ab initio* methods**

DAMMIN - *ab initio* shape determination using a dummy atom model
DAMMIF - rapid shape determination
GASBOR - reconstruction of a protein structure by a chain-like ensemble of dummy residues
MONSA - shape determination using a multiphase dummy atom model

Rigid body modelling

SASREF - modelling of multisubunit complexes
BUNCH - modelling of multidomain proteins against multiple data sets
CORAL - modelling of multidomain protein complexes against multiple data sets
MASSHA - interactive modelling of atomic structures and shape analysis
GLOBSYMM - rigid body modelling of symmetric oligomers

Mixtures and flexible systems

OLIGOMER - volume fractions of mixtures with known scattering intensities from the components
MIXTURE - modelling of multicomponent systems
BUNCH - modelling of multidomain proteins against multiple data sets
EOM - Ensemble Optimization Method for flexible proteins

PDB oriented tools

CRYSOL - X-ray scattering patterns from known hi-res structures
CRYSON - neutron scattering patterns from known hi-res structures
SUPCOMB - superimposes one 3D structure onto another
DAMAVER - align *ab initio* models, select the most typical one
DARA - database for rapid search of structural neighbours

ATSAS online and SAXIER forum

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

EMBL HAMBURG

Biological
Small Angle Scattering

Home > Software > ATSAS online

ATSAS online

Create an account
Change password
Forgot your password?

As a courtesy to other users please do not submit more than 50 jobs at a time.
Following services are available for registered users:

DAMMIN - *ab initio* shape determination by simulated annealing using a bead model
DAMMIF - rapid *ab initio* bead model shape determination
GASBOR - *ab initio* reconstruction of protein structure by a chain-like ensemble of dummy residues
MONSA - multiphase *ab initio* modelling

CRYSOL - evaluation of X-ray solution scattering curves from atomic models
SASREF - modelling of multisubunit complexes from contrast variation and X-ray data
Utility for generation of a contact conditions file to be used in SASREF 6.0 offline

EOM - Ensemble Optimisation Method (for flexible proteins)

DANESSA - Automated data analysis system (alpha version)

My Projects - List of your recent projects (to check/re-run/report a problem)

Questions and feedback

SAXS
Small Angle X-ray Scattering Initiative for Europe :: Forum

Login Register

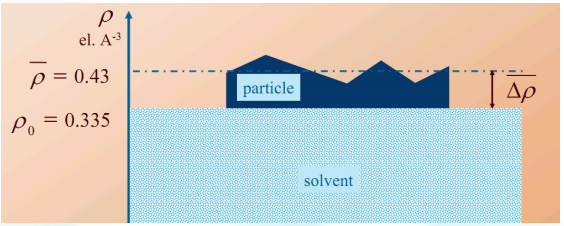
View unanswered posts | View active topics

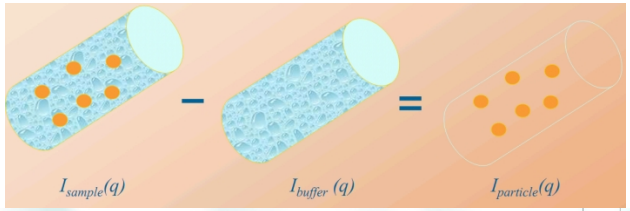
Board index

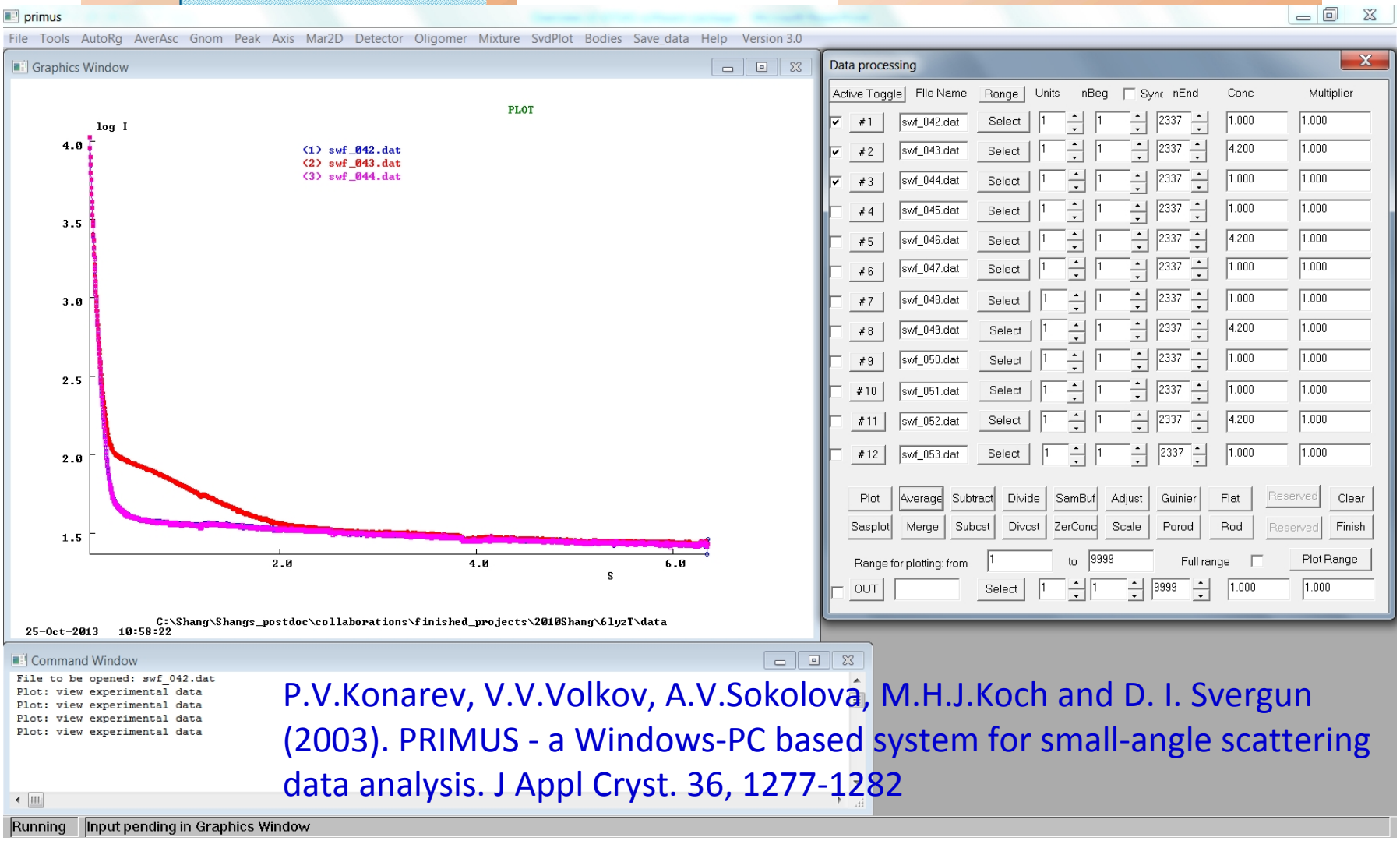
Forum	
General	
	Announcements Courses and workshops, new software releases
	SAS Software Small angle scattering software (except ATSAS)
	SAS Hardware/Instrumentation Detectors, sample changers, home sources, nanomanipulation tools, a phase scattering etc.
	SAS Experiment Design Plan your small angle X-ray or neutron scattering measurements
	Literature New to small angle scattering? Got recommendations for beginners? W

http://www.saxier.org/forum/index.php

PRIMUS: buffer subtraction

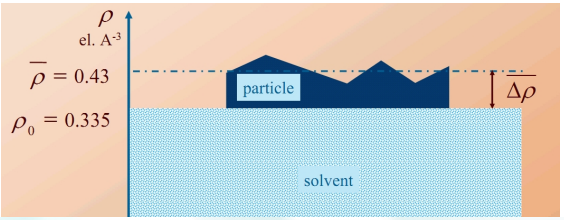


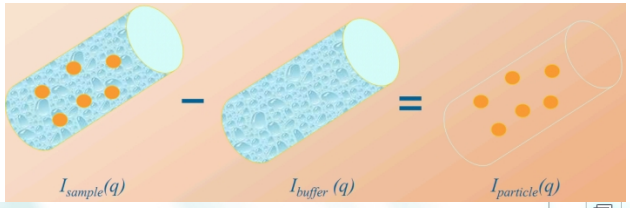


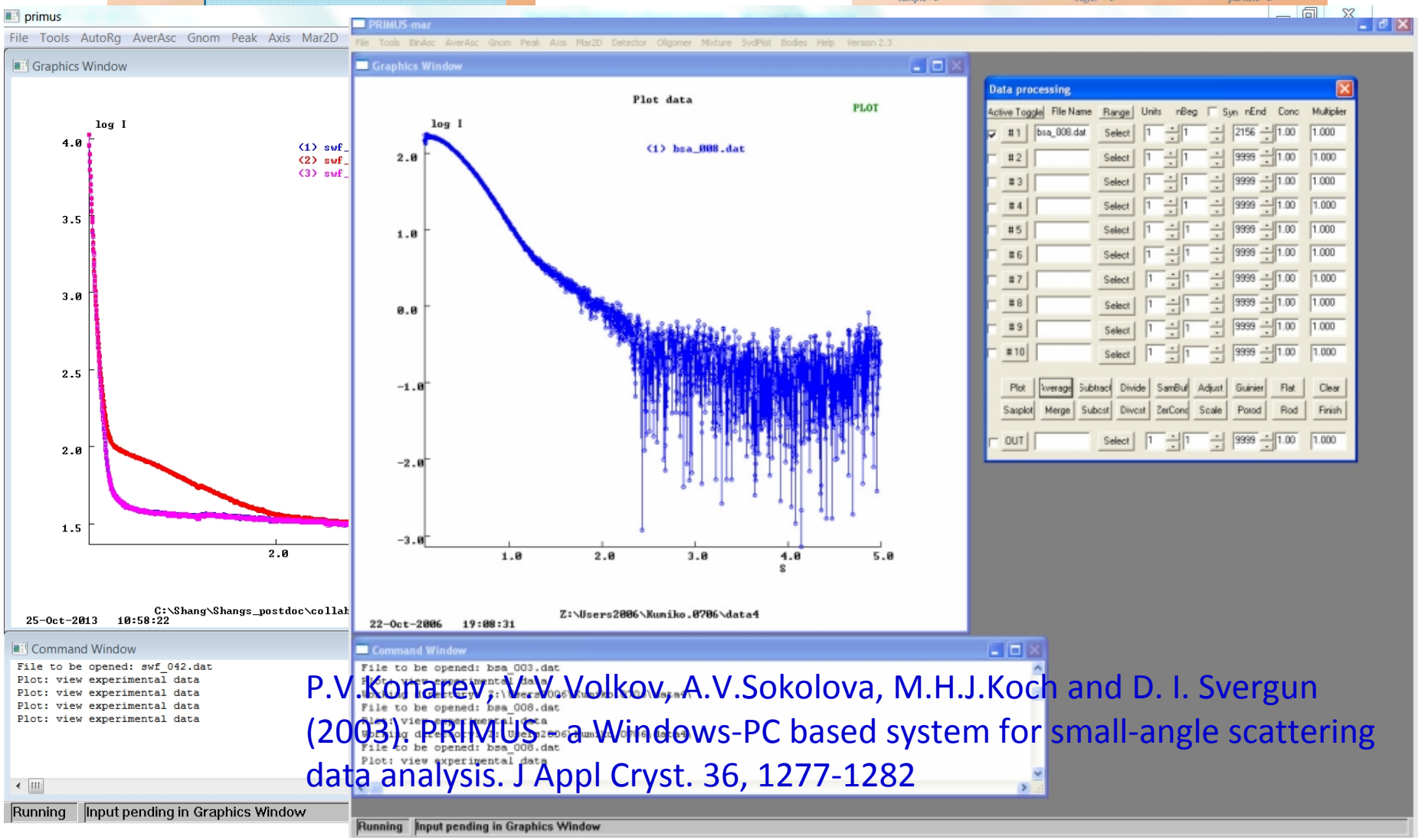


P.V.Konarev, V.V.Volkov, A.V.Sokolova, M.H.J.Koch and D. I. Svergun (2003). PRIMUS - a Windows-PC based system for small-angle scattering data analysis. J Appl Cryst. 36, 1277-1282

PRIMUS: buffer subtraction

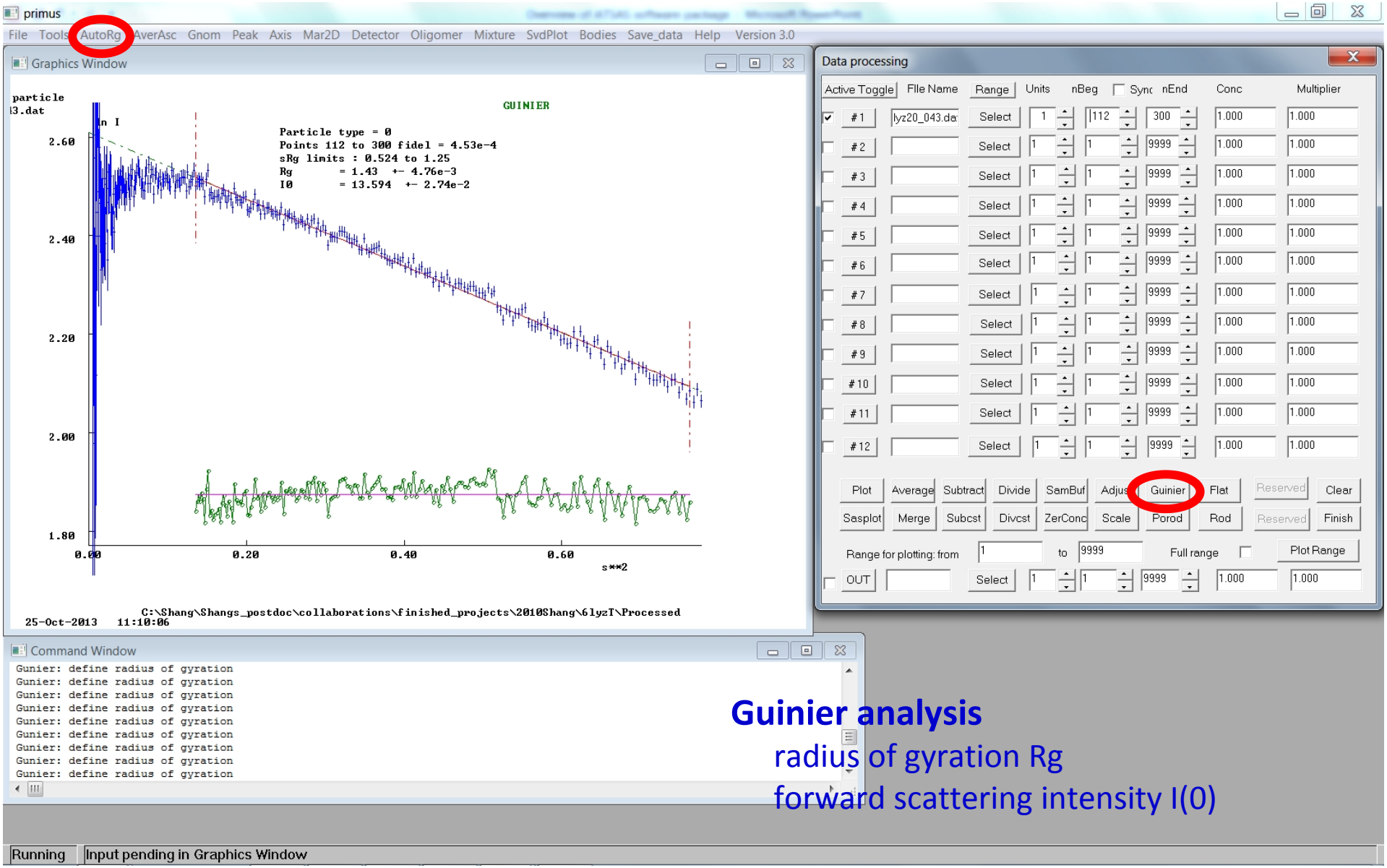






P.V.Konarev, V.V.Volkov, A.V.Sokolova, M.H.J.Koch and D. I. Svergun (2003). PRIMUS – a Windows-PC based system for small-angle scattering data analysis. J Appl Cryst. 36, 1277-1282

PRIMUS: Guinier plot and AutoRg



Guinier analysis

radius of gyration Rg

forward scattering intensity I(0)

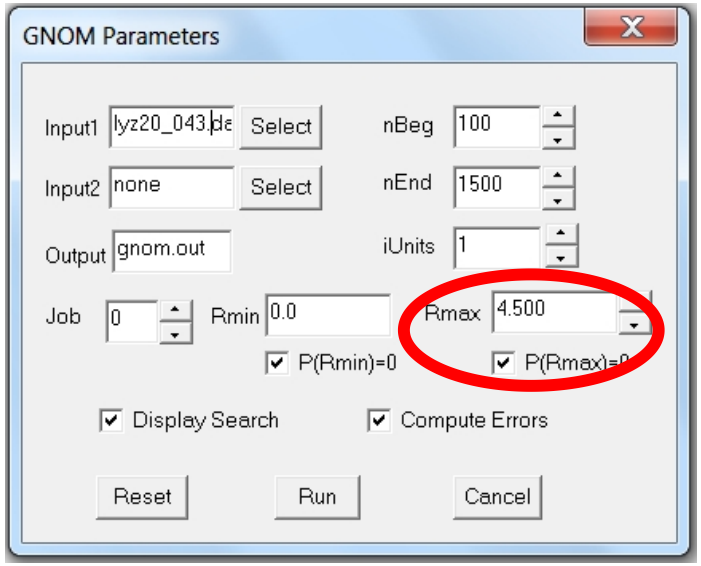
GNOM: p(r) calculation

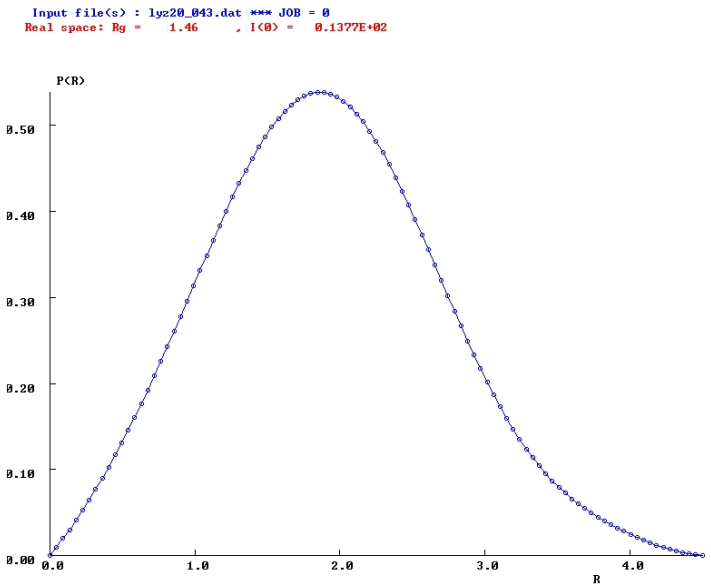
Pair distance distribution funciton p(r)
histogram of distances between any pair of scattering points within the particle weighted by the respective electron density

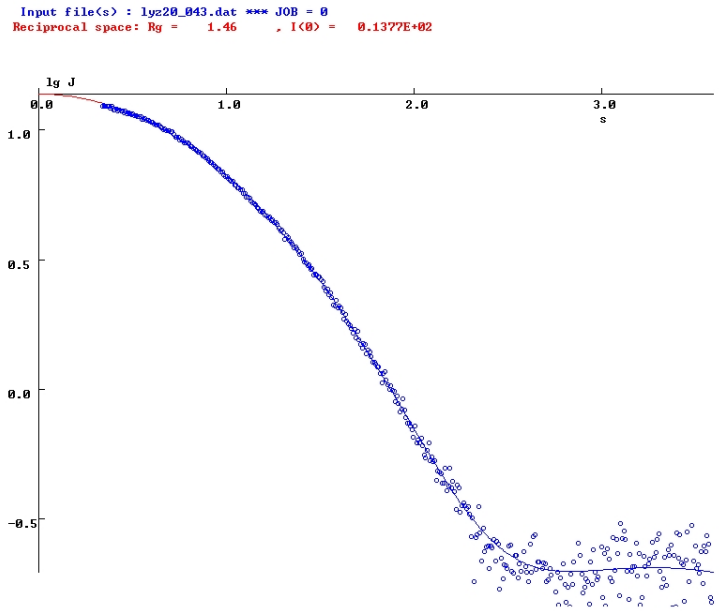
$$p(r) = \frac{r^2}{2\pi^2}\int_0^\infty q^2\, \mathrm{I}(q)\frac{\sin(qr)}{qr}dq$$

Experimental data I(q): **incomplete and noisy!**

Svergun D.I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495-503





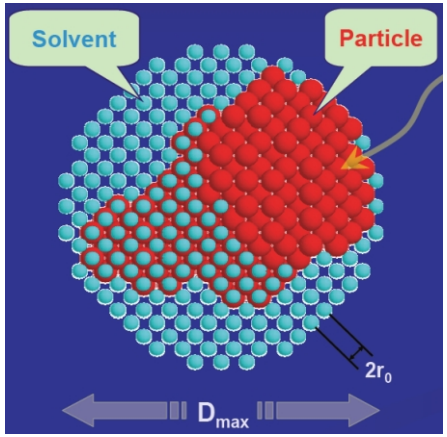


ab initio modeling programs: DAMMIN/MONSA

A search volume, e.g. a sphere of diameter D_{max}, is filled by densely packed beads of radius $r_0 << D_{max}$

Each bead is assigned to either solvent or particle phase

Number of model parameters ~ 1000 -2000



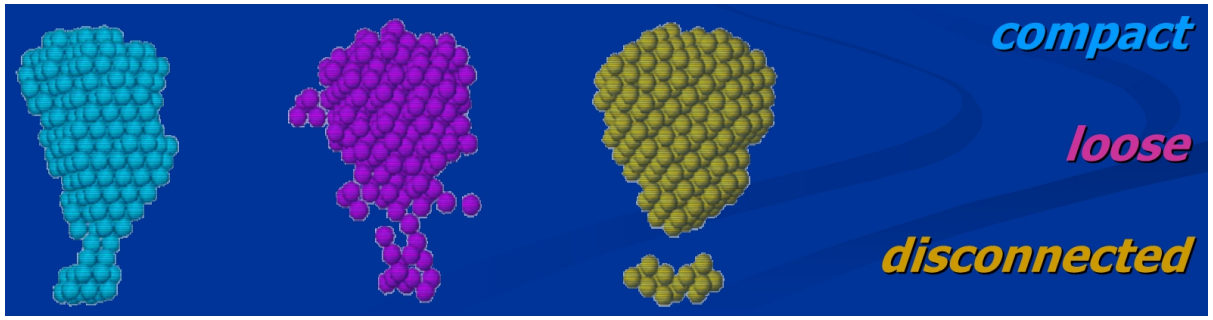
Using the simulated annealing method to minimize the goal function:

$f(X) = \chi^2 + \Sigma \alpha P(X)$

χ^2: Discrepancy between the experimental and calculated SAXS profiles;

P(X): Penalty term, e.g. the looseness of the dummy atom configuration;

α: Weighting factor, e.g. for the looseness penalty



D. I. Svergun (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* 76, 2879-2886.

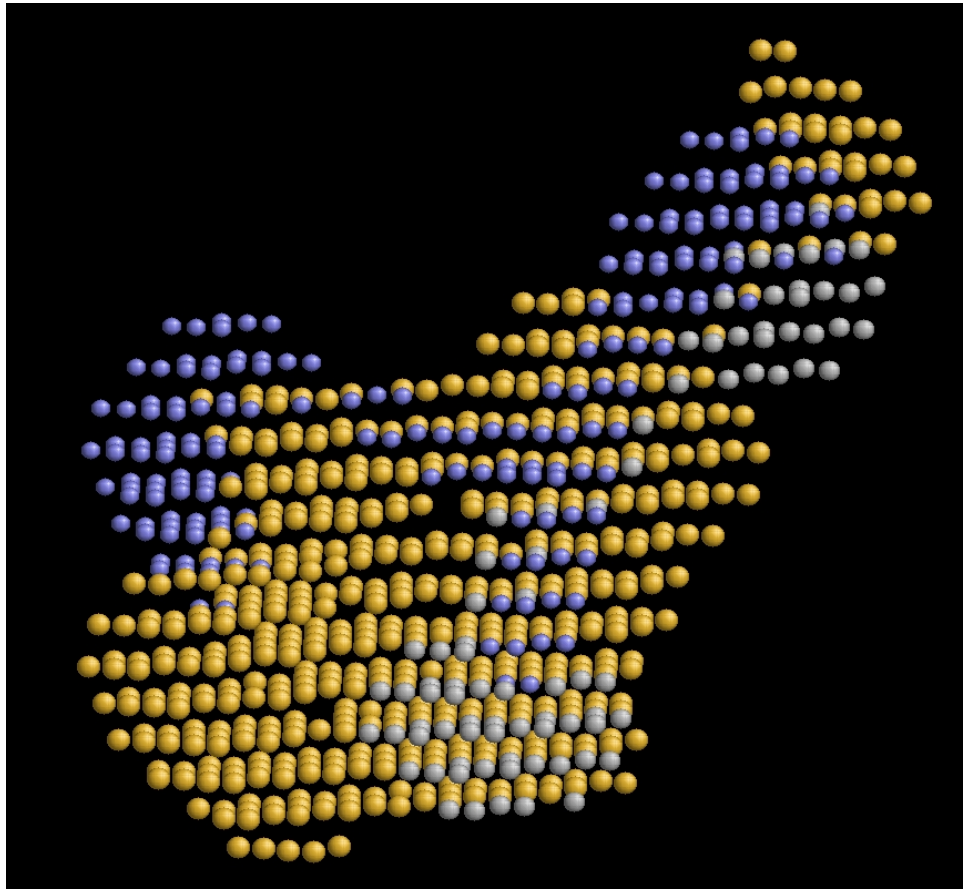
Multiphase *ab initio* modeling: MONSA

Four constructs:

1. Full length (1305 AA)
2. Construct 1-1010
3. Construct 133-1010
4. Construct 133-1305

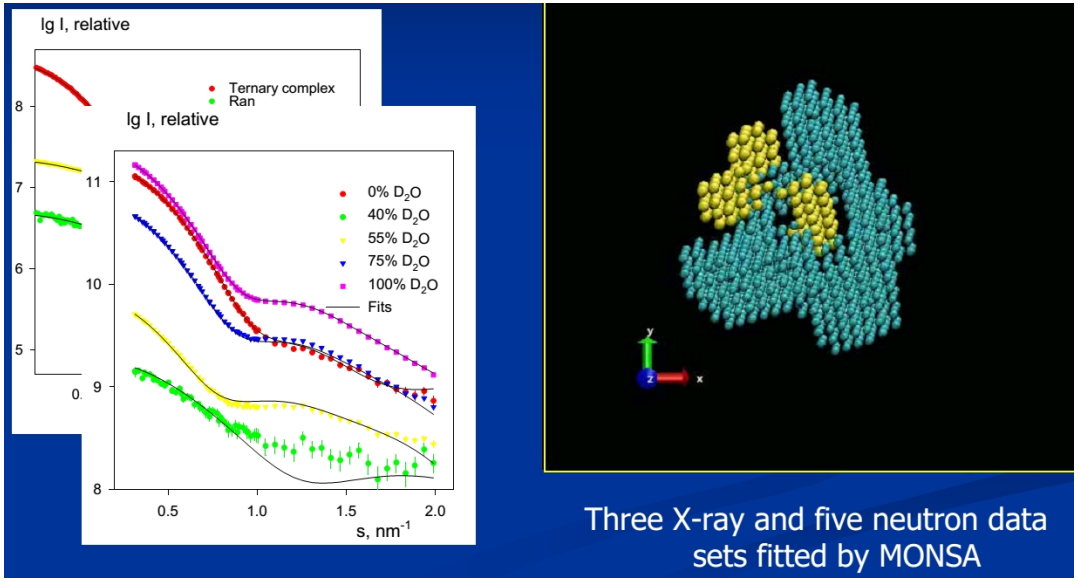
Phases:

1. Residues 1-133
2. Residues 133-1010
3. Residues 1010-1305



Multiphase *ab initio* modeling: MONSA

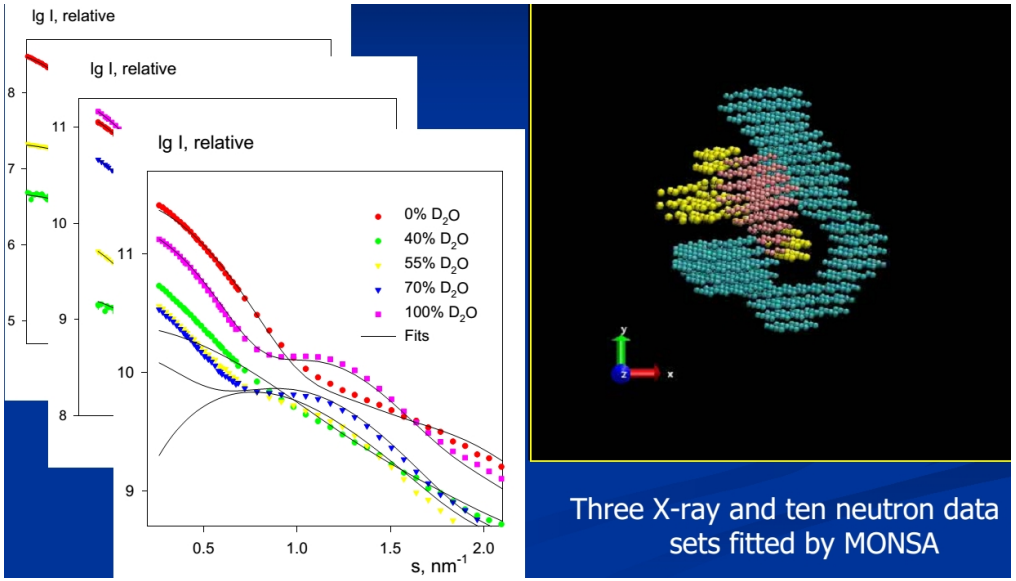
Ternary complex: Exportin-t/Ran/tRNA



Three X-ray and five neutron data sets fitted by MONSA

Localization of tRNA

Deuterated Ran protein



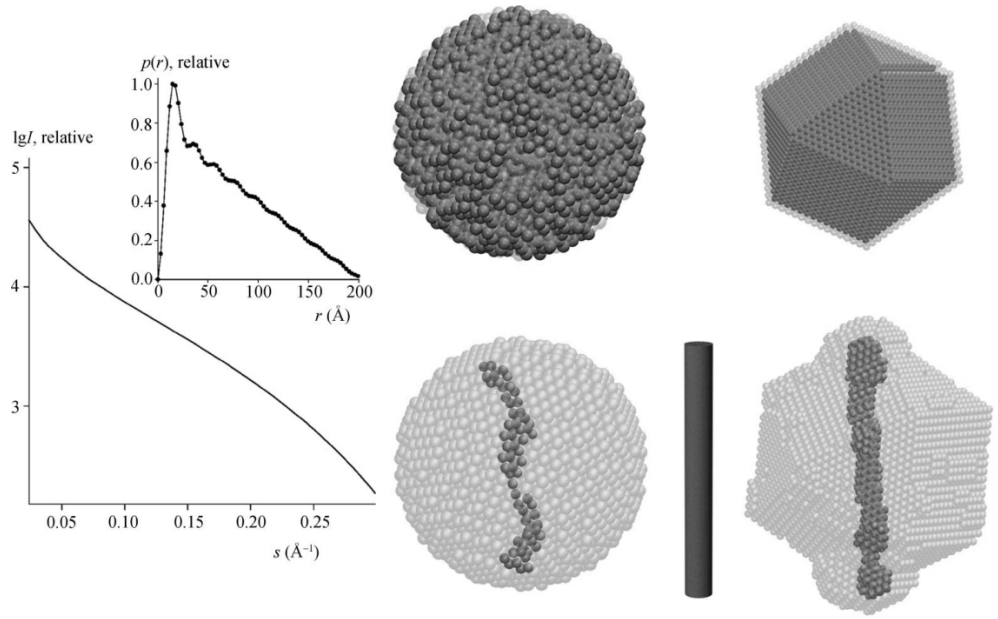
Three X-ray and ten neutron data sets fitted by MONSA

ab initio modeling programs: DAMMIF

Similar approach as DAMMIN (beads, simulated annealing etc.)

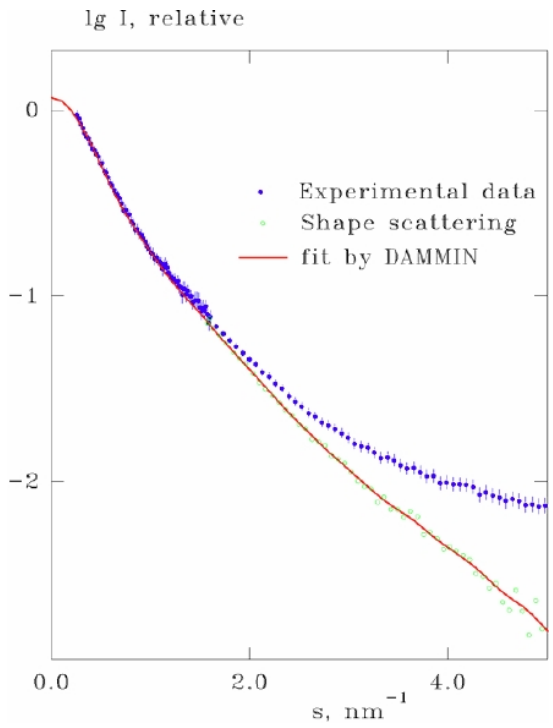
What is different?

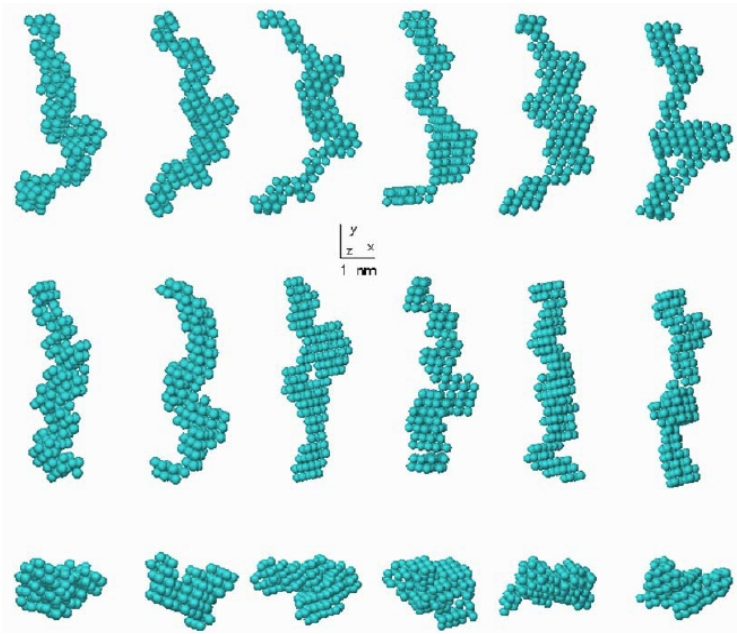
- Source code and compiler optimization for speed (parallelization for multi-core/CPU systems)
- Test of connectivity at each SA step before scattering calculation/comparison
- Adaptive search volume
- R_g penalty term



Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.*, **42**, 342-346.

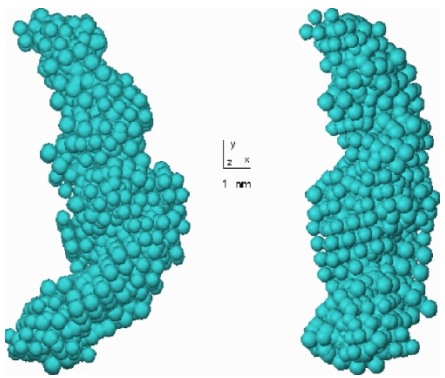
Average dummy atom models: DAMAVER



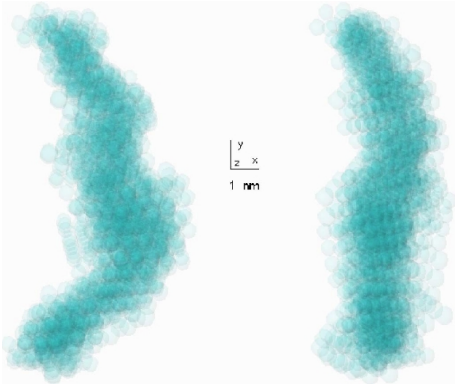


All models aligned and superimposed with the program SUPCOMB

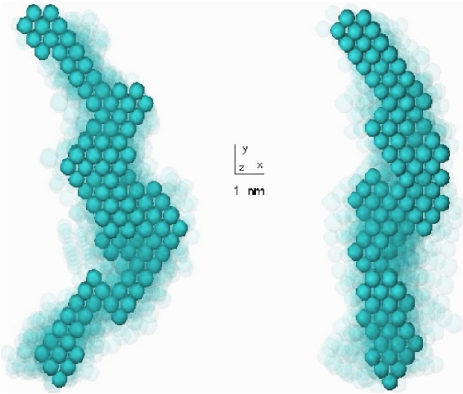
Solution spread region



Most populated region

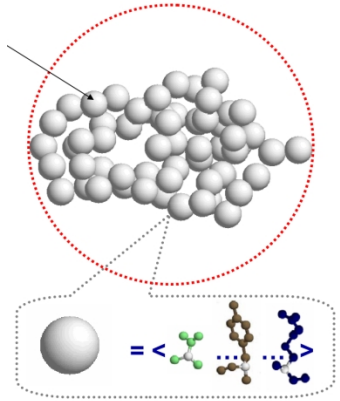


final averaged model

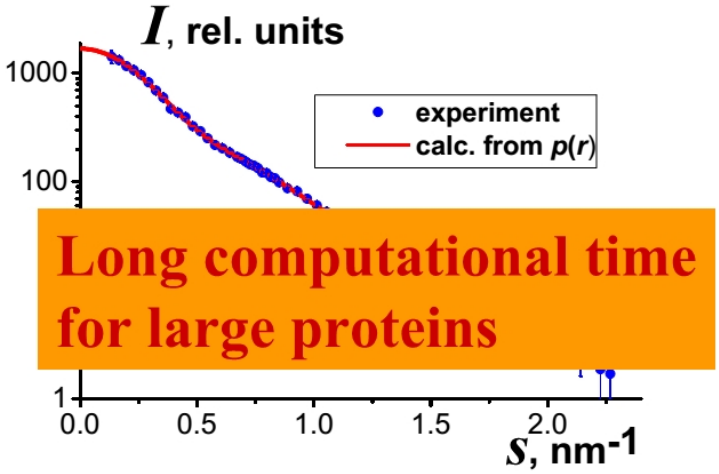


ab initio modeling programs: GASBOR

Proteins consist of folded polypeptide chains of amino acid residues.
At a resolution of 5 Å each AA residue can be represented as a sphere (dummy residue)
Nonlinear search for a chain-like arrangement of dummy residues which fits the experimental data.

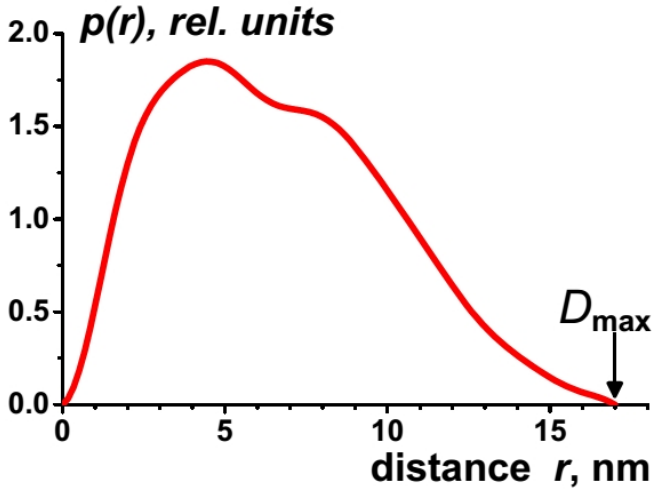


$$f(X) = \chi^2 + \Sigma\alpha_i P_i(X)$$



Long computational time for large proteins

Gasbor reciprocal space

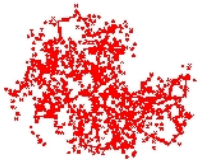


Gasbor real space

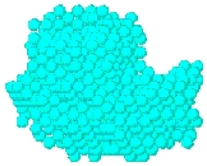
Svergun, D.I., Petoukhov, M.V. and Koch, M.H.J. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.*, **80**, 2946-2953.

Scattering calculated from atomic coordinates: CRYSOL

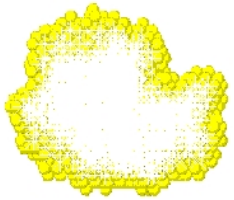
$$I(s) = \left\langle \left| A(\mathbf{s}) \right|^2 \right\rangle_\Omega = \left\langle \left| A_a(\mathbf{s}) - \rho_s A_s(\mathbf{s}) + \delta\rho_b A_b(\mathbf{s}) \right|^2 \right\rangle_\Omega$$



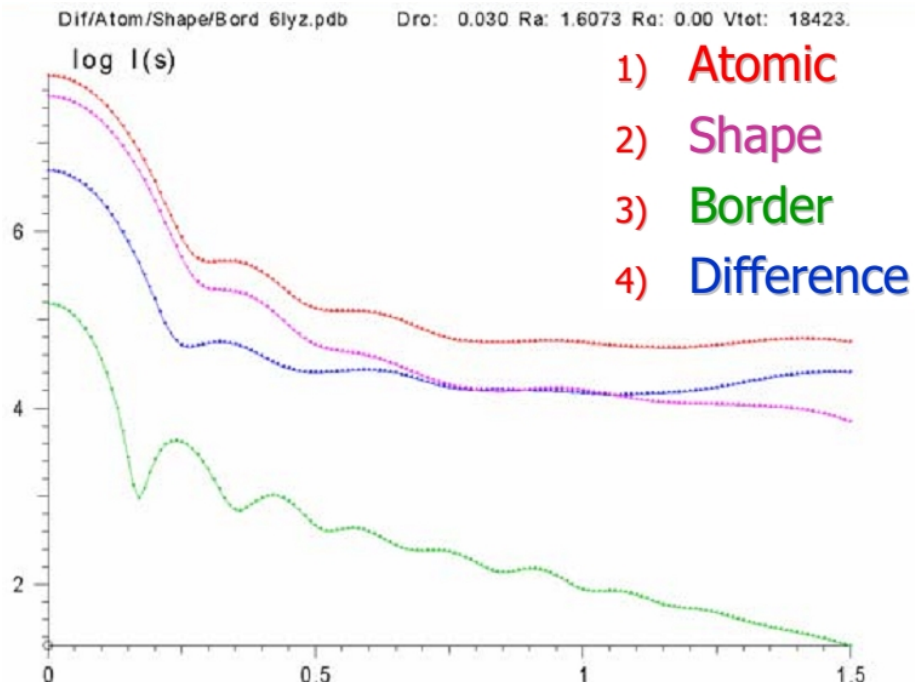
$A_a(s)$: atomic scattering amplitude in vacuum



$\rho_s A_s(s)$: scattering amplitude from the excluded volume of the solvent



$\delta\rho_b A_b(s)$: scattering amplitude of the hydration shell

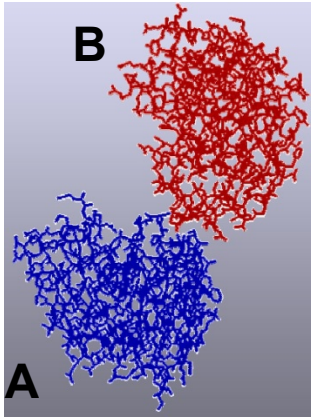


Predict SAXS profile

Validate crystal structures in solution (fit the experimental data using two free parameters, i.e. average displaced volume per atomic group and the contrast of the hydration shell)

Similar program CRYSON for neutron scattering!

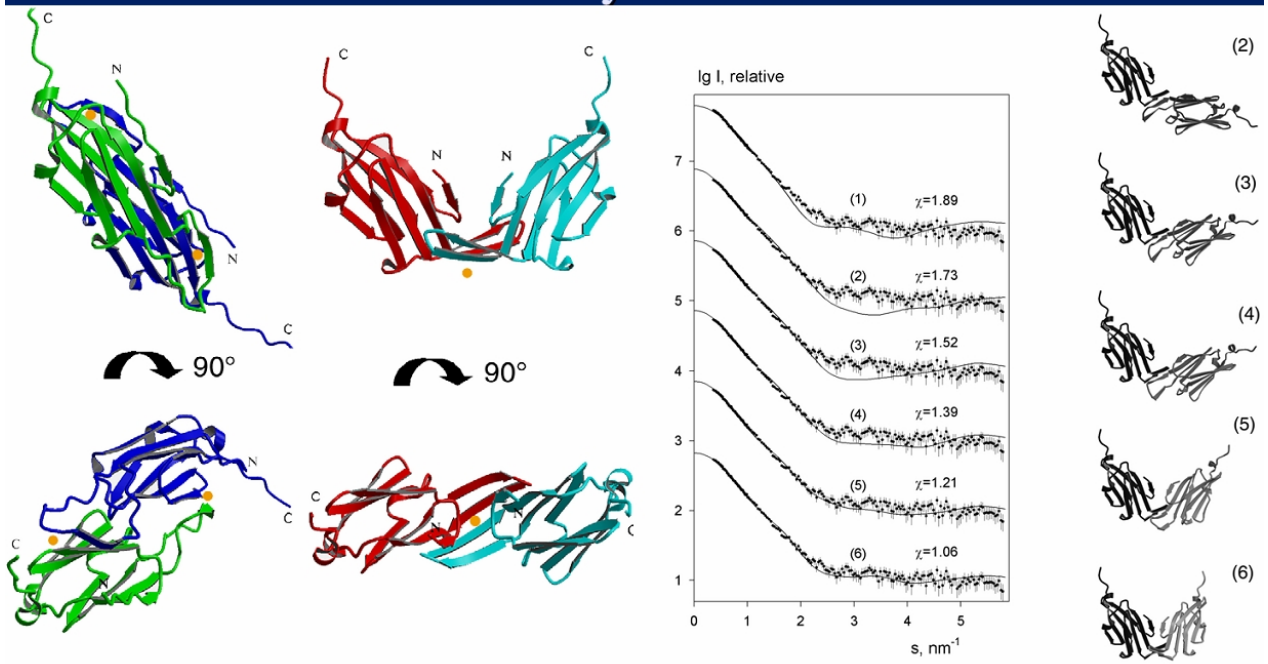
Rigid body modeling: MASSHA



Atomic structure of the two subunits are known, but the relative orientation of the overall structure unknown.

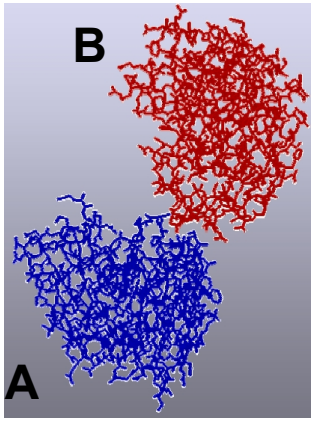
Fix the subunit A and change the coordinates of the subunit B (shift x, y, z, rotation α, β, γ)

Manual refinement: quaternary structure of the dimeric α-crystallin domain



Feil, I.K., Malfois, M., Hendle, J., van der Zandt, H. & Svergun, D.I. (2001) *J. Biol. Chem.* **276**, 12024-12029

Rigid body modeling: SASREF



Atomic structure of the two subunits are known, but the relative orientation of the overall structure unknown.

Fix the subunit A and change the coordinates of the subunit B (shift x, y, z, rotation α, β, γ)

Simulated annealing

Aim: find a vector of *M* variables *{x}* minimizing a function *f(x)*

1. **Start from a random configuration *x* at a "high" temperature *T*.**
2. **Make a small step (random modification of the configuration) $x \rightarrow x'$ and compute the difference $\Delta = f(x') - f(x)$.**
3. ***If* $\Delta < 0$, accept the step; if $\Delta > 0$, accept it with a probability $e^{-\Delta/T}$**
4. ***Make another step from the old (if the previous step has been rejected) or from the new (if the step has been accepted) configuration.***
5. ***Anneal the system at this temperature, i.e.* repeat steps *2-4* *"many" (say, 100M* tries or *10M* successful tries, whichever comes first) times, then decrease the temperature *(T' = cT, c<1).***
6. ***Continue cooling the system until no improvement in f(x)* is observed.**

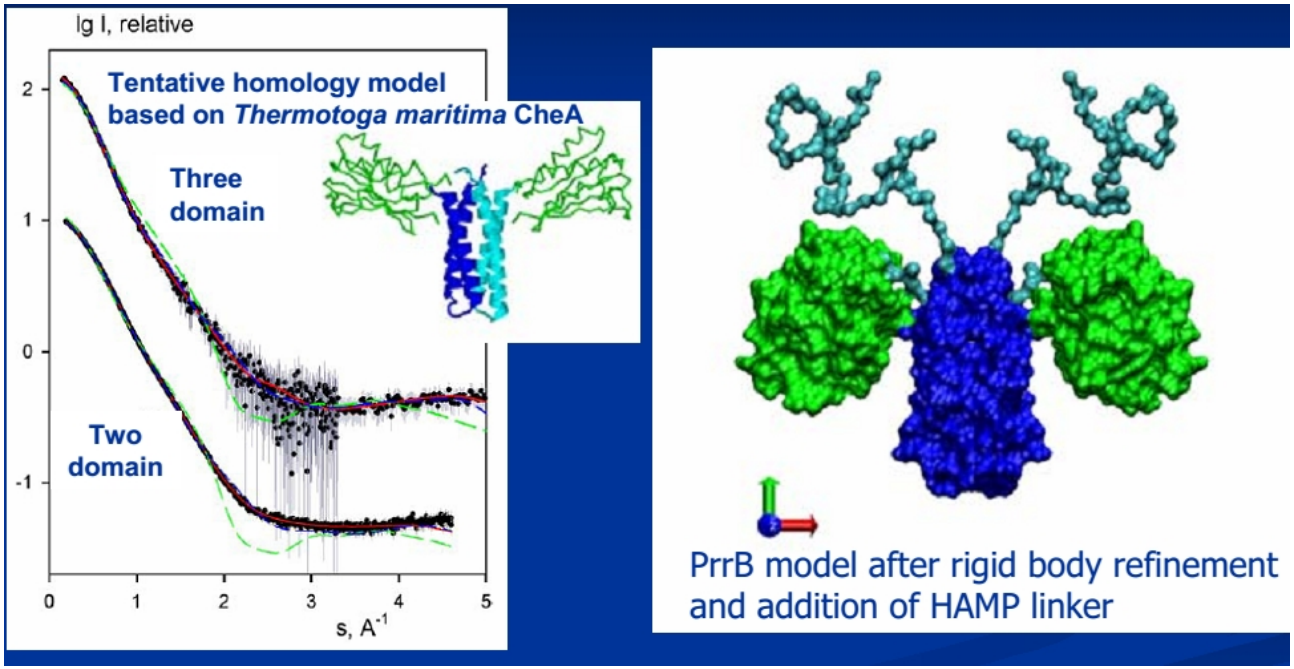
Shape determination: M≈ 10^3 variables (e.g. 0 or 1 bead assignments in DAMMIN

Rigid body methods: M≈ 10^1 variables (positional and rotational parameters of the subunits)

f(x) is always (Discrepancy + Penalty)

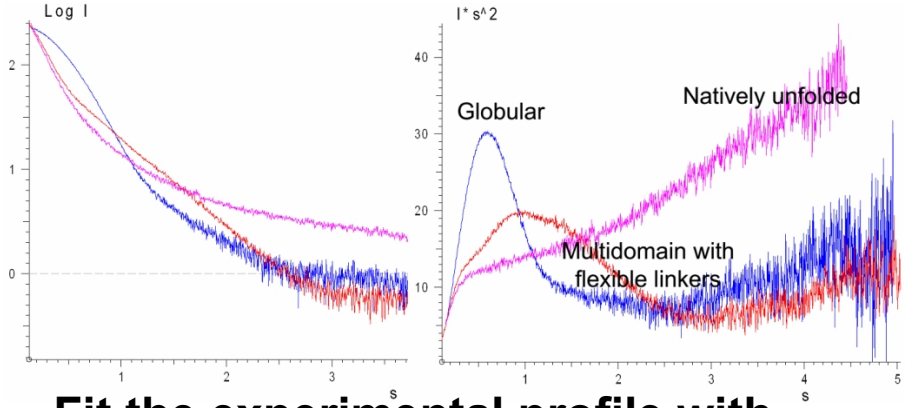
hybrid modeling: BUNCH

BUNCH combines rigid body and *ab initio* modeling to find the positions and orientations of rigid domains and the probable conformations of the flexible linker represented as a chain of dummy residues.



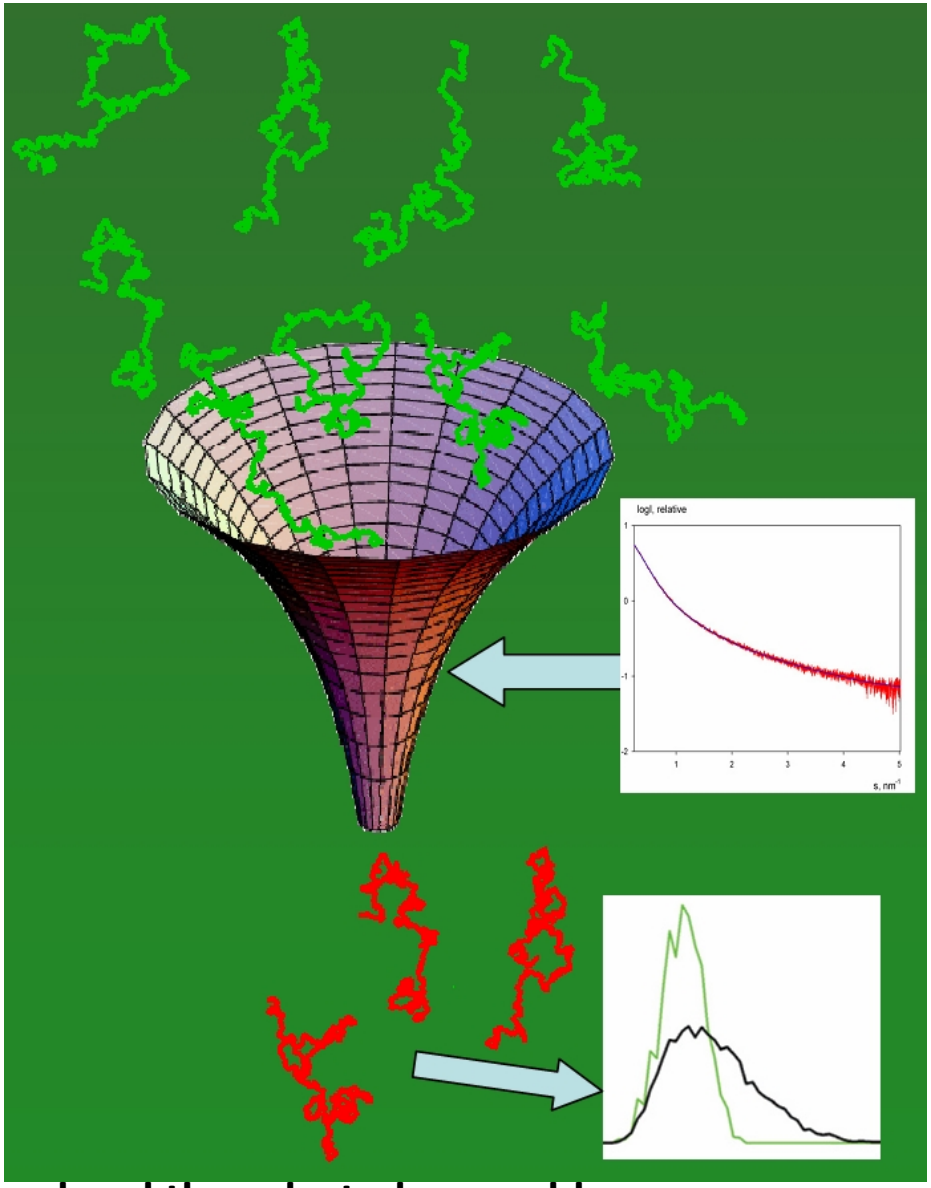
The dimeric sensor histidine-kinase PrrB from Mycobacterium tuberculosis contains ATP binding and dimerization domains and a 59 aas long (flexible) HAMP linker
Nowak, E., Panjikar, S., Morth, J. P., Jordanova R., Svergun, D. I. & Tucker, P. A. (2006) Structure, 14, 275

Ensemble Optimization Method (EOM) for flexible systems



Fit the experimental profile with an ensemble of models selected from a randomly generated model pool.

1. **RanCh (generate the random pool)**
2. **CRYSOL (calculate the scattering profile of each conformation in the pool)**
3. **GAJOE (select an ensemble from the pool fitting the data using genetic algorithm)**



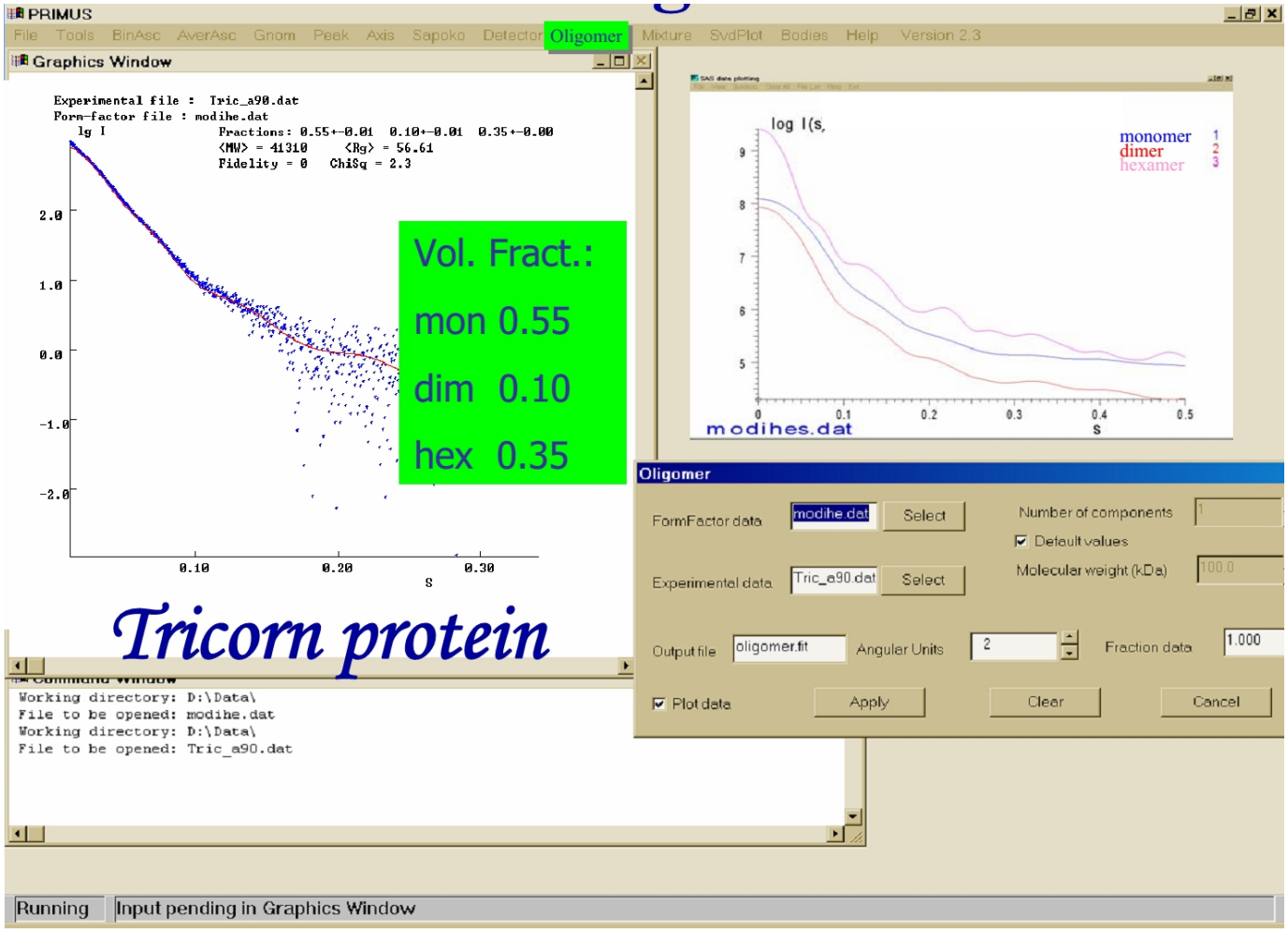
Result

Distribution of Rg and Dmax for the random pool and the selected ensemble.

Mixtures

Monomer-Dimer equilibrium
Complex formation

$$I(s)=\sum_k v_k I_k(s)$$



Goettig, P. et.al., JBC (2005) 280, 33387-33396

What SAS can do for macromolecules

! Mono-disperse system !

nothing known, single scattering pattern

low-resolution shape (DAMMIN/DAMMIF, GASBOR)

nothing known, multiple scattering patterns

low-resolution structure (MONSA)

incomplete atomic structure (PDB) available

addition of missing domains (CREDO)

addition of missing residues (BUNCH)

complete atomic structure (PDB) available

validation of the structure in solution (CRYSOL)

conformation changes? (CRYSOL)

atomic structures (PDB) for the subunits available

quaternary structure (SASREF)