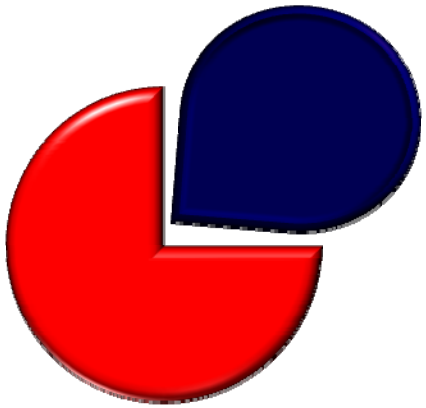
Mixtures, Assemblies, Flexible Systems

Maxim Petoukhov
EMBL, Hamburg Outstation

Complex vs Mixture

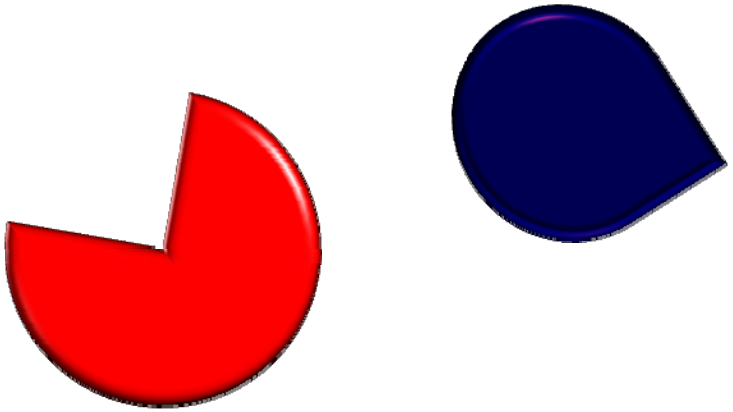
- Equimolar ratio A and B



Correlated

$A_{a+b}(\boldsymbol{s}) = A_a(\boldsymbol{s})+A_b(\boldsymbol{s})$

$I_0 \sim M_a+M_b$



Independent

$I_{a+b}(s)=c_a I_a(s)+c_b I_b(s)$

$I_0 \sim (M_a^2+M_b^2)/(M_a+M_b) \leq MAX(M_a, M_b)$

Scattering from mixtures

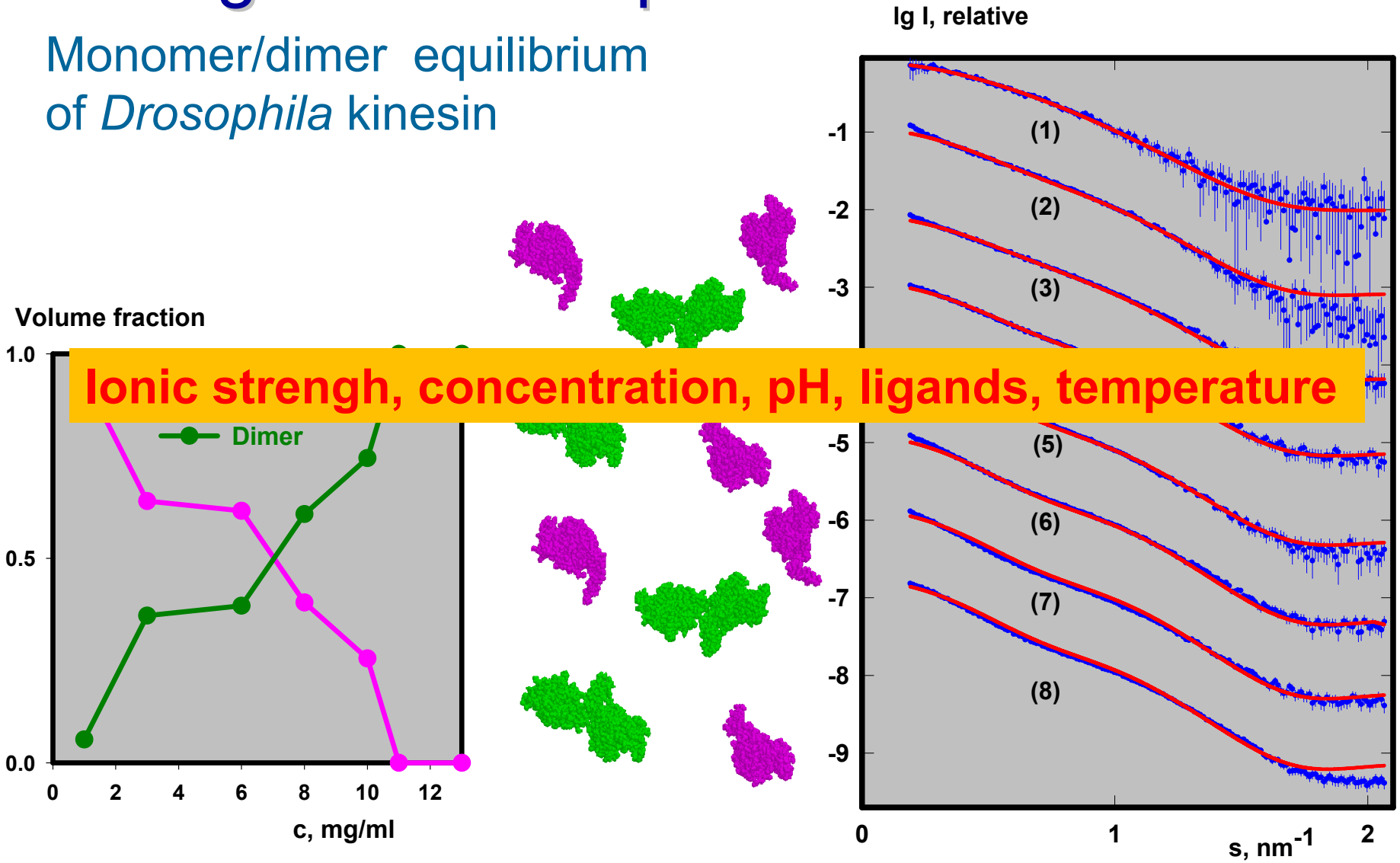
$$I(s)=\sum_k v_k I_k(s)$$

For mixtures, solution scattering permits to determine the number of components and, given their scattering intensities $I_k(s)$, also the volume fractions

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003) *J. Appl. Cryst.* **36**, 1277

Oligomeric composition of mixtures

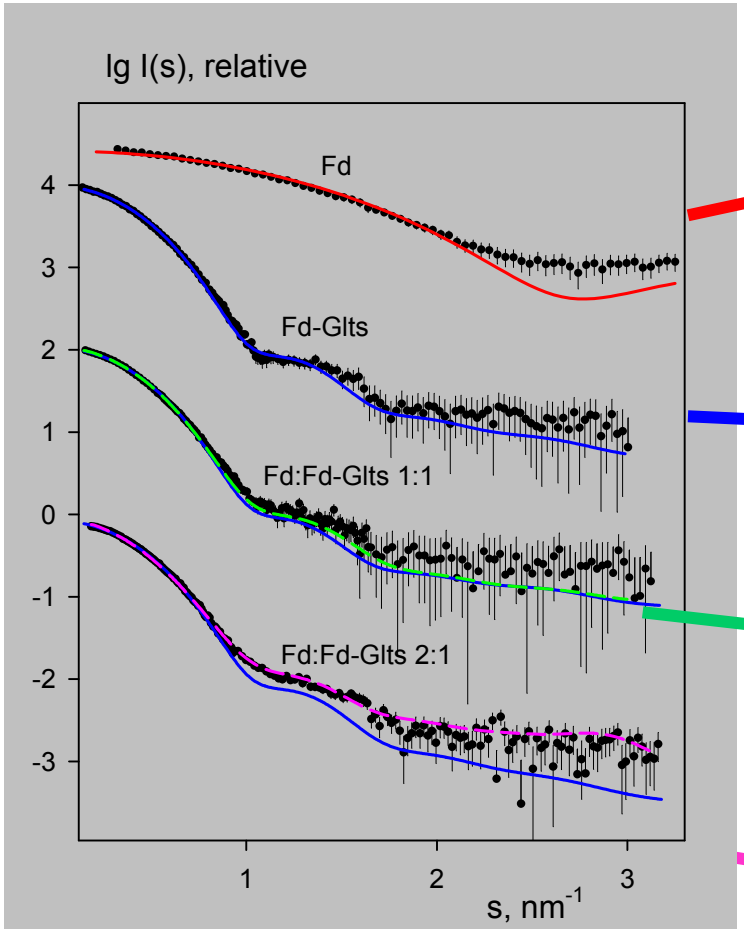
Monomer/dimer equilibrium of *Drosophila* kinesin

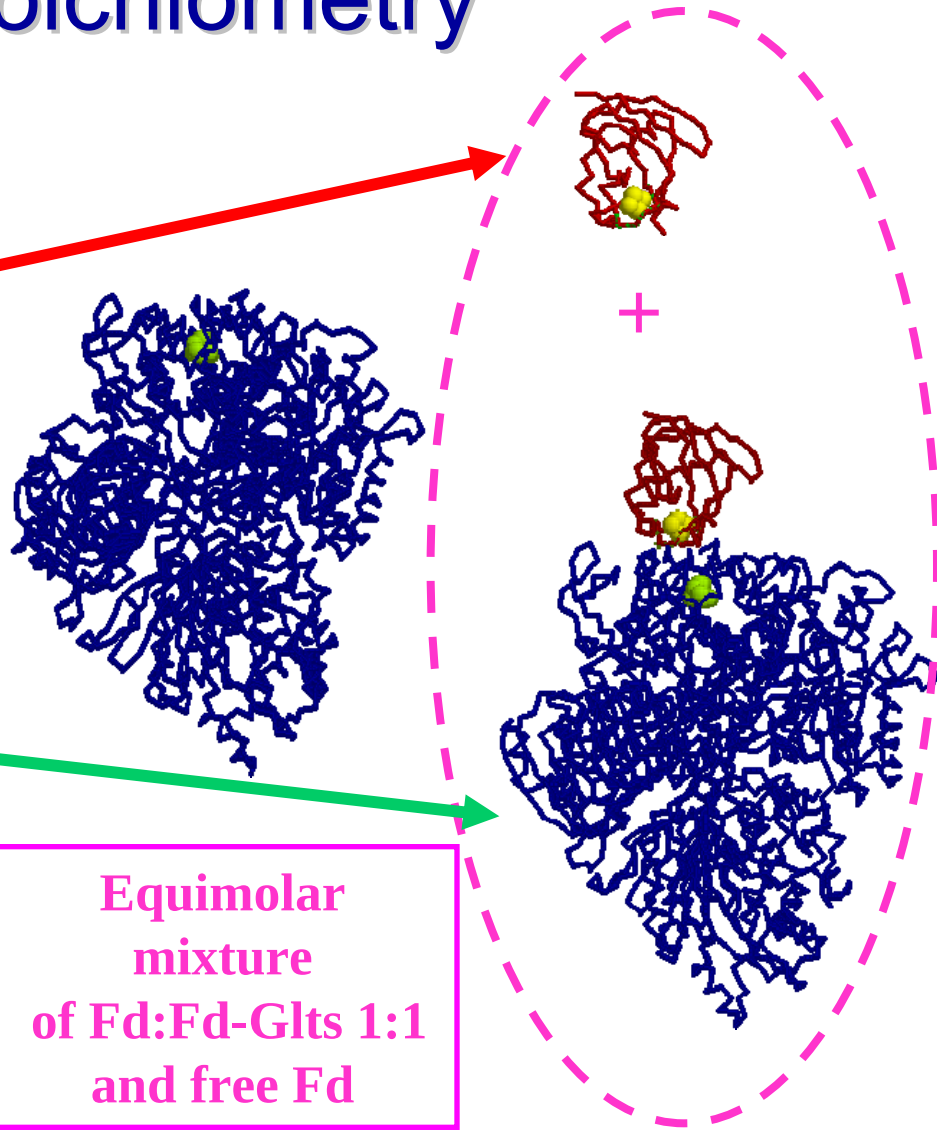


Ionic strengh, concentration, pH, ligands, temperature

Kozielski, F., Svergun, D.I., Zaccai, J. Wade, R.H. & Koch, M.H.J. (2001) *J. Biol. Chem.* **276**, 1267

Complex Stoichiometry

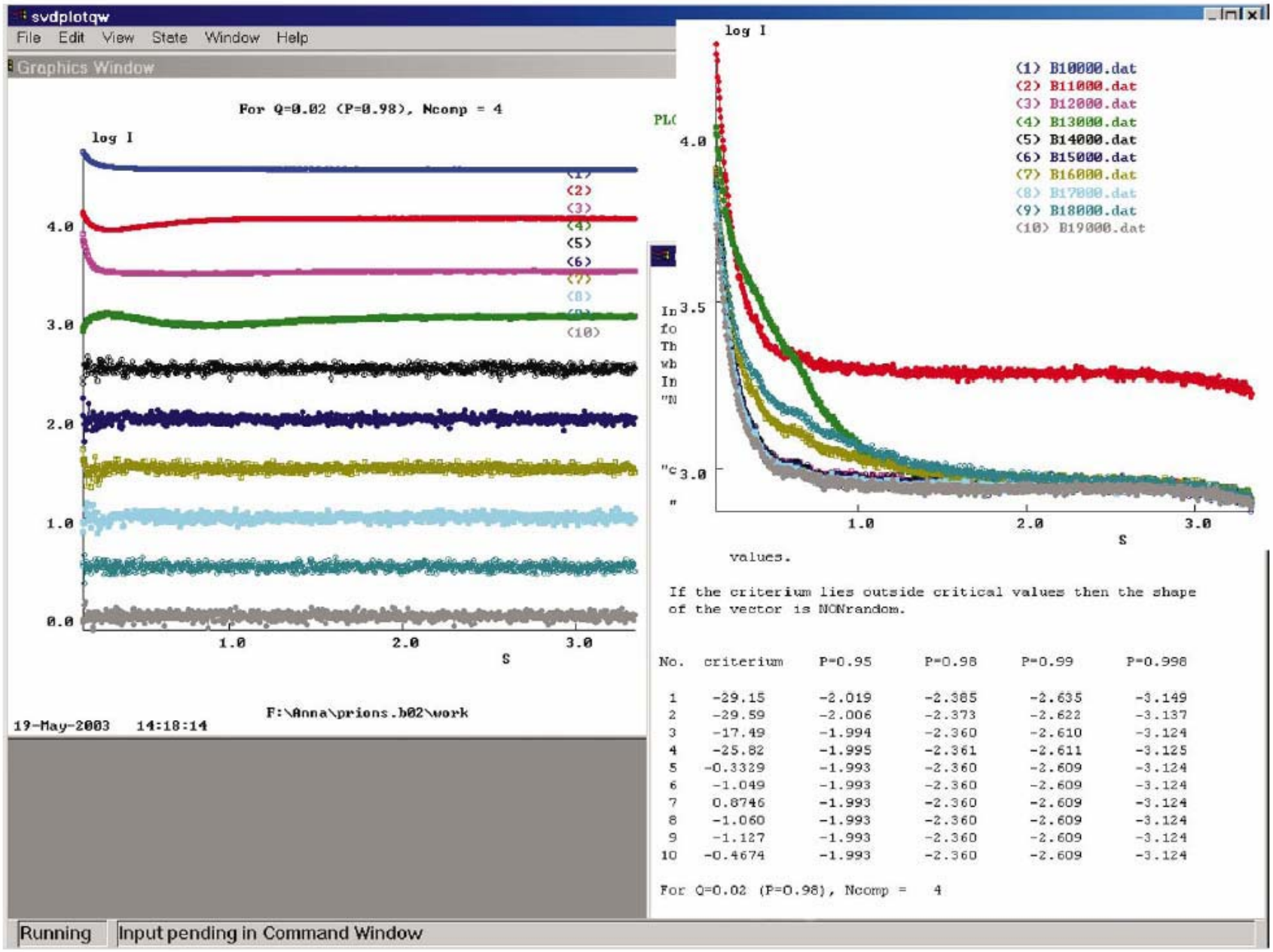




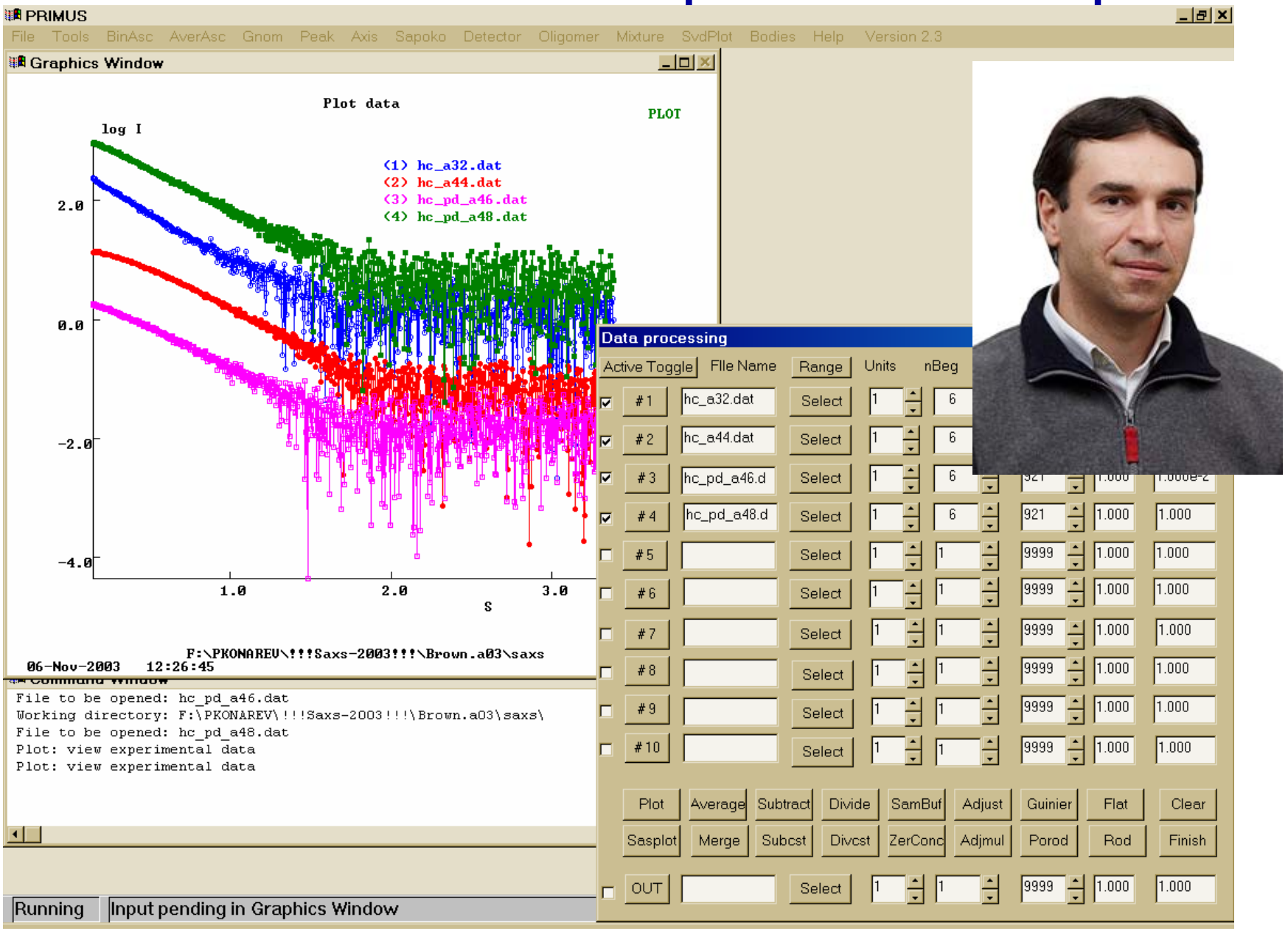
Equimolar mixture of Fd:Fd-Glts 1:1 and free Fd

R.H.H. van den Heuvel, D.I. Svergun, M.V. Petoukhov, A. Coda, B. Curti, S. Ravasio, M.A. Vanoni & A. Mattevi (2003). *J. Mol. Biol.* **330**, 113-128

SVD Analysis



SVD: Number of independent components



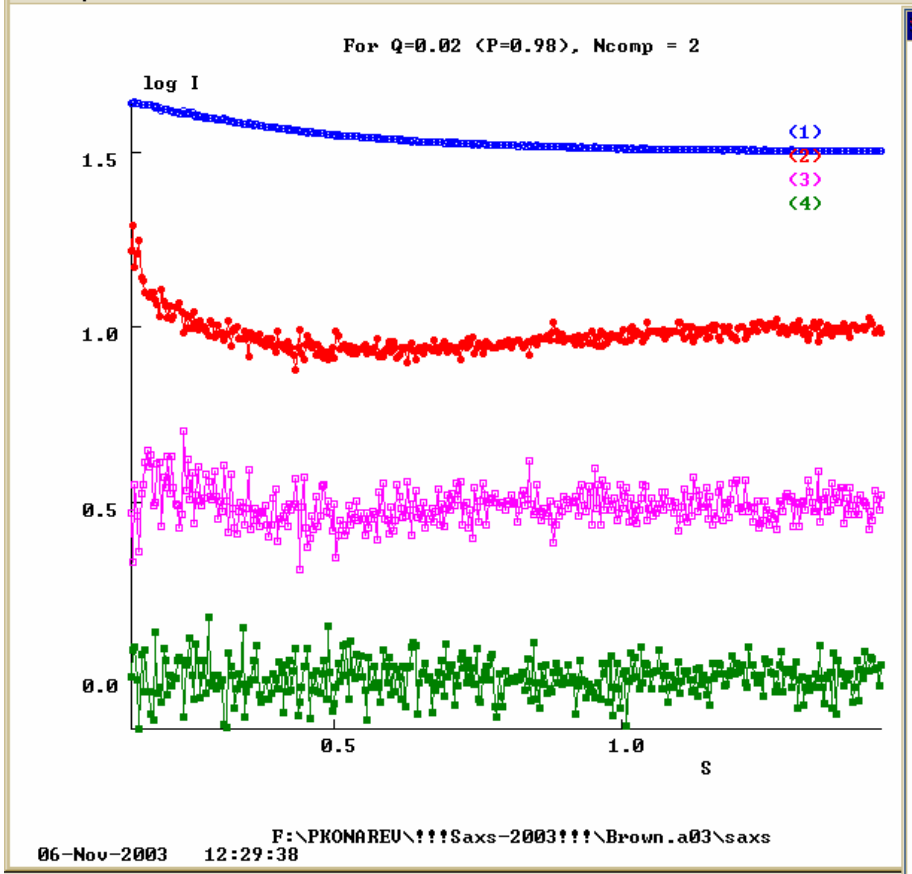
Mixture of monomers and dimers

SVDplot in “old” Primus

svdplotqw

File Edit View State Window Help

Graphics Window



Command Window

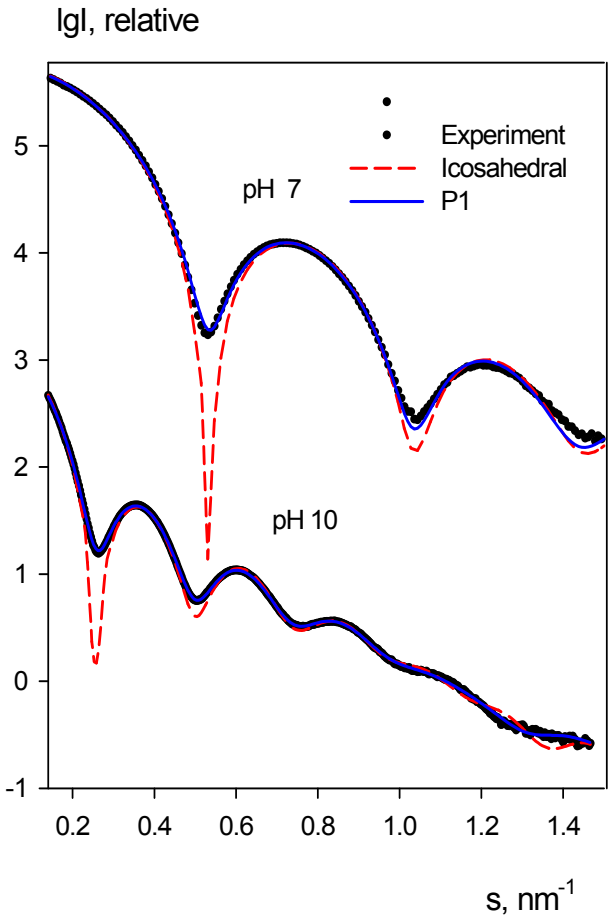
```
        Non-parametric test for the number of components.

In this test shapes of left-side singular vectors are tested
for randomness of sequencing positive and negative elements.
The number of components equals the number of singular vectors
which shape is nonrandom (non-noisy).
In this table:
"No.":
        ordinal number of the singular vector. All the vectors
        are arranged correspondingly to singular values of the
        data matrix arranged in decreasing order.
"criterium":
        testing statistics.
 "P":
        confidence level. In the corresponding columns two
        numbers (columnwise pairs) are left and right critical
        values.

 If the criterium lies outside critical values then the shape
 of the vector is NONrandom.


No.  criterium    P=0.95      P=0.98      P=0.99      P=0.998

 1    -18.92      -2.017      -2.384      -2.633      -3.148
 2    -12.11      -2.013      -2.379      -2.628      -3.143
 3    -1.185      -2.012      -2.379      -2.628      -3.142
 4    0.9518      -2.012      -2.378      -2.628      -3.142

For Q=0.02 (P=0.98), Ncomp =   2
Press to continue
```

NCOMP = 2

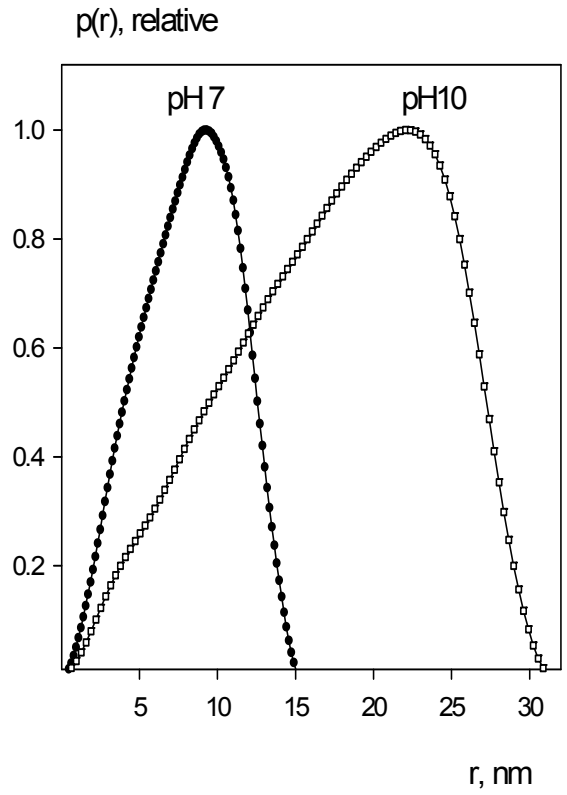
Running | Input pending in Command Window

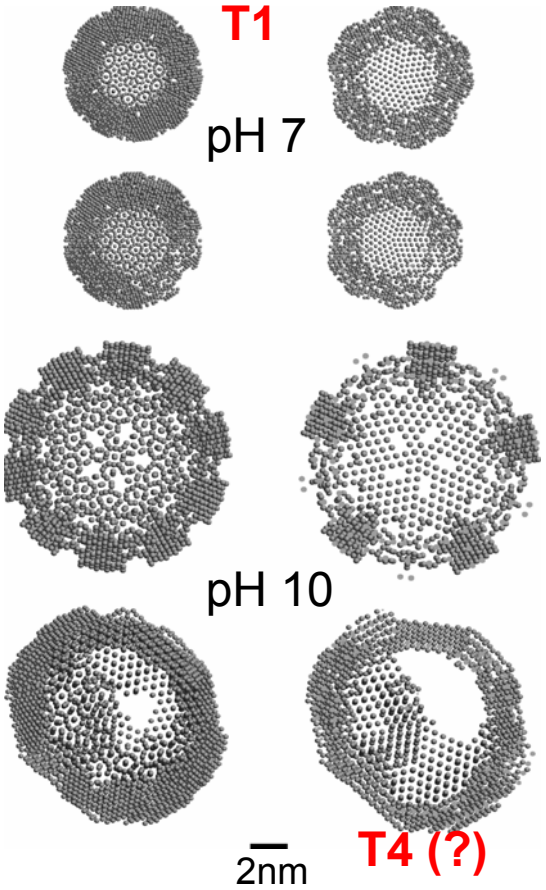
Molecular Assembly of Lumazine Synthase

LS catalyzes the formation of 6,7-dimethyl-8-ribityllumazine in the penultimate step of riboflavin biosynthesis. Depending on the buffer it forms pentamers, dimers of pentamers and icosahedral capsids



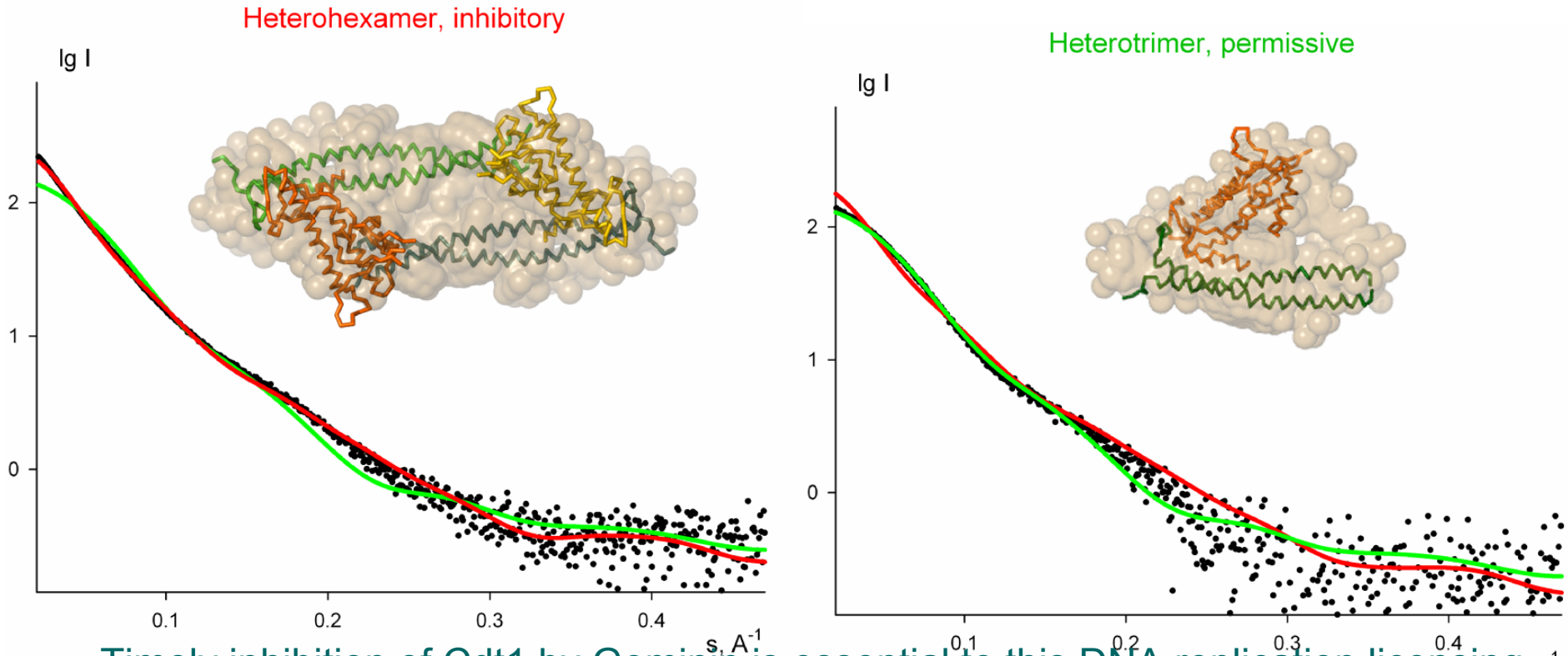
Wild type LS in borate buffer





Zhang, X., Konarev, P.V., Petoukhov, M.V., Svergun, D.I., Xing, L., Cheng, R.H., Haase, I., Fischer, M., Bacher, A., Ladenstein, R. & Meining, W. (2006) *J Mol Biol*. **362**, 753-770

Quaternary structure of the human Cdt1-Geminin complex regulates DNA replication licensing

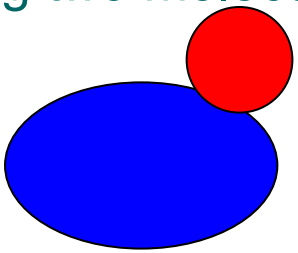


- Timely inhibition of Cdt1 by Geminin is essential to this DNA replication licensing
- The mechanism of DNA licensing inhibition by Geminin, is analyzed by combining MX, SAXS and functional studies.
- The Cdt1:Geminin complex can exist in two distinct forms, a ‘‘permissive’’ heterotrimer and an ‘‘inhibitory’’ heterohexamer

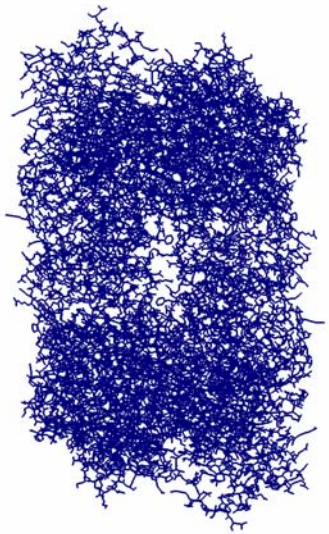
V. De Marco, P. J. Gillespie, A. Lib, N. Karantzelis, E. Christodouloua, R. Klompmaker, S. van Gerwen, A. Fish, M. V. Petoukhov, M. S. Iliou, Z. Lygerou, R. H. Medema, J. J. Blow, D. I. Svergun, S. Taraviras & A. Perrakis (2009) PNAS USA, **106,** 19807

Word of Caution: Glutamate Synthase Story

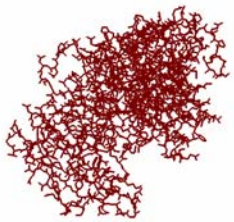
GltS is a complex iron-sulfur flavoprotein that catalyse the reductive transfer of L-glutamine (L-Gln) amide group to the C_2 carbon of 2-oxoglutarate (2-OG), yielding two molecules of L-glutamate (L-Glu)



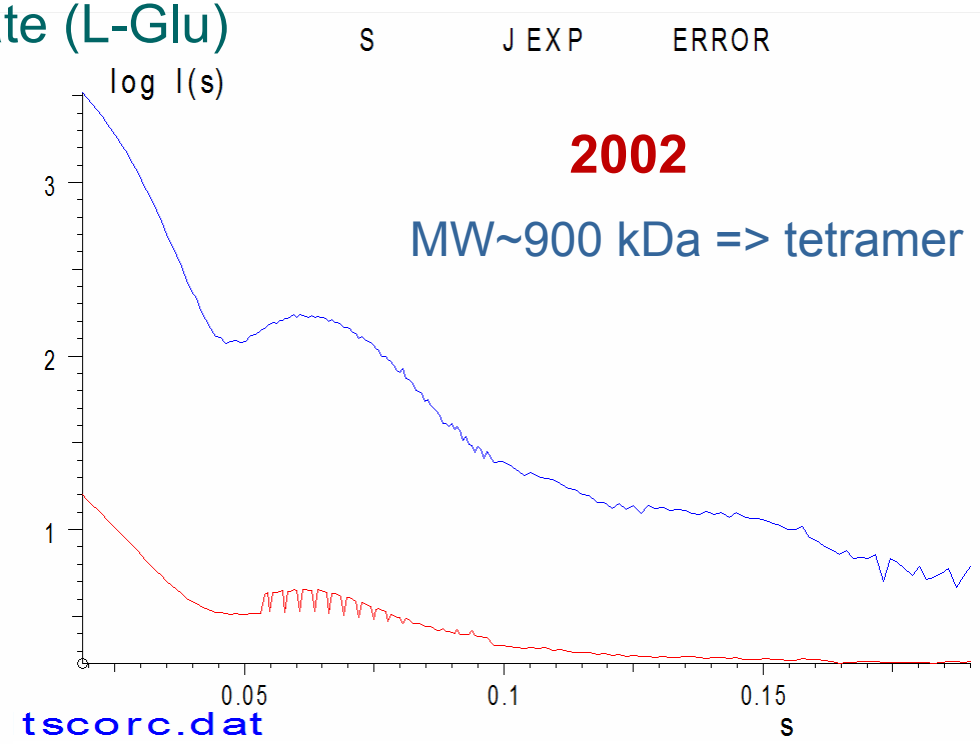
αβ protomer: 160+50=210 kDa



α Xtal dimer



β homology model

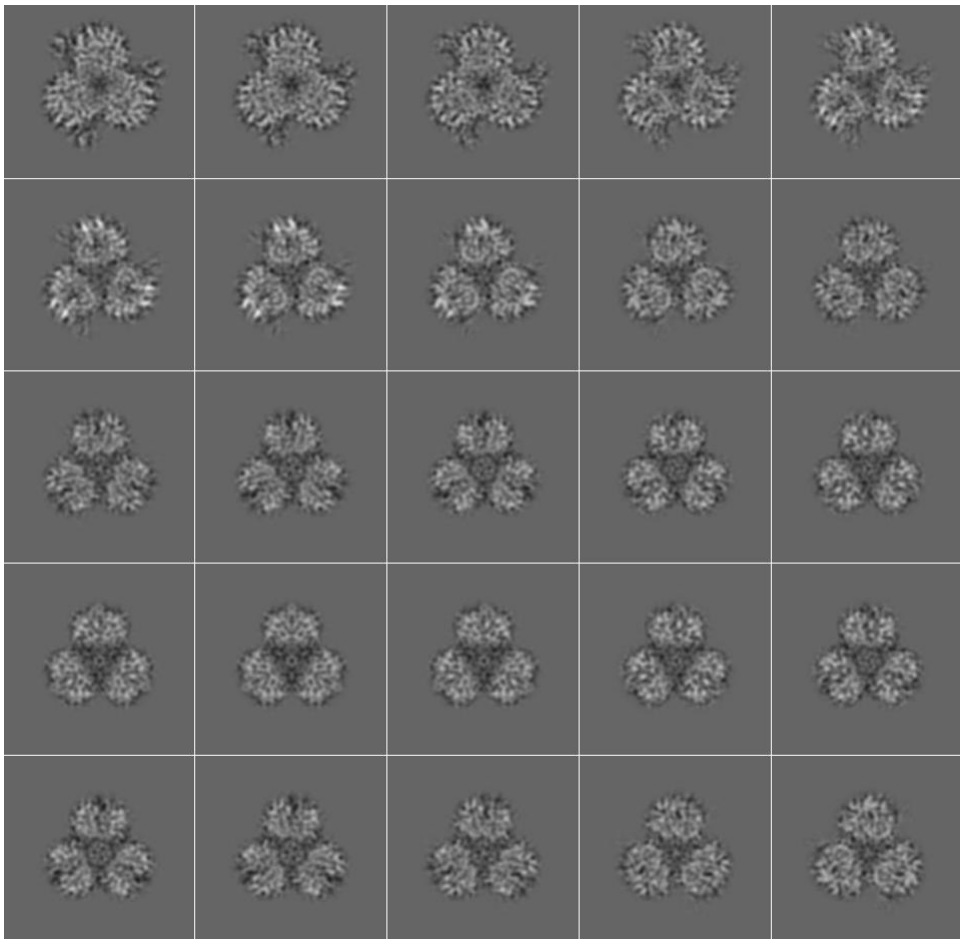


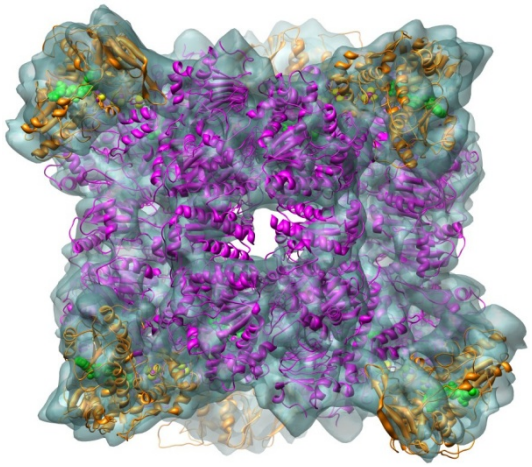
2002

MW~900 kDa => tetramer

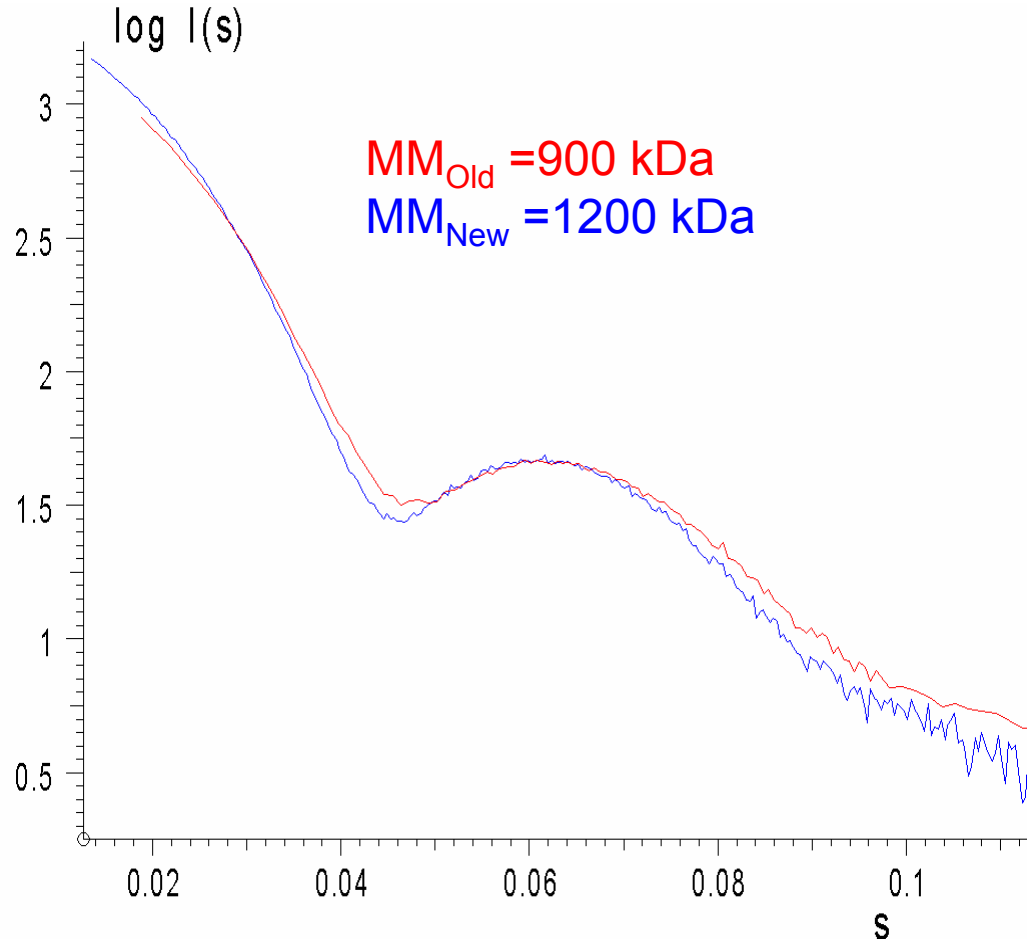
Collaboration: M.A. Vanoni (Milano Univ.)

Glutamate synthase: EM model

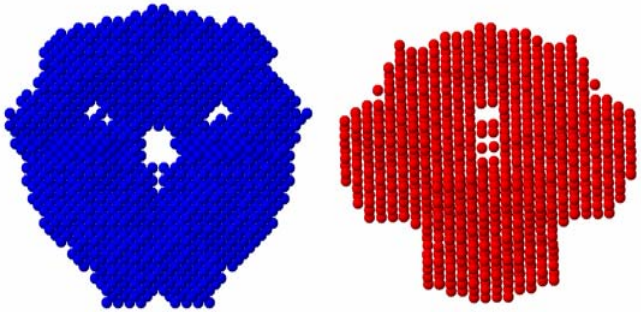


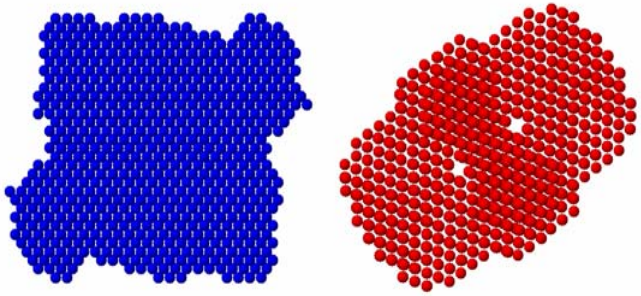


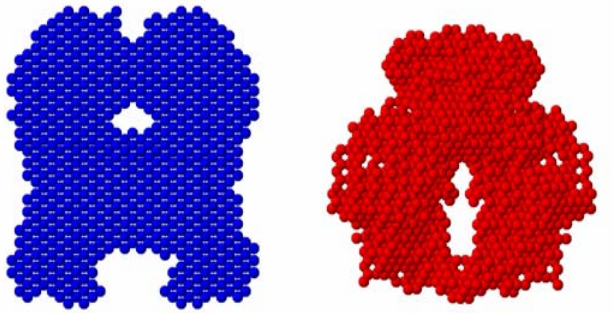
Comparison of the “Old” and “New” data



EM hexamer Old SAXS tetramer







Glutamate synthase: salt / substrate dependence

OLIGOMER with EM model

GltS => 90%$(\alpha\beta)_6$ +10% αβ

GltS+NaCl => 100% αβ

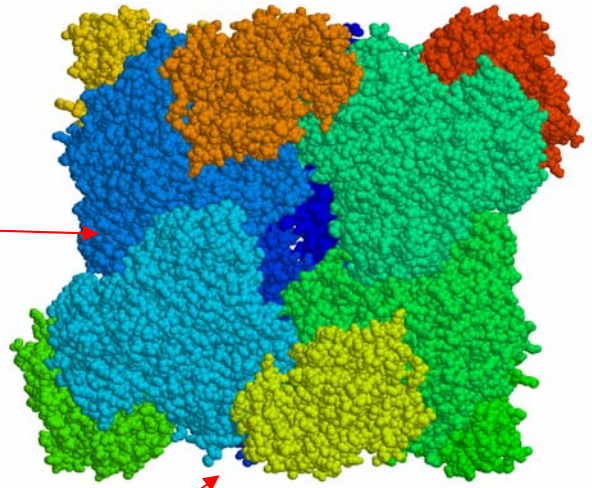
GltS+2OG => 90%$(\alpha\beta)_6$ +10% αβ

GltS+L-MetS => 90%$(\alpha\beta)_6$ +10% αβ

GltS+L-MetS+2OG => 90%$(\alpha\beta)_6$ +10% αβ

GltS+NADP=> 70%$(\alpha\beta)_6$ +20%$(\alpha\beta)_2$+10% αβ

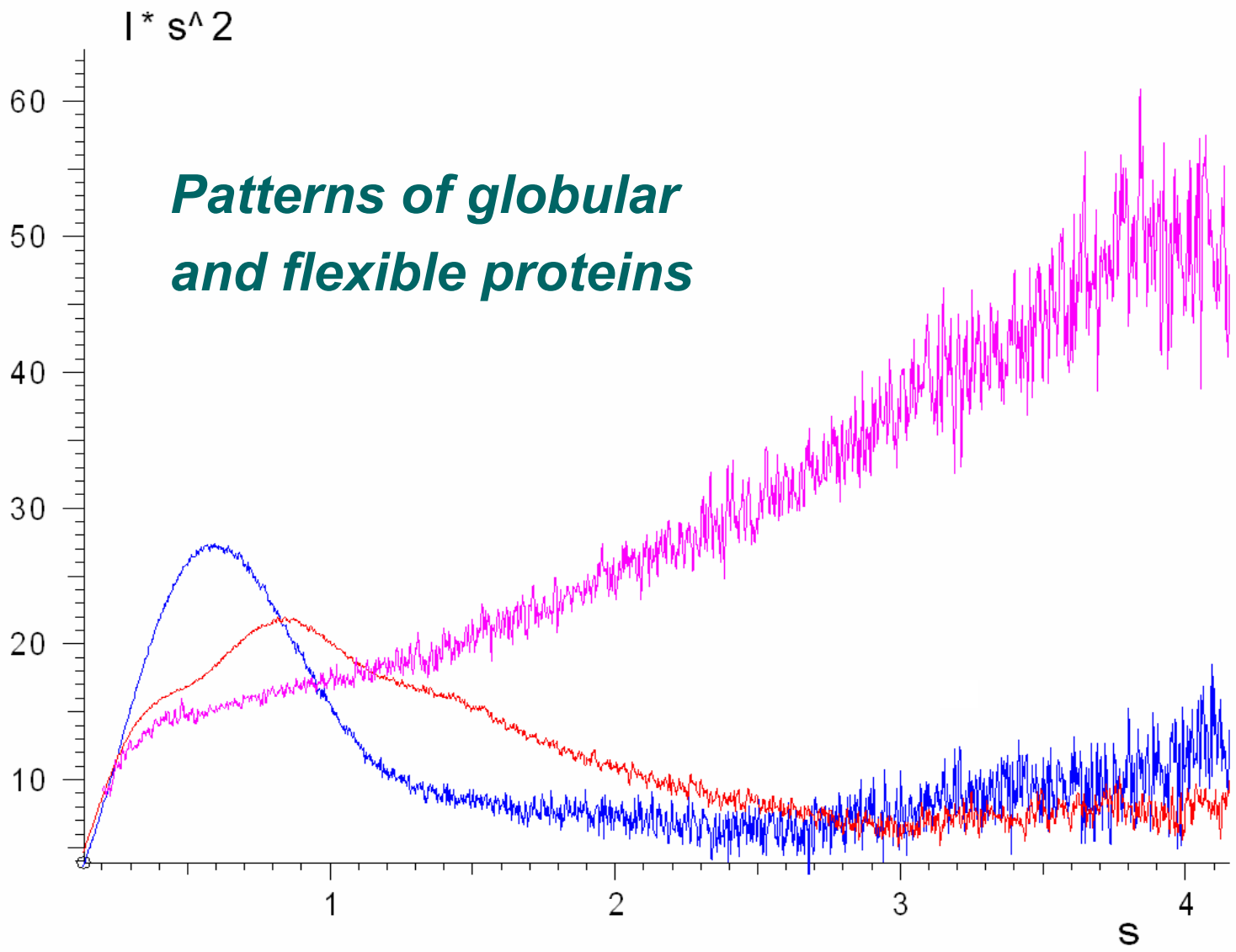
GltS+NADP+L-MetS+2OG=> 70%$(\alpha\beta)_6$ +20%$(\alpha\beta)_2$+10% αβ



In case of a mixture, SAS permits to study

- Oligomeric equilibrium
- Conformation changes
- Complex stoichiometry

Flexible systems



Unfolded

Folded

Multi-domains with flexible linkers

Kratky plot permits to detect disorder

Biomolecules are Dynamic Entities

- Local Protein Fluctuations: Backbone and side-chains

Protein vibrations, loop motions and breathing to facilitate interactions and catalysis

- Concerted domain motions: Linkers as Hinges

Specific and limited domain jumps between relative positions often linked with catalysis in on/off mechanisms

- Highly flexible regions or domains

An astronomical number of conformations are available.
Flexible multi-domain proteins (MD) and Intrinsically Disordered Proteins (IDPs)
Linked with signaling and regulation

Flexible Proteins: Why to Care?

► Many **biological functions** such as transcription, regulation, cell cycle control, **require extensive flexibility**

► Is more common in higher organisms that have to perform more and more controlled functions. **Disorder is correlated with complexity**

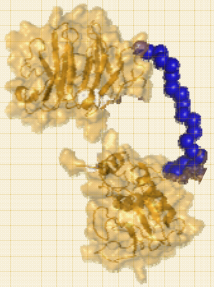
► High selectivity, moderate affinity, and promiscuity are properties often linked to flexibility

► In these systems partially structured conformations or transient long range interactions can be crucial for biological activity. Structural studies are important

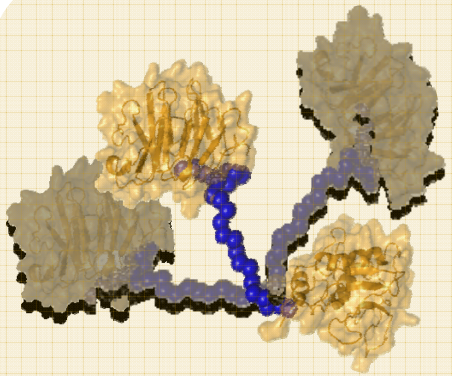
Indications (not Proofs!) of Flexibility

- Smooth Scattering profiles and featureless Kratky Plots
- Large R_g and D_{max}
- Absence of correlation peaks in the *p(r)* function
- Low correlation densities in *ab initio* reconstructions
- Isolated domains in rigid body modelling
- Prediction of disorder using bioinformatics

Detection of Flexibility



Rigid Scenario



Flexible Scenario

SAXS curves

Analysis of the overall size descriptors (R_g, *p(r)*, Kratky)

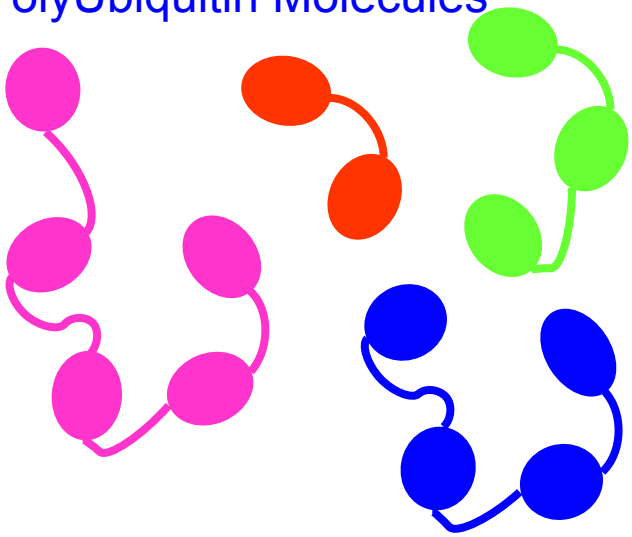
Go for flexibility!

Modelling: *ab initio* (DAMMIN/DAMMIF) and Rigid body (BUNCH/CORAL)

Analysis of the differences

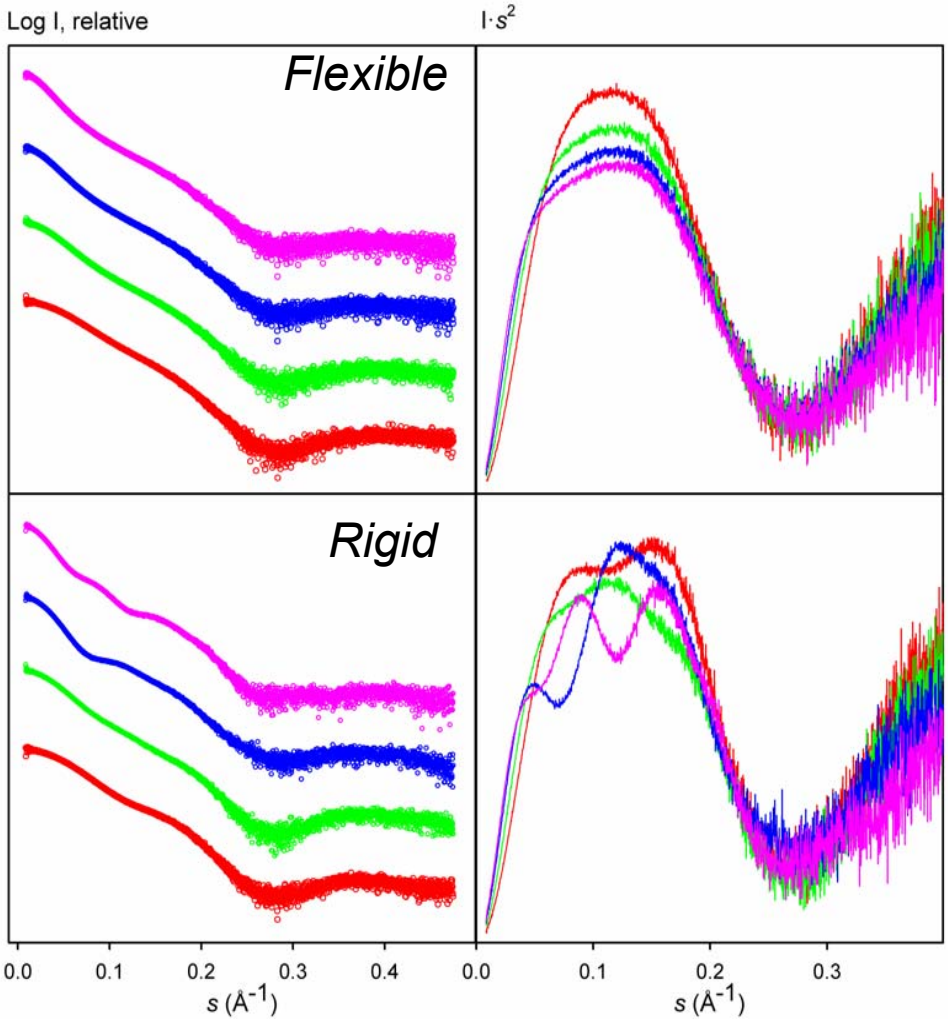
Detection of Flexibility

PolyUbiquitin Molecules



2,3,4 and 5 Ubiquitin (72 AA) domains connected by 20 AA linker (RanCH)

Flexible Multidomain Proteins present less features than rigid counterparts

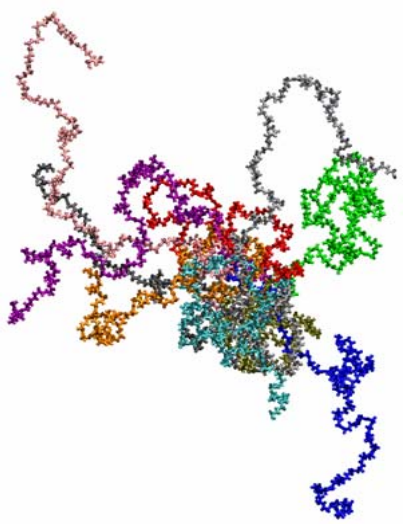


Bernadó *Eur. Biophys. J.* 2009, 39, 769

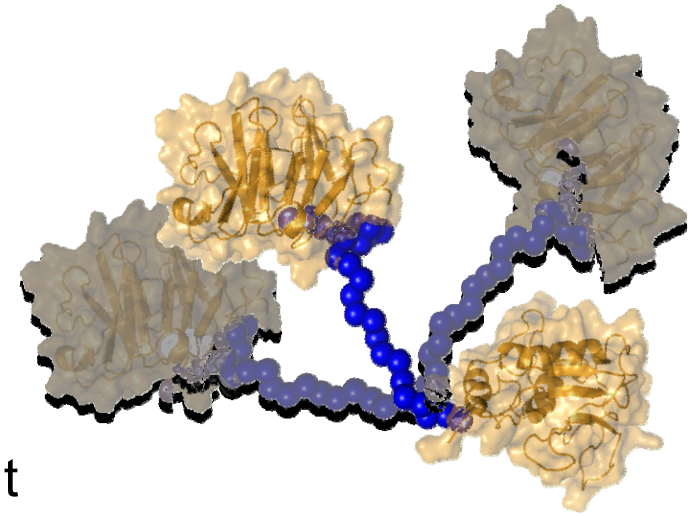
Flexibility as a mix of different conformations

$$I(s) = 4\pi \int_0^D p(r) \frac{\sin sr}{sr} dr$$

For monodisperse systems the scattering is proportional to that of a single particle averaged over all orientations



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



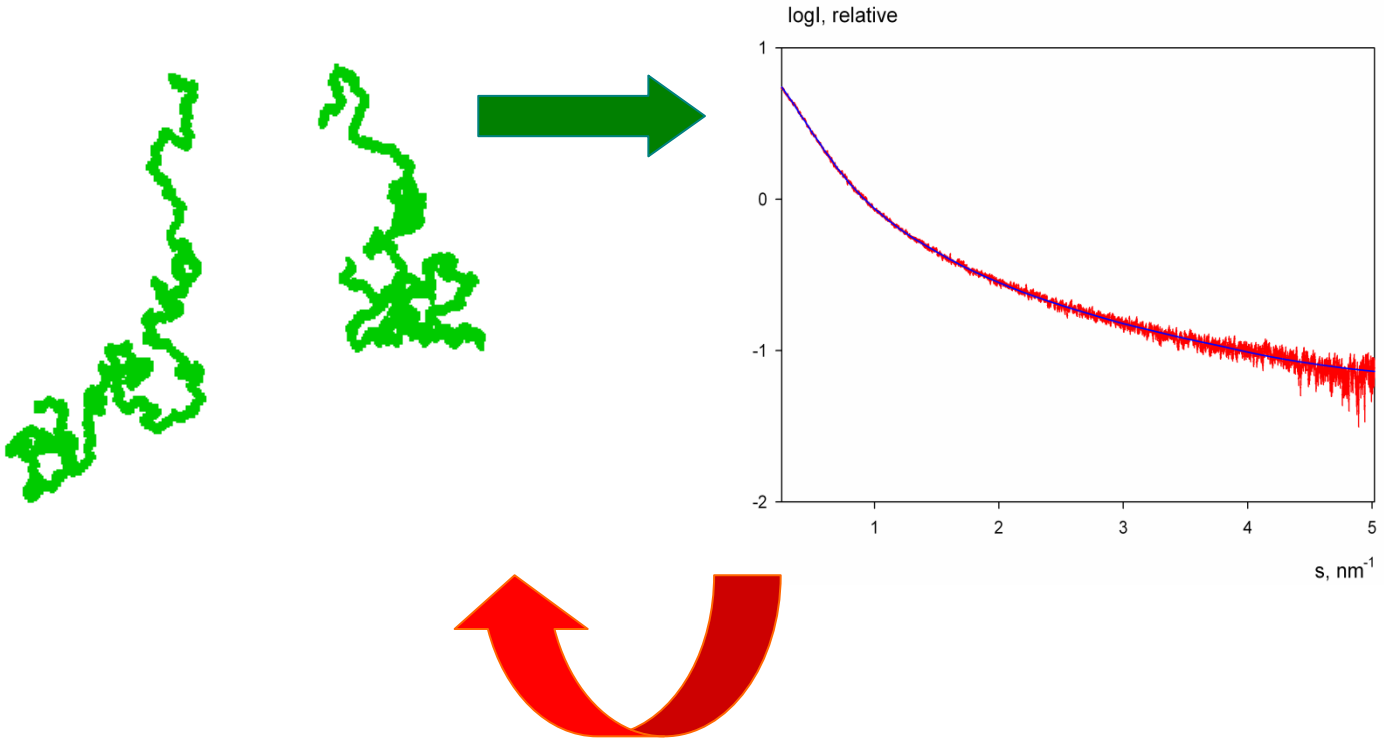
ν_k = volume fraction

$I_k(s)$ = scattering intensity from the k-th component

Ensemble Optimization Method

structures

curve



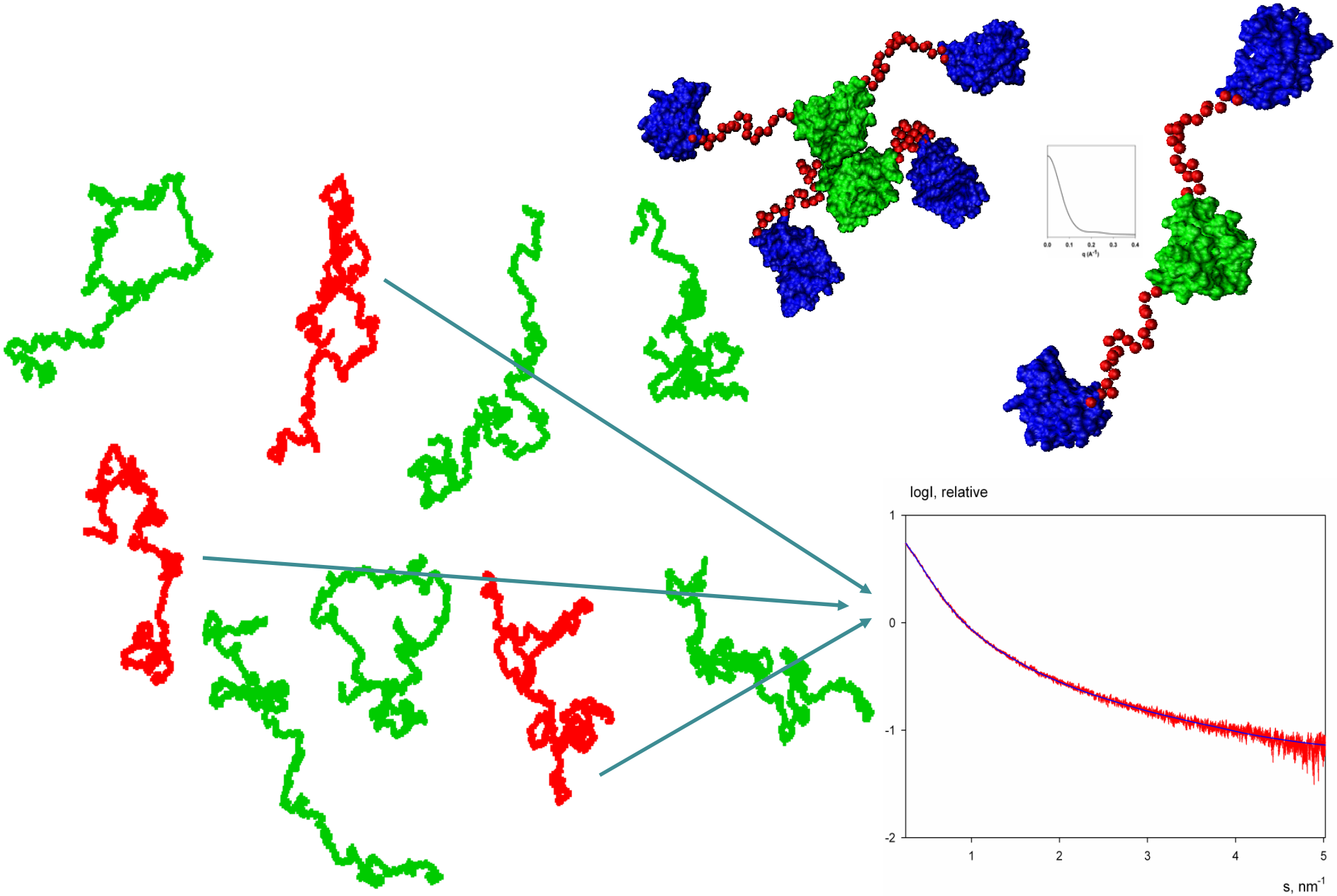
Instead: find an ensemble with the same characteristics as the sample

Ensemble optimization method for structural characterization of flexible proteins

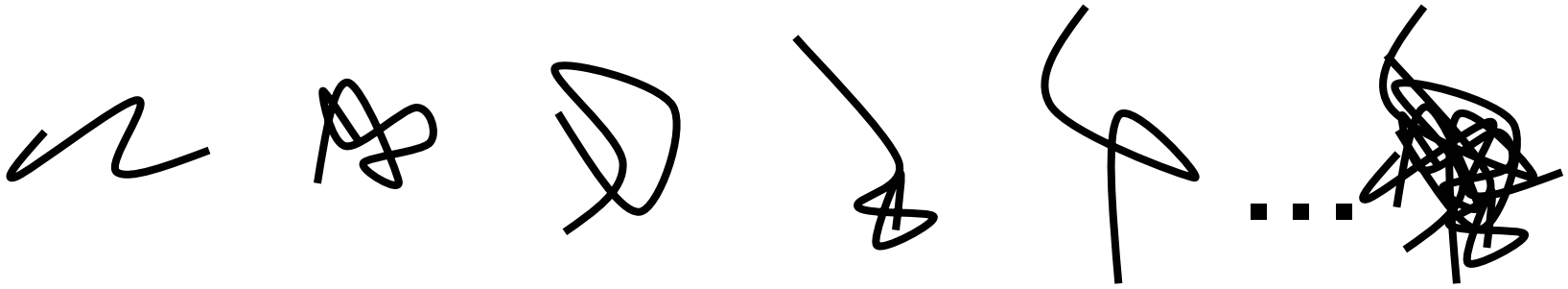
- When?
  - SAS data cannot be described by a single conformation
  - To assess the flexibility
- How?
  - Assume that a representative ensemble exists that fits the data
  - Large number of random structures is created
  - From the full ensemble of structures select the ones that together fit the scattering curve

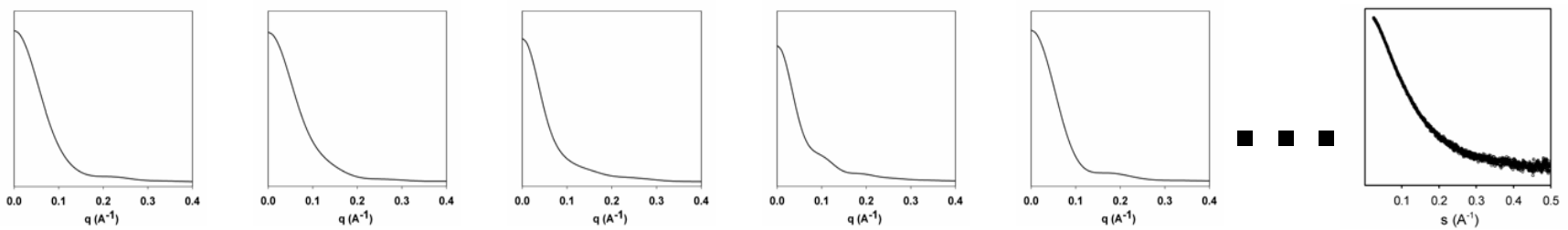
Bernado, P., E. Mylonas, M.V. Petoukhov, M. Blackledge & D.I. Svergun (2007) *JACS.* **129**, 5656-64

Ensemble Optimization Method

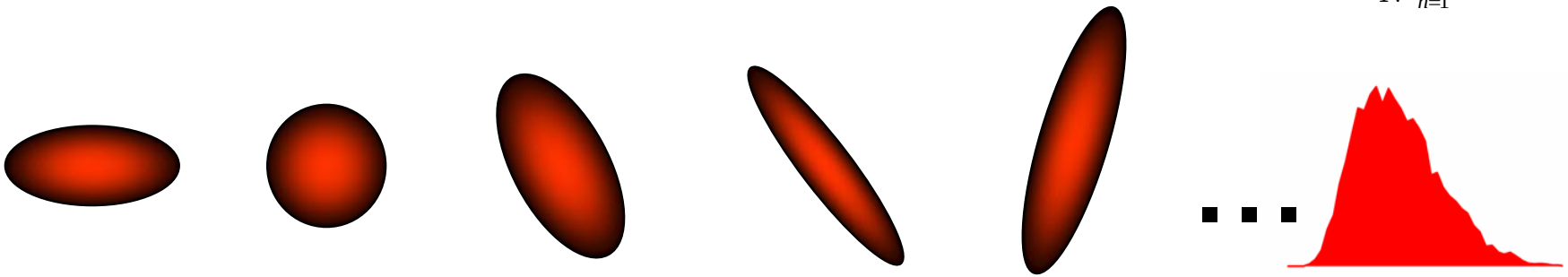


Ensemble methods in SAXS





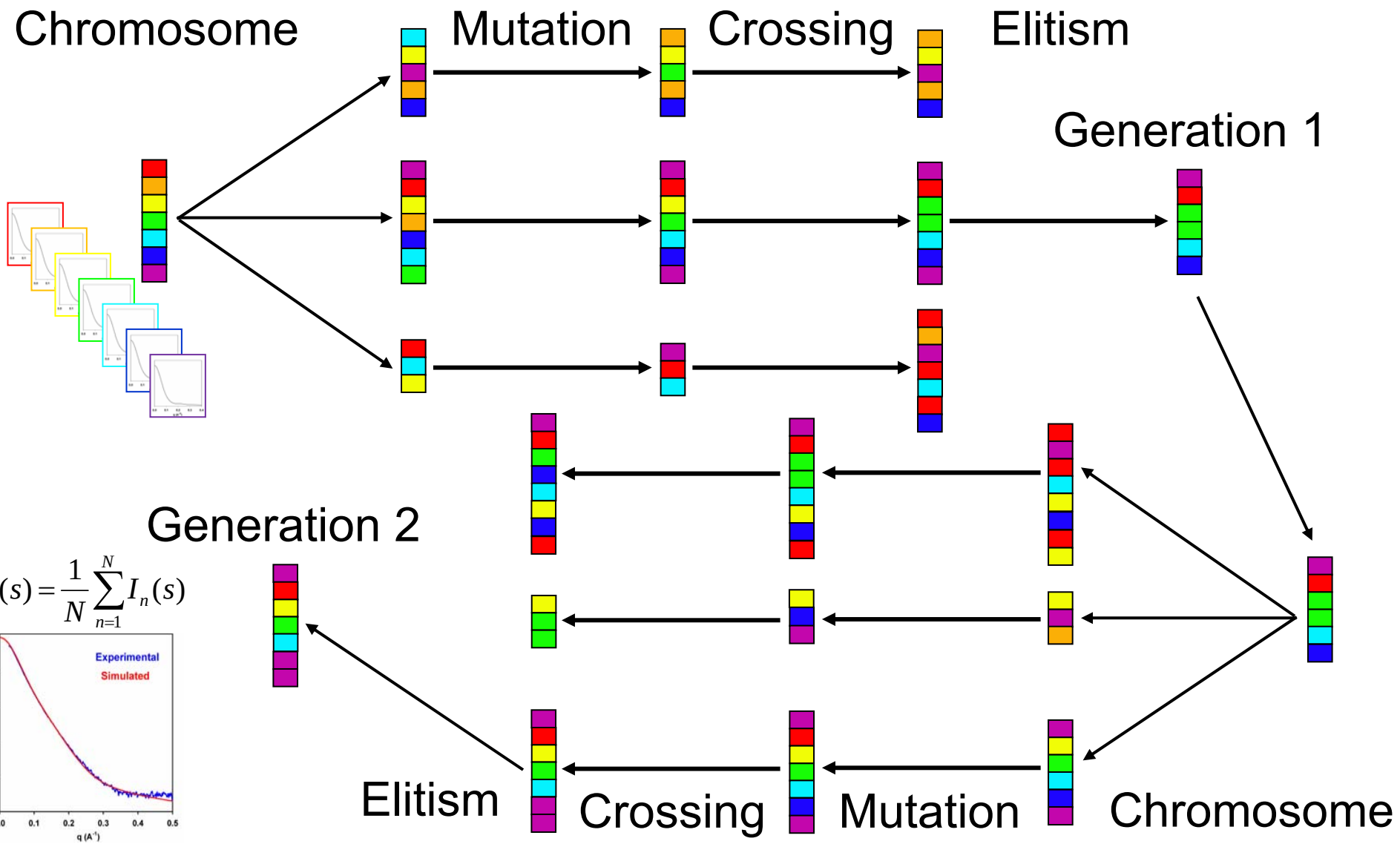
$$I(s) = \frac{1}{N}\sum_{n=1}^{N} I_n(s)$$

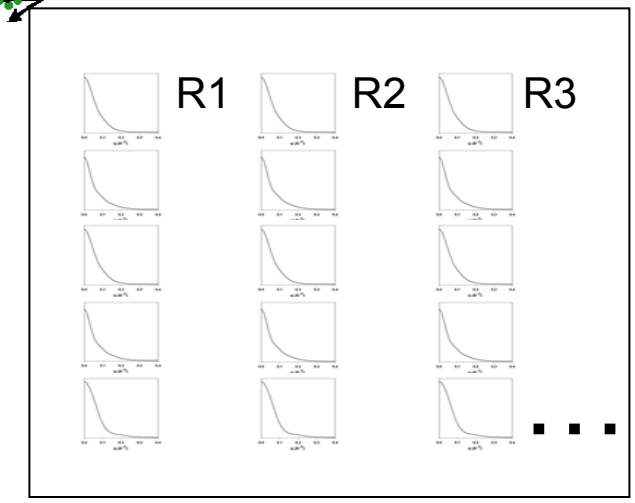


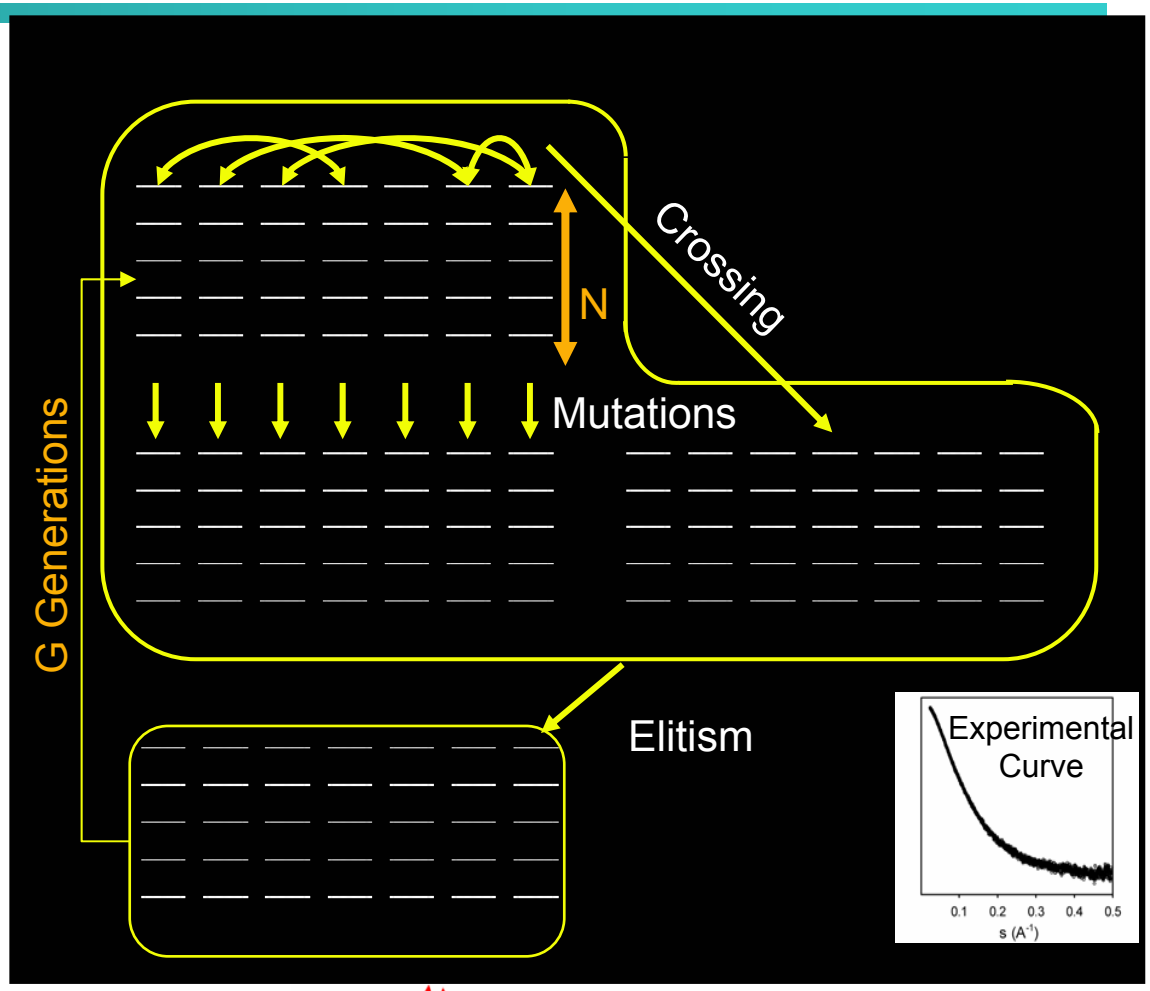
R_g^1 R_g^2 R_g^3 R_g^4 R_g^5

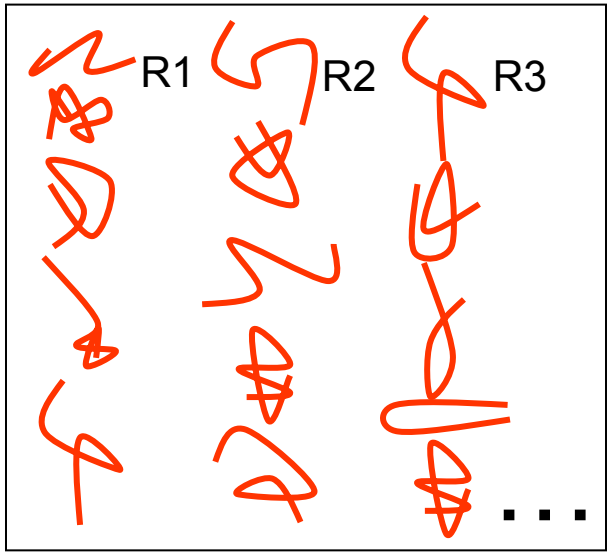
$\rho\ (R_g)$

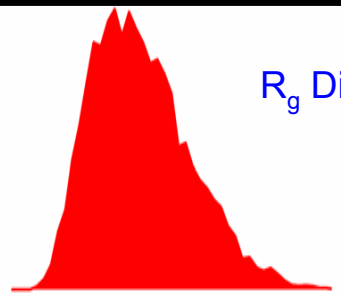
Genetic Algorithm *(optimized ensemble size)*











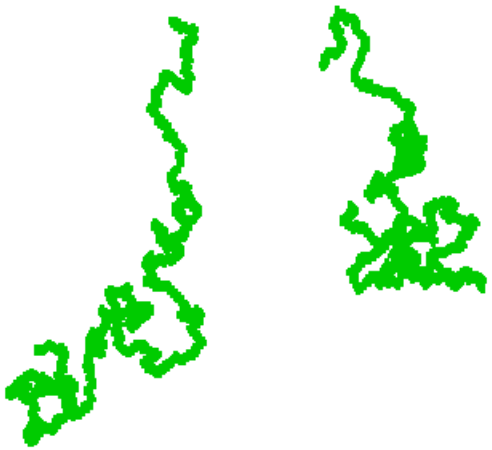
R_g Distribution

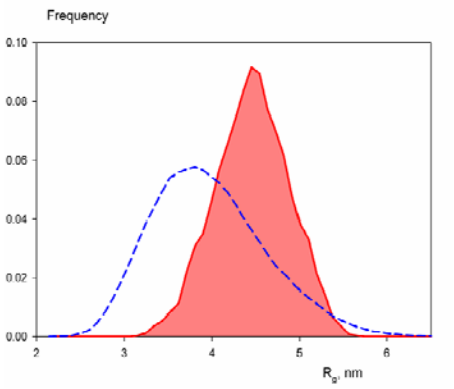
Ensemble Optimization Method (EOM)

Bernadó, Mylonas, Petoukhov, Blackledge, and Svergun.
J Am Chem Soc 2007, **129**:5656-64.

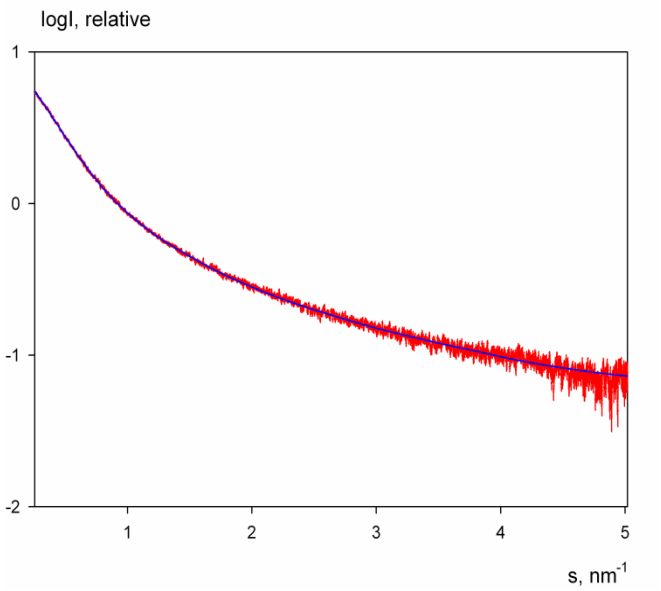
Ensemble Optimization Method

actual structures

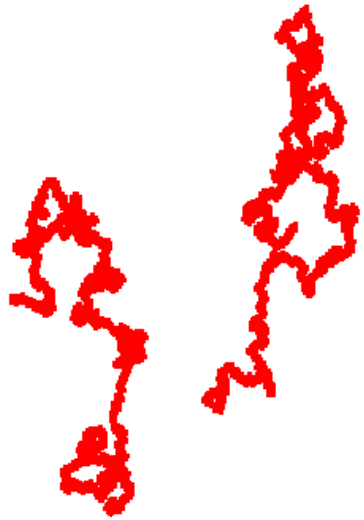


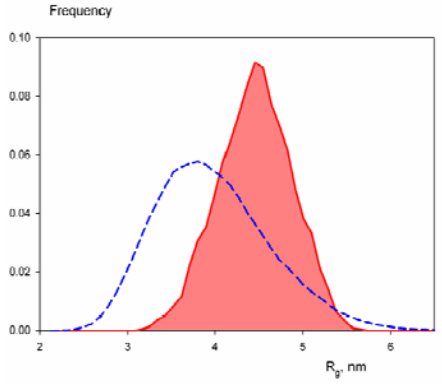


curve

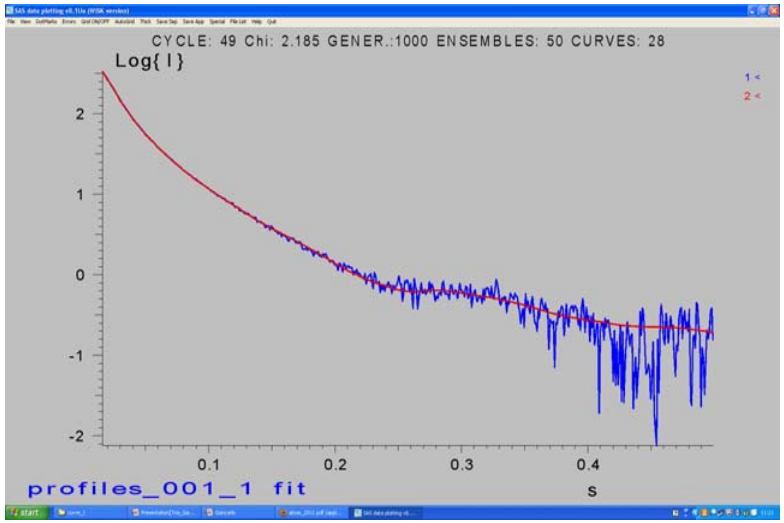


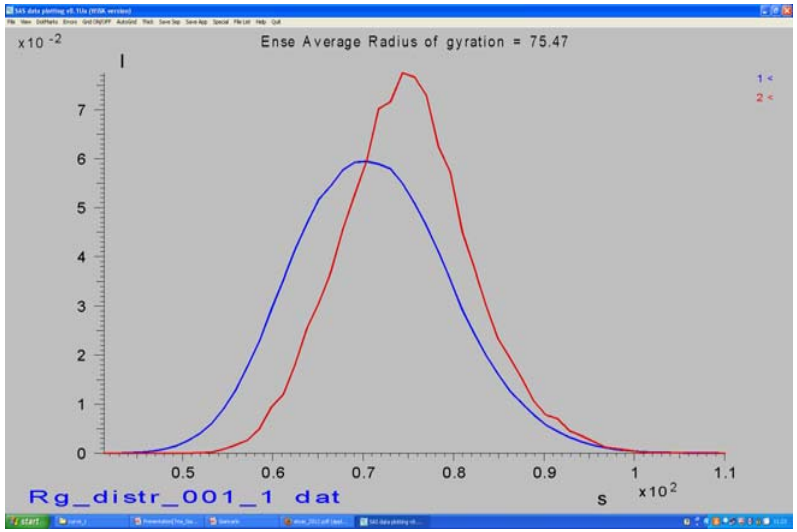
Optimized ensemble

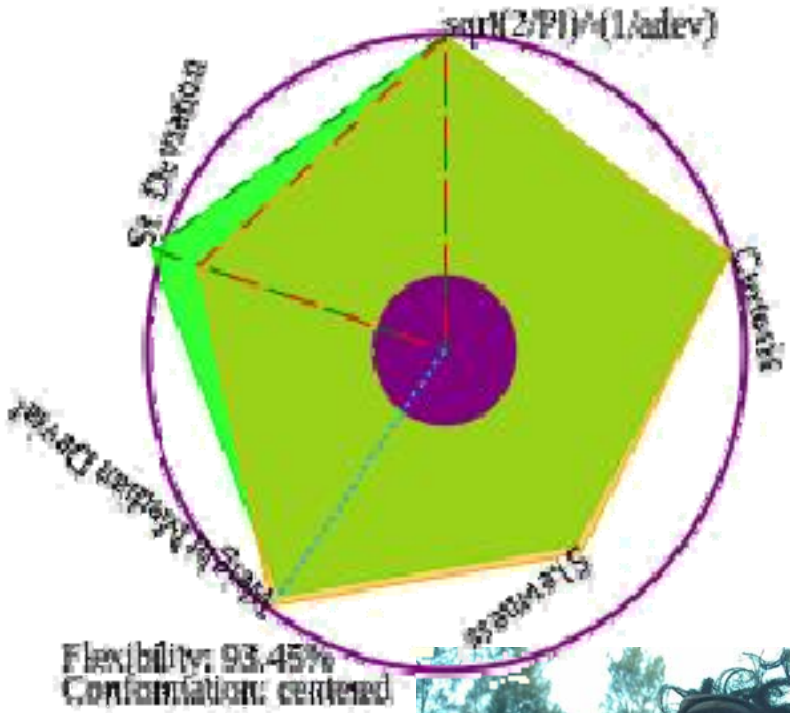




… the beauty of the *Pentagon*

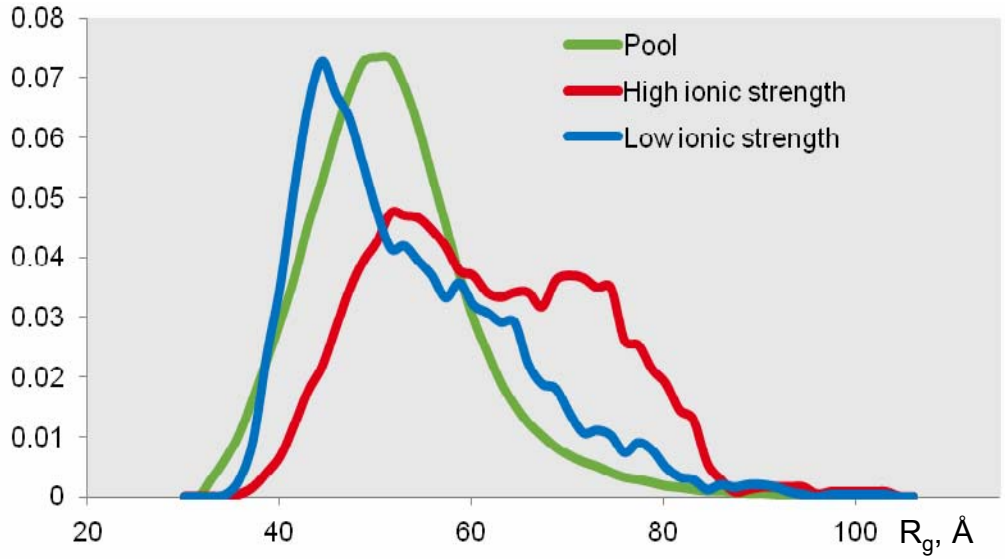


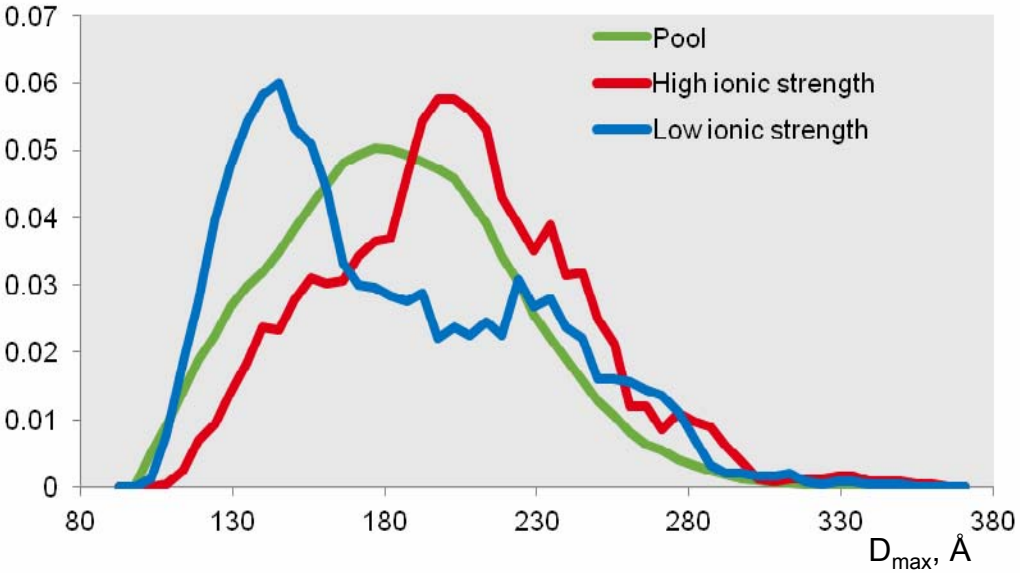


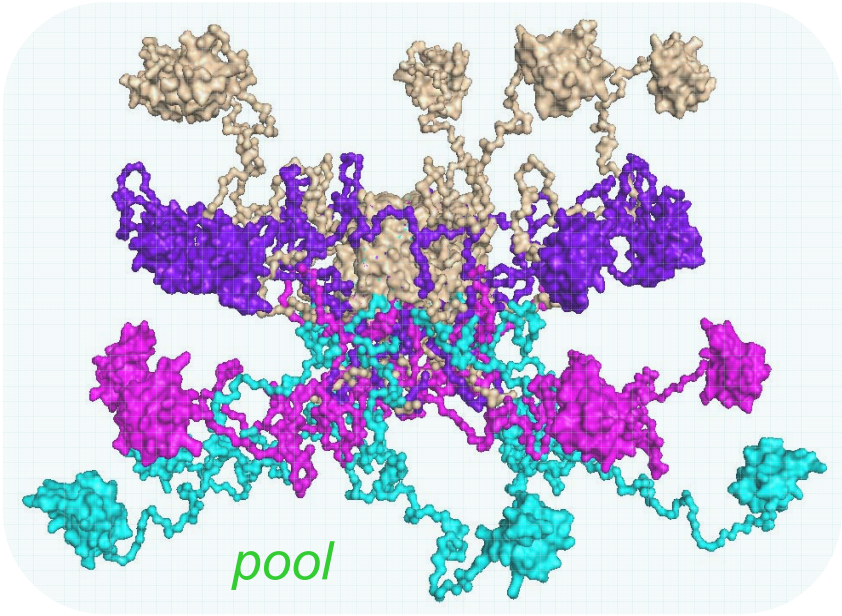


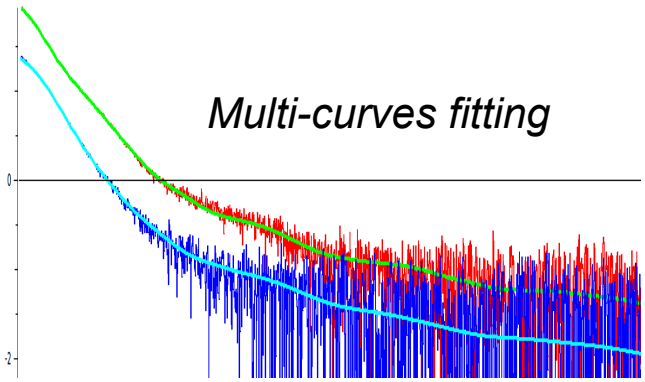
EOM 2.0 can handle multimeric assemblies

(full length protein measured in two buffers with low and high ionic strength respectively)

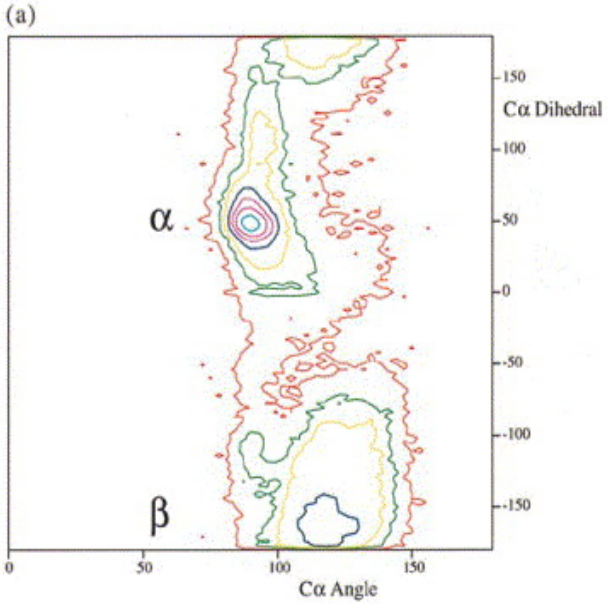




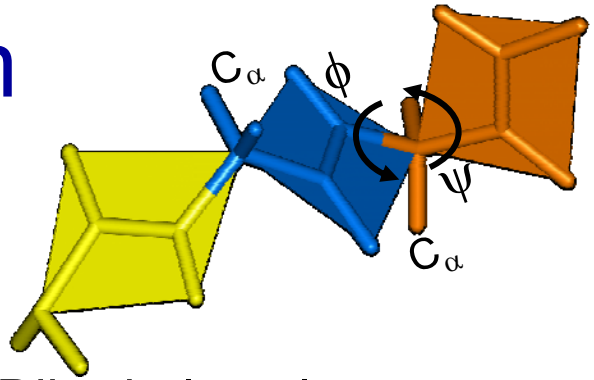




Modelling: Native vs. Random

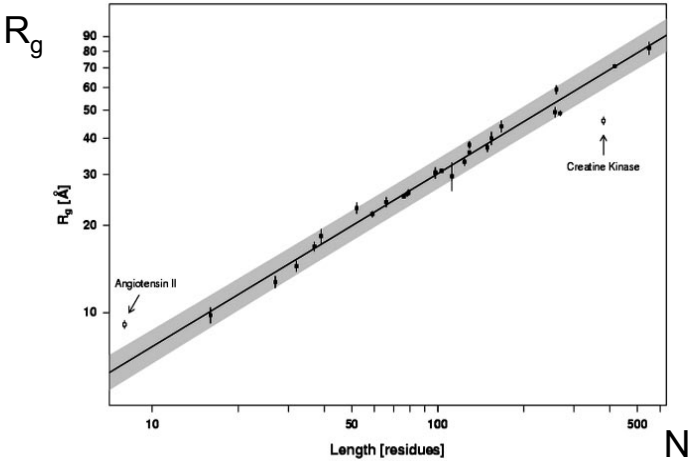


Quasi C_α-C_α Ramachandran plot



Bond angles vs. Dihedral angles

G. Kleywegt , *Validation of protein models from C_α coordinates alone*, *JMB*, 1997, 273, 371-376



Theoretical distribution of the bond and dihedral angles for random chains

$$R_g = R_0 \cdot N^{\nu}$$

R_0 Persistence Length

ν Solvent 'quality'

Several experimental and theoretical studies establish $\nu \approx 0.588$ as an indication of the 'random coil' in chemically denatured (Urea or GuHCl) proteins.

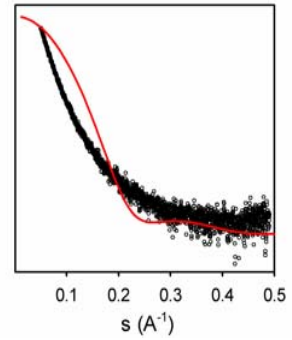
Kohn et al. *PNAS*, 2004, 101, 12491

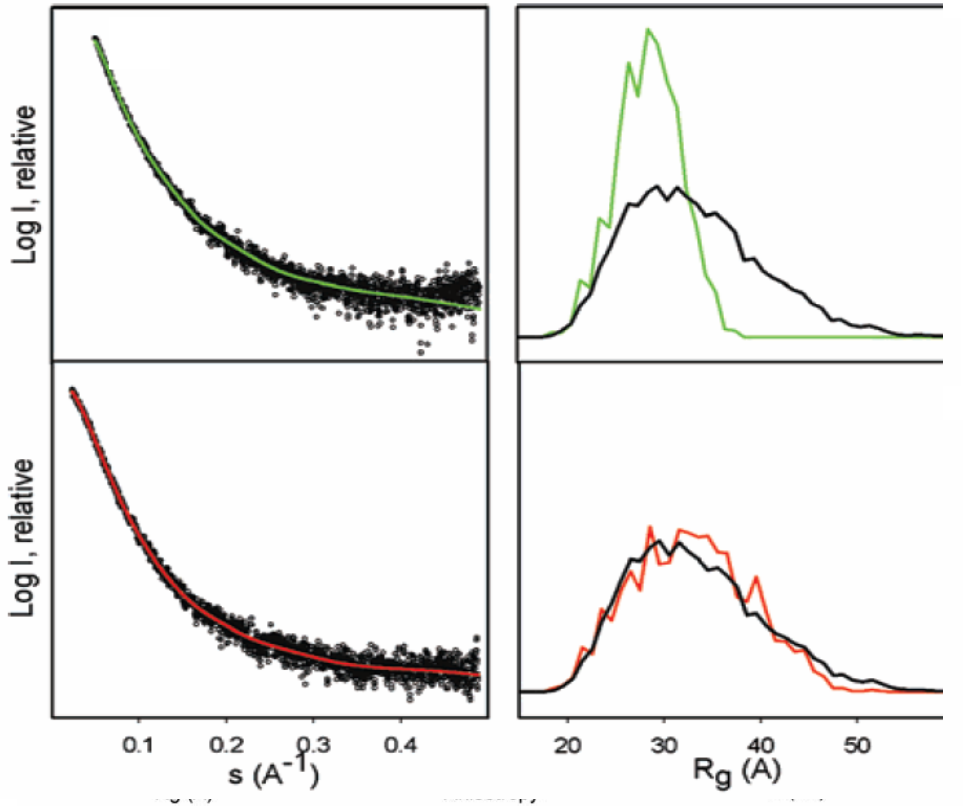
Assessing conformation variability: lysozyme unfolding

8M Urea, 10 mM DTT

R_g = 26.3 Å

R_g (native) = 15.1 Å



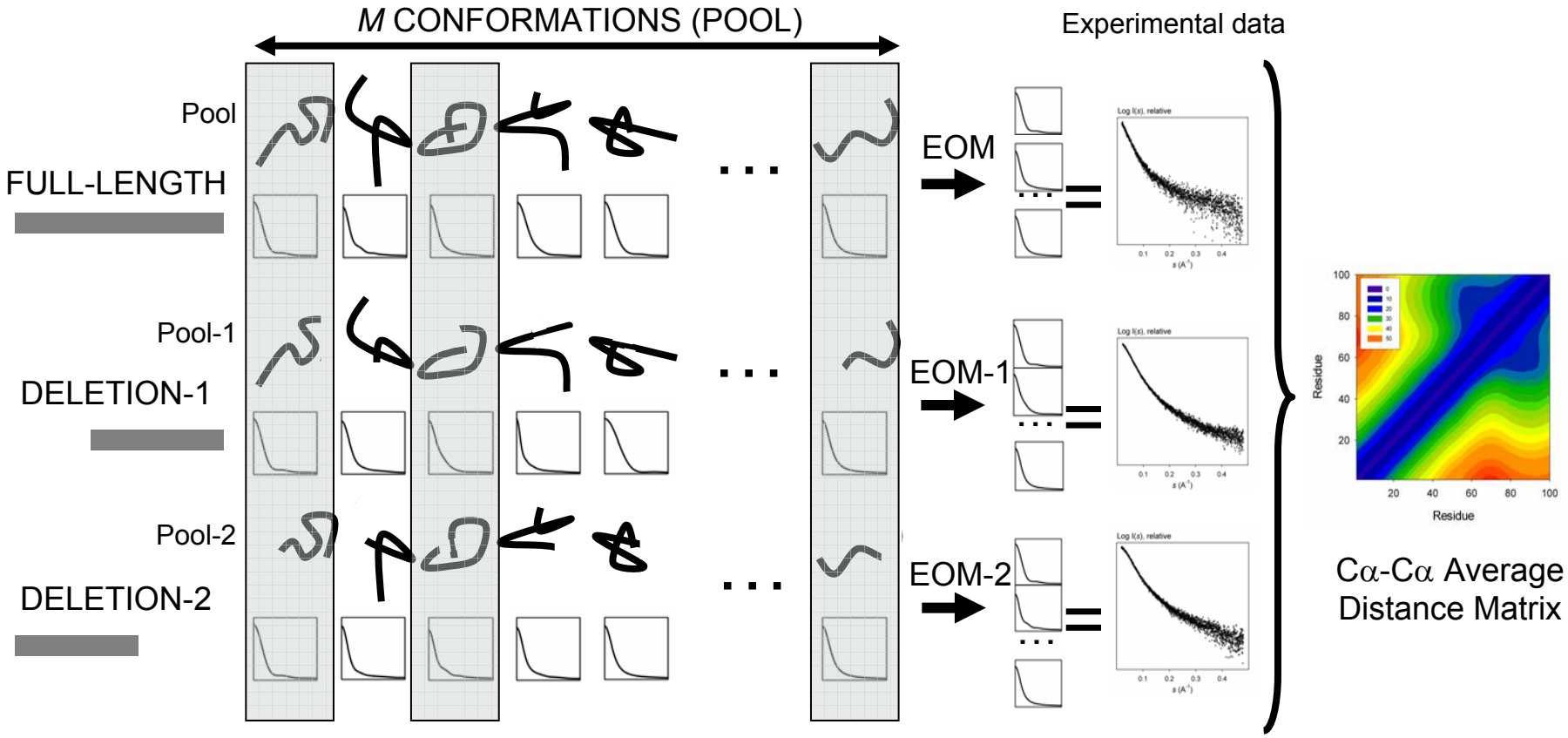


8M Urea, 10 mM DTT

8M Urea, 100 mM DTT

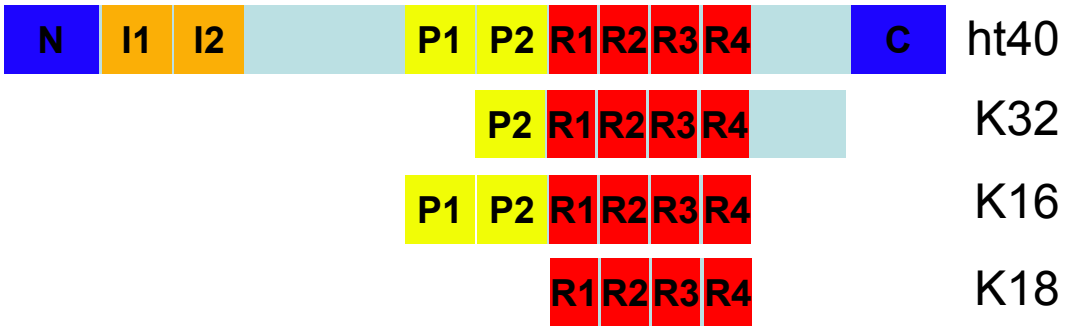
R_g = 30.0 Å

Multiple Curve Fitting with EOM: Reaching «Higher Resolution»

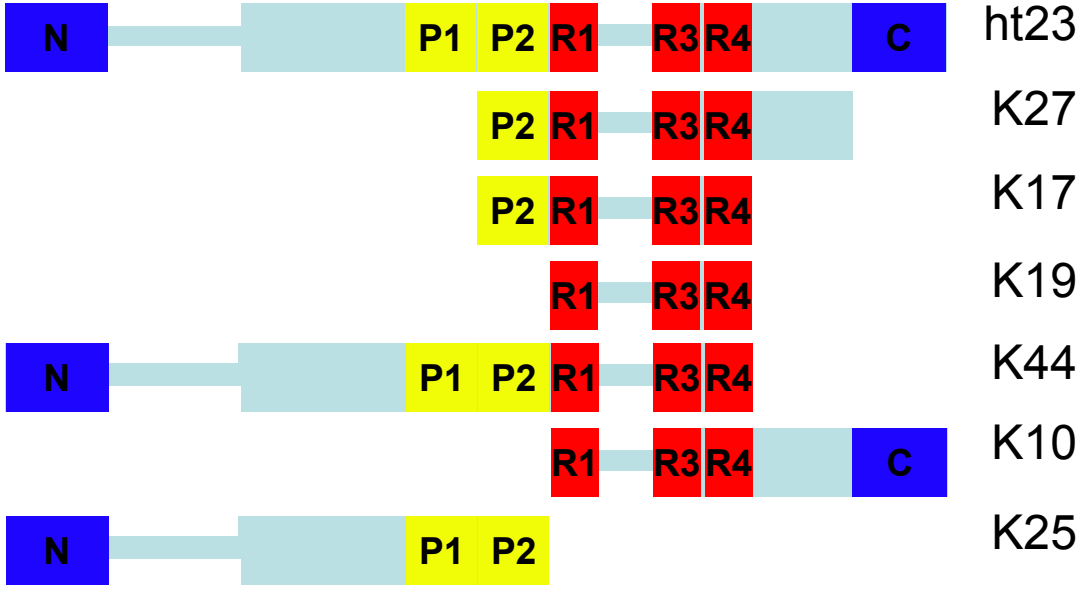


Application to Tau Protein

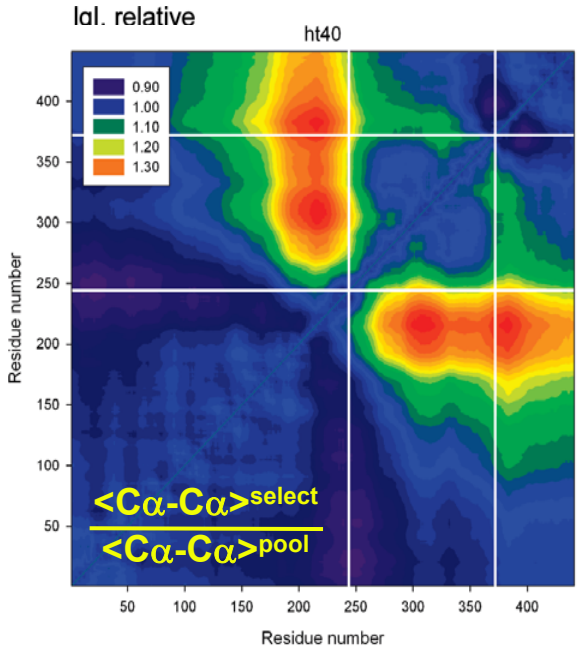
Adult Tau

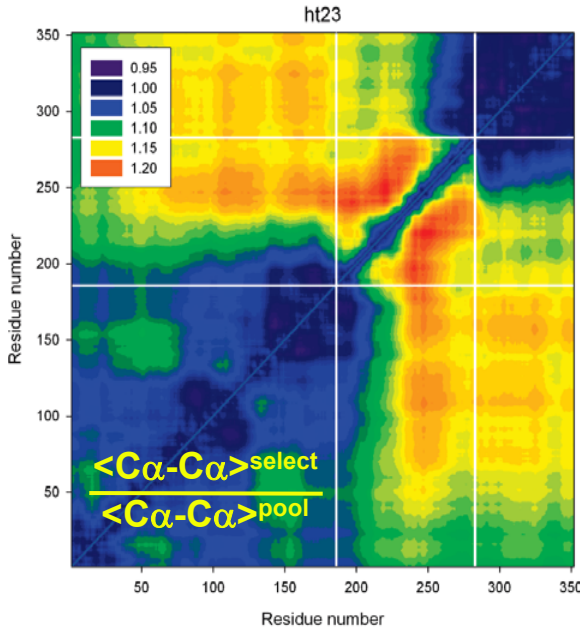


Fetal Tau



Mylonas et al. Biochemistry 2008, 47, 10345

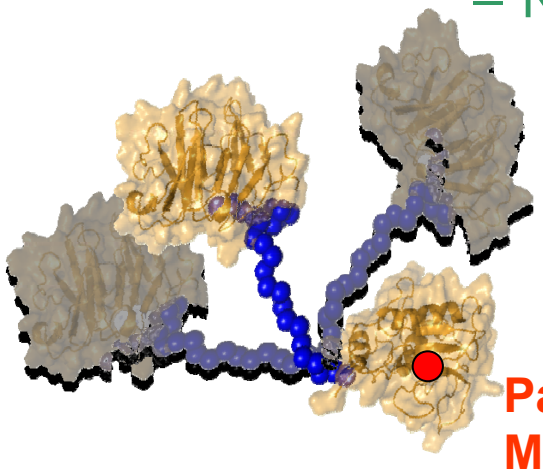




Other ensemble approaches

Maximum Occurrence Approach

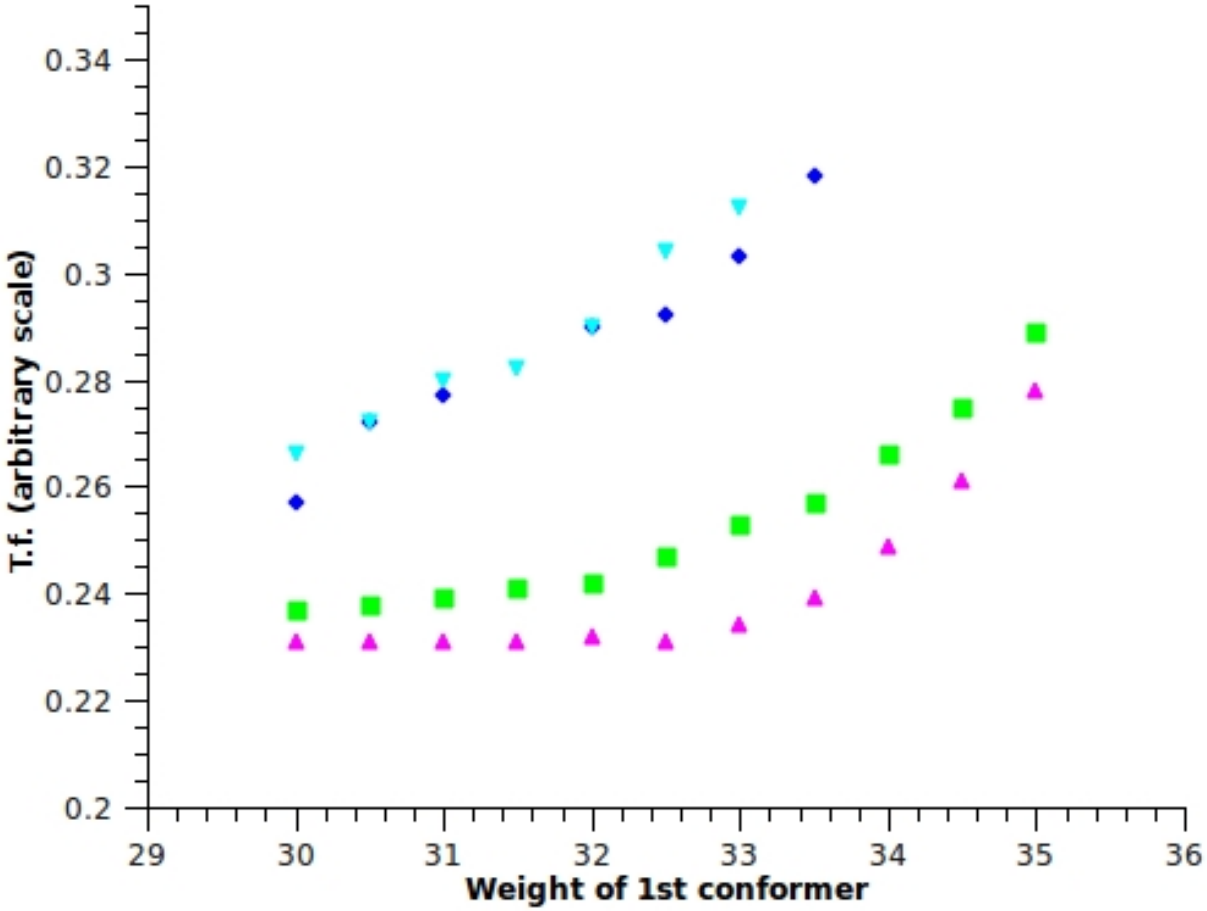
- The MO of a conformation can be calculated as the **maximum weight that the conformation can have** and be in agreement with the experimental data
  - NMR data
    - » Residual dipolar couplings
    - » Pseudo contact shifts
    - » SAXS

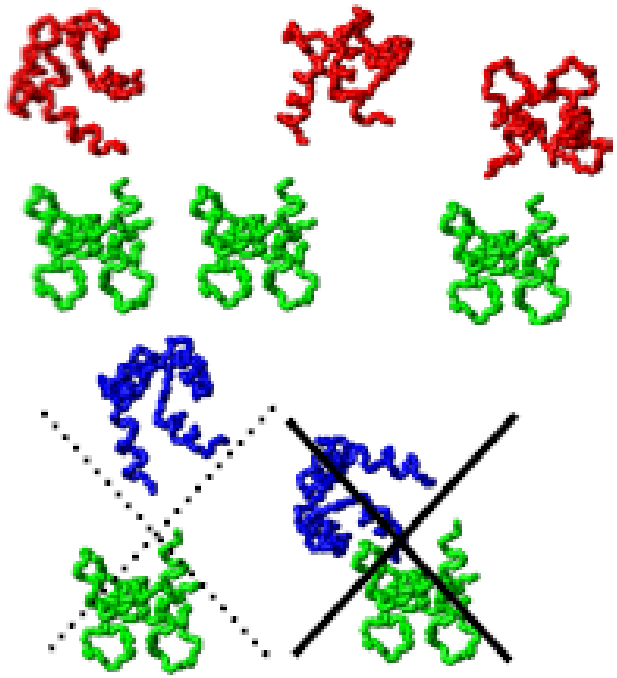


Paramagnetic Me

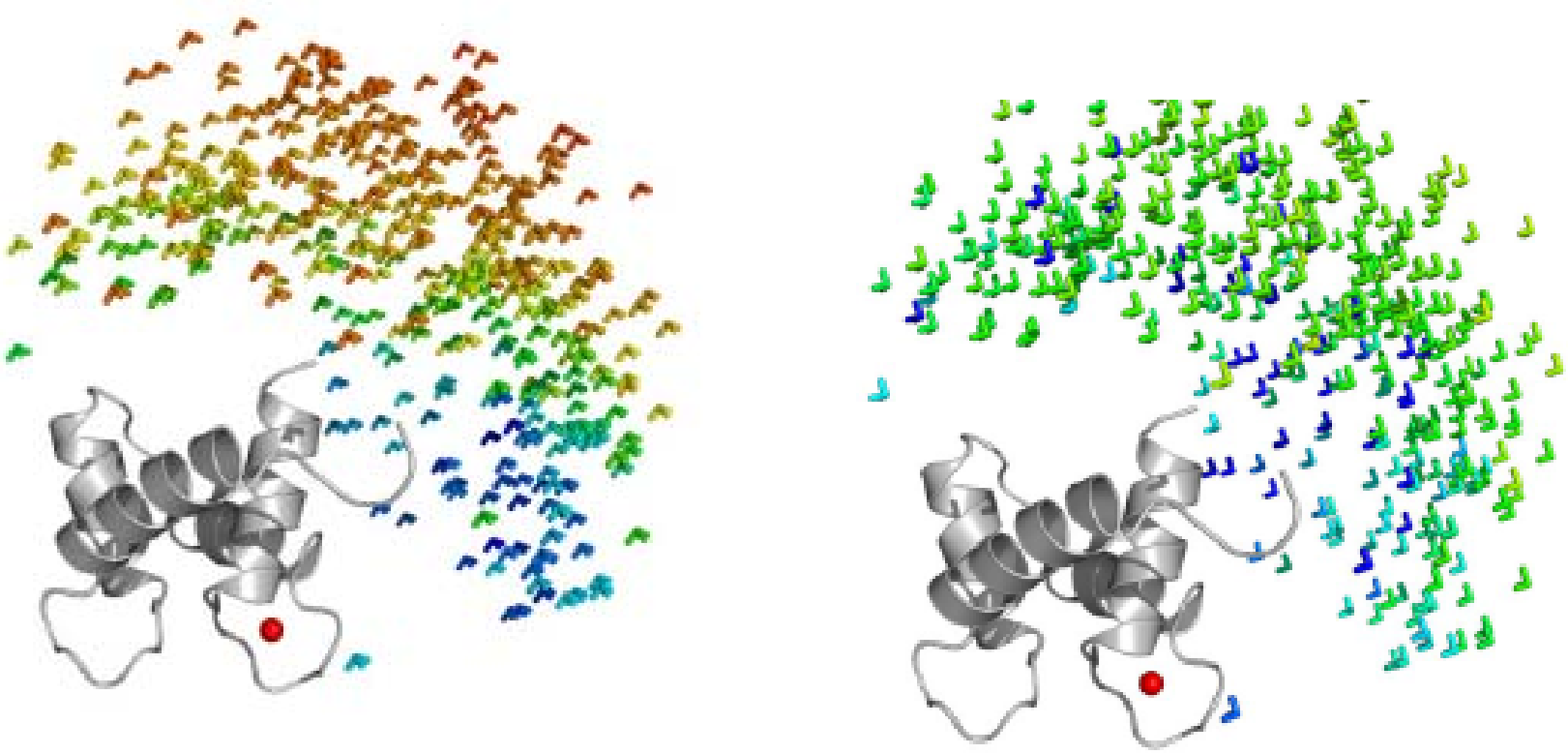
Bertini, Giachetti, Luchinat, Parigi, Petoukhov, Pierattelli, Ravera, Svergun, JACS, 2010

Maximum Occurrence Approach





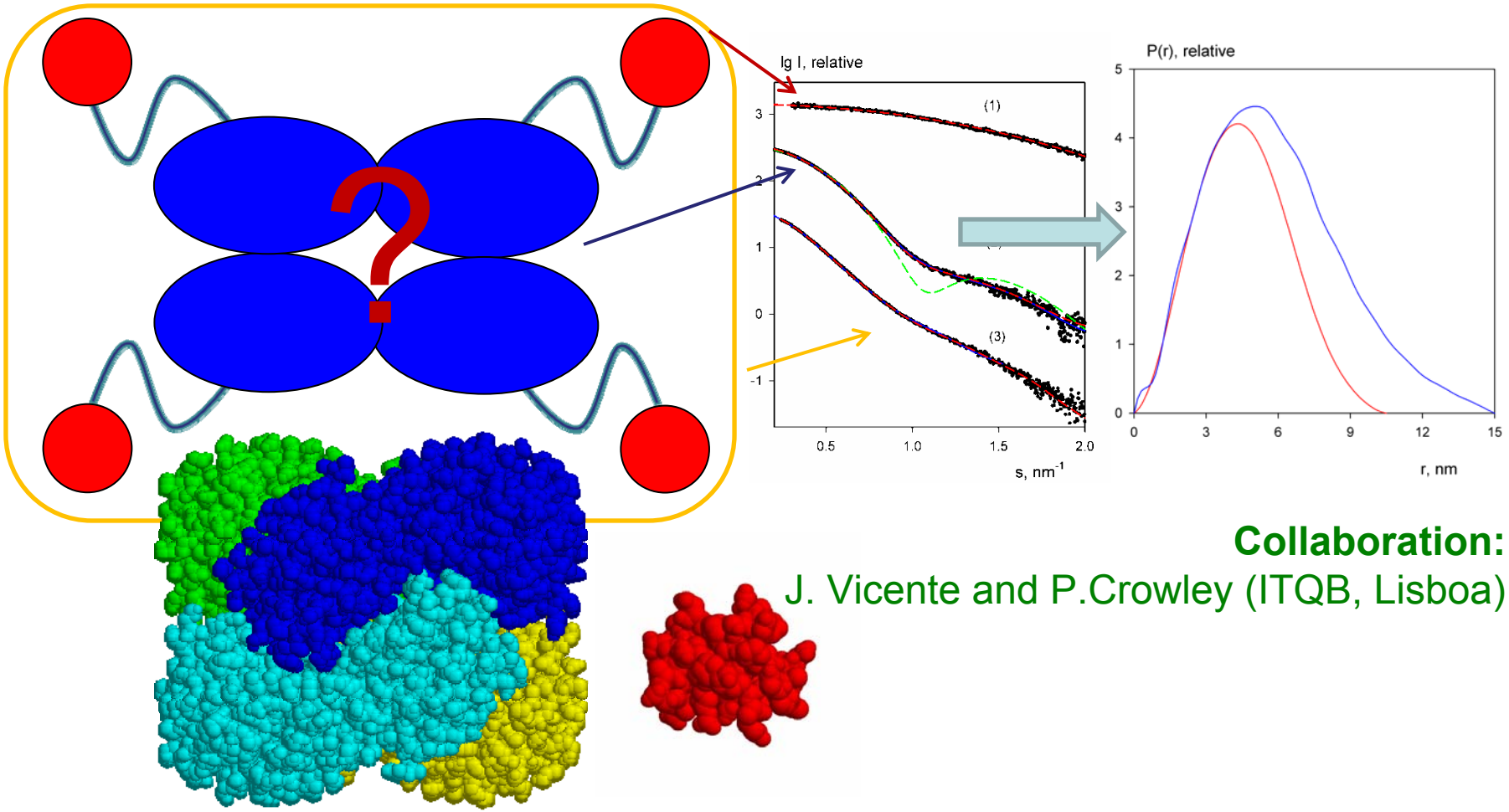
Probe of Calmodulin Conformations by Combining SAXS with NMR



Blue: 5% Red: 40%

E. coli Flavorubredoxin

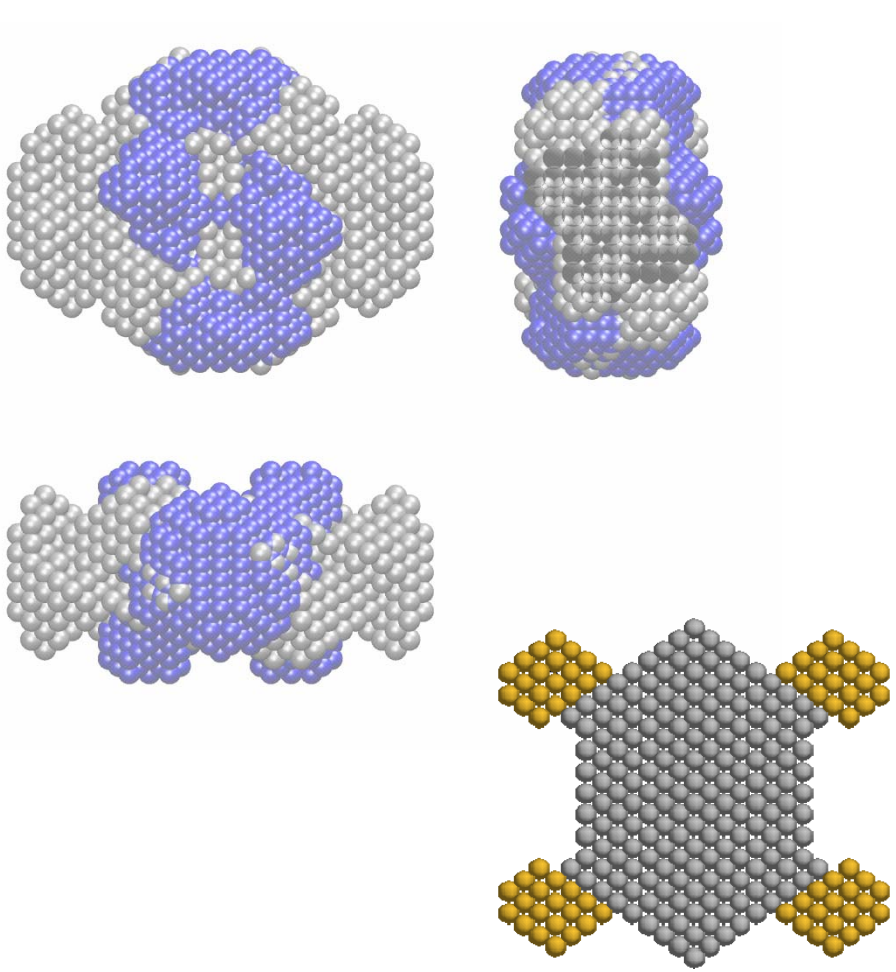
Modular enzyme endowed with nitric oxide and/or oxygen reductase activity

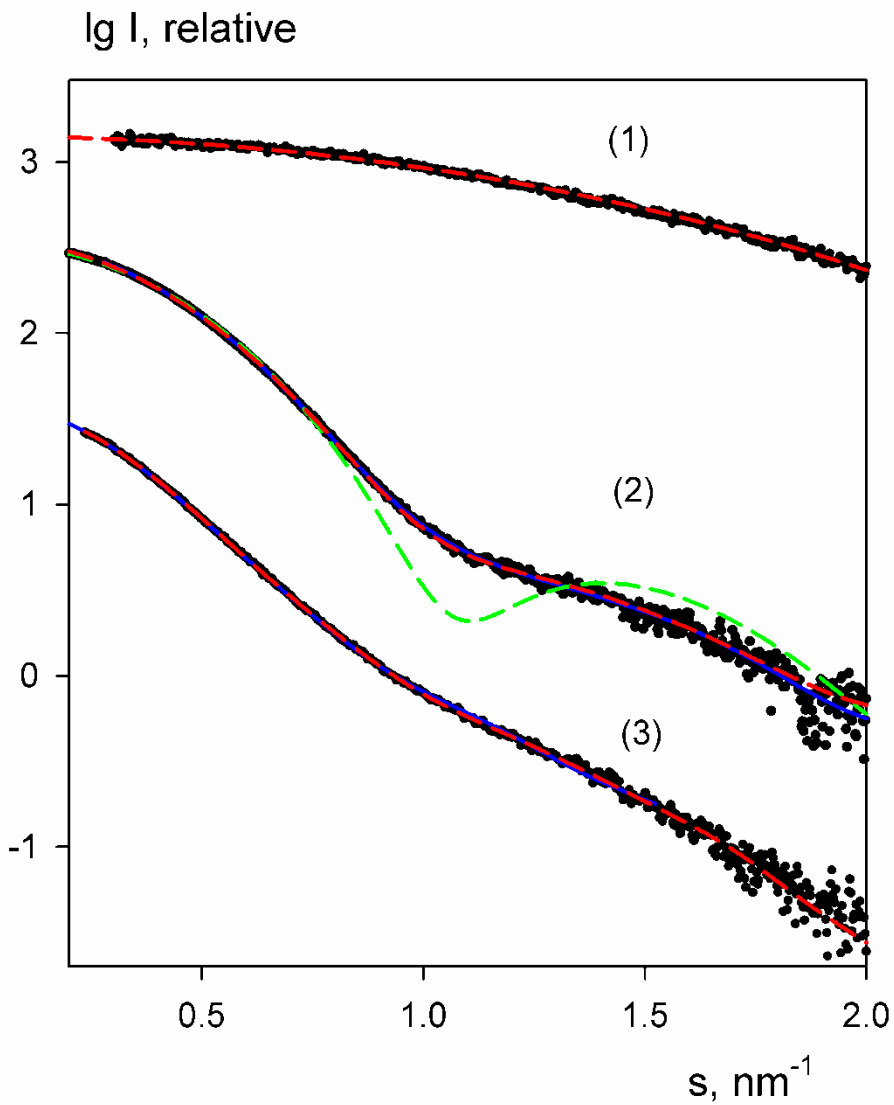


Collaboration:

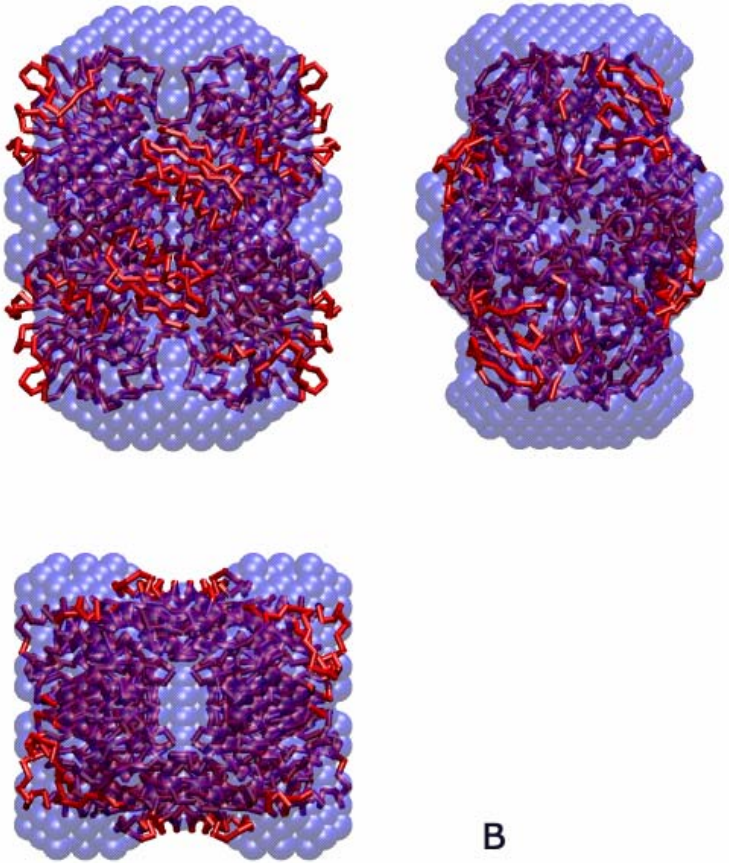
J. Vicente and P.Crowley (ITQB, Lisboa)

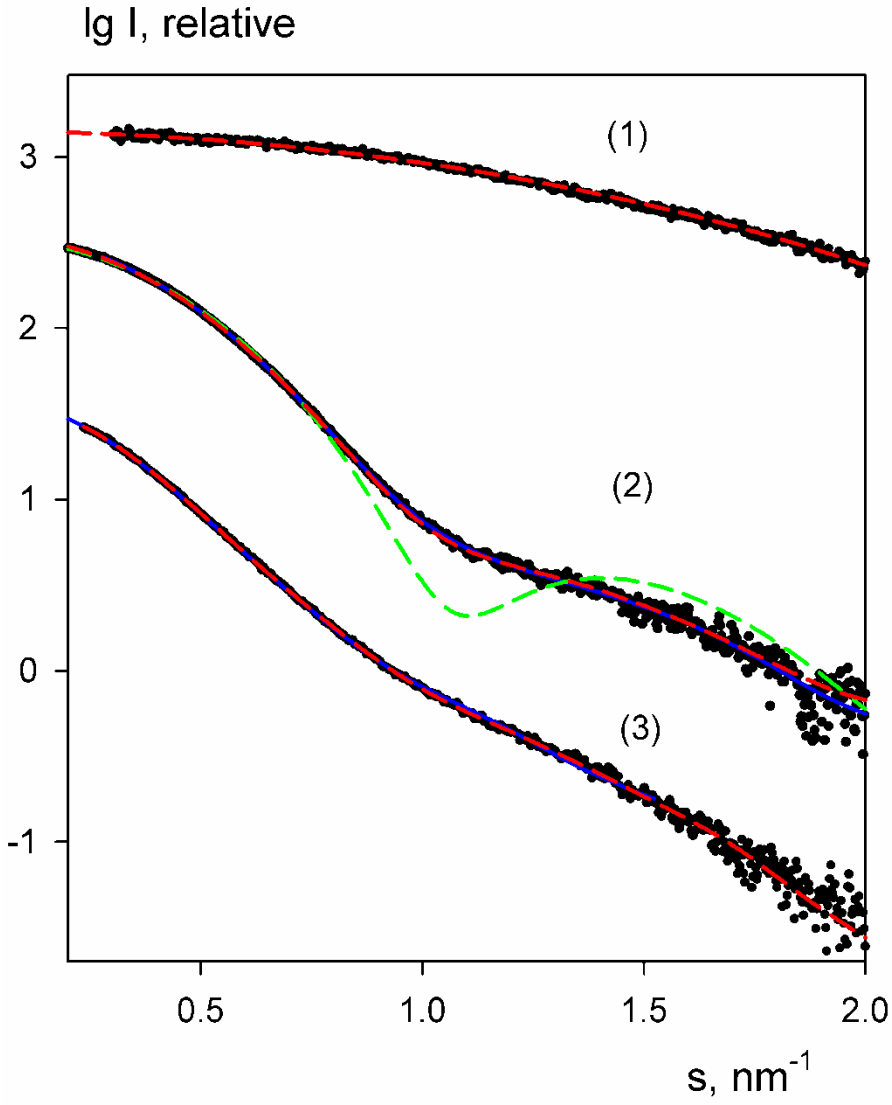
E. coli Flavorubredoxin: *ab initio* modelling



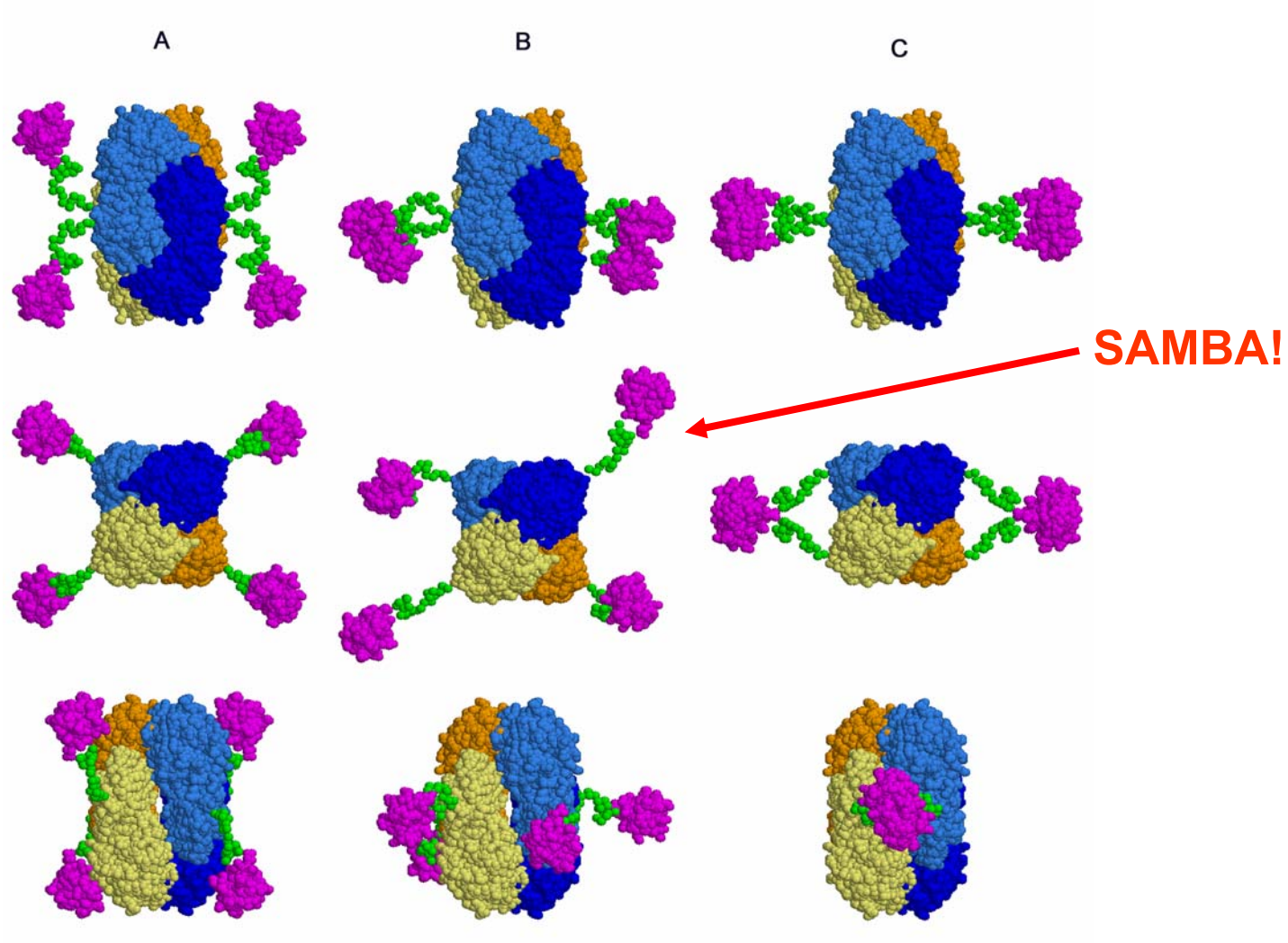


E. coli Flavorubredoxin: Xtal vs SAXS

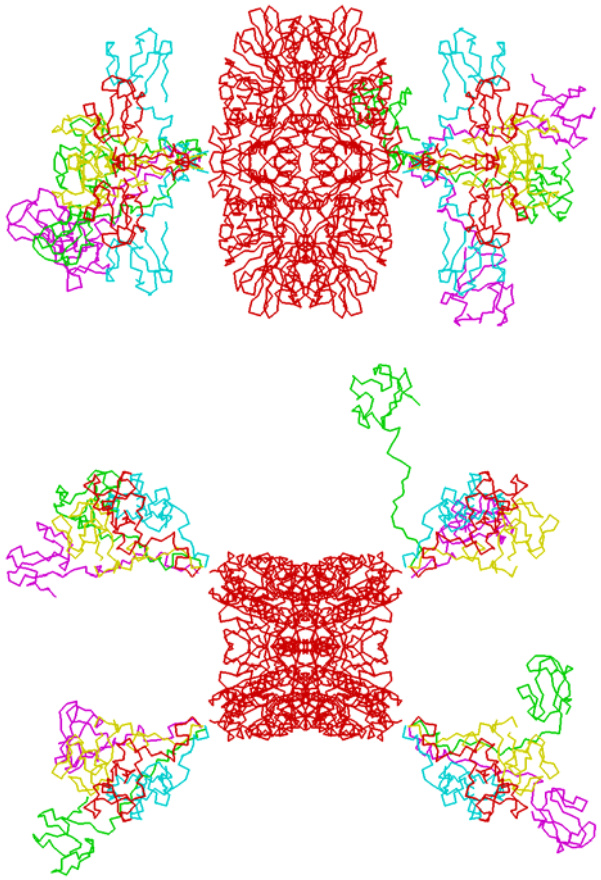


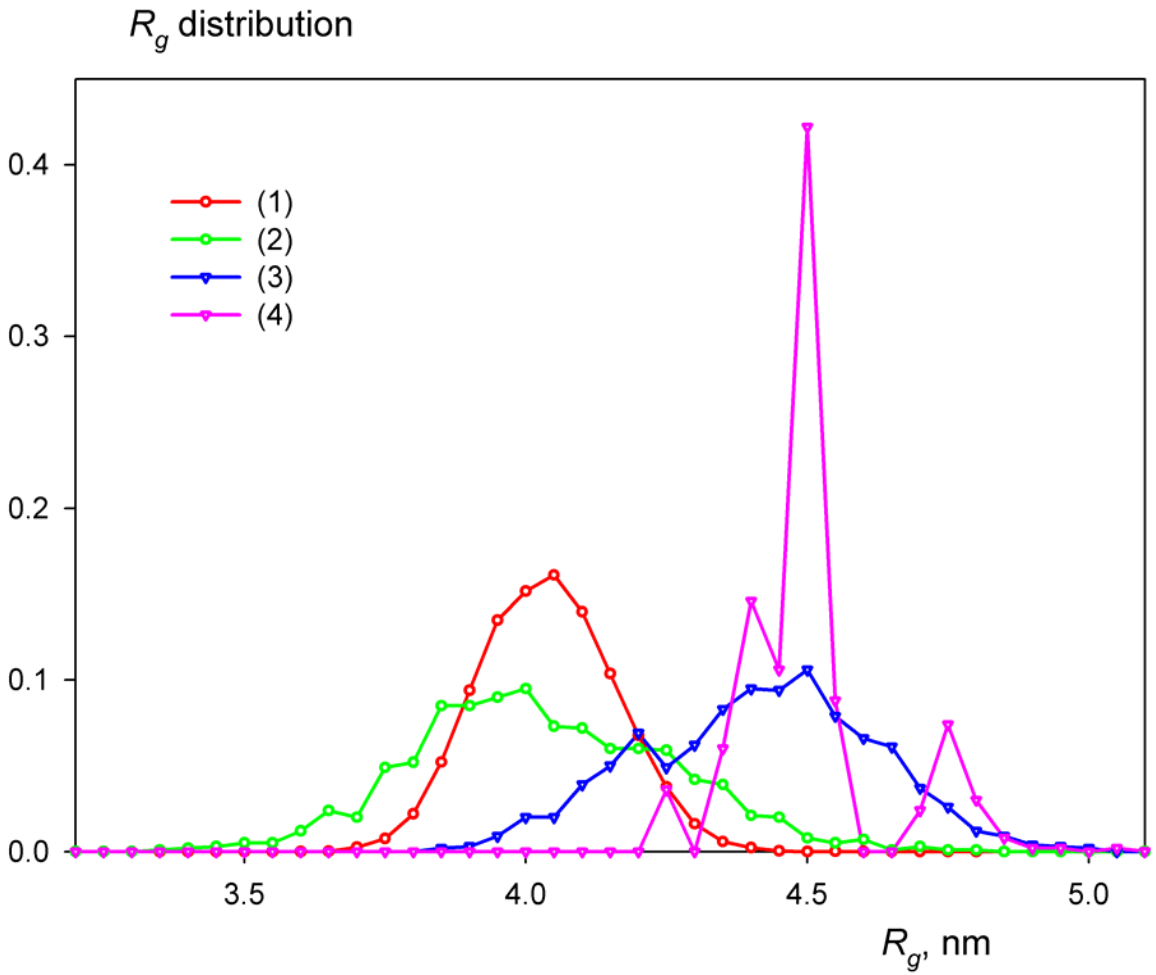


E. coli Flavorubredoxin: various constraints

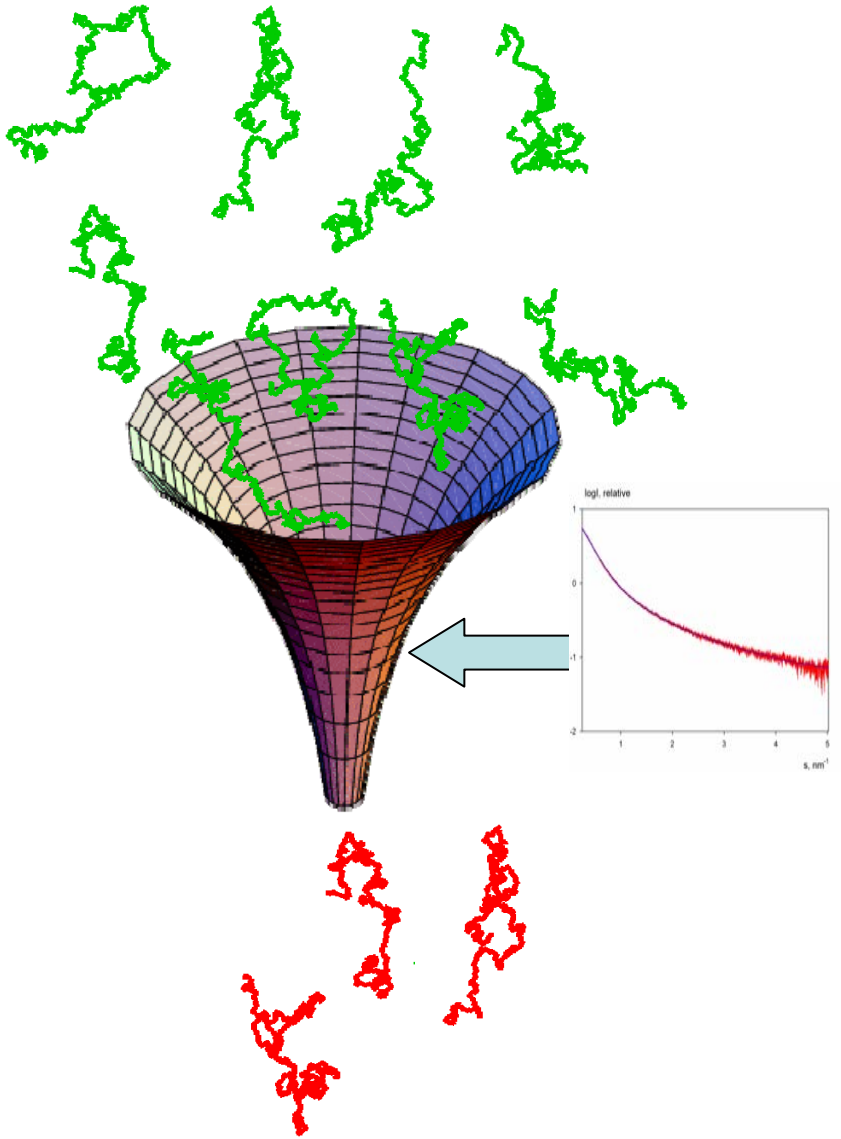


E. coli Flavorubredoxin: EOM



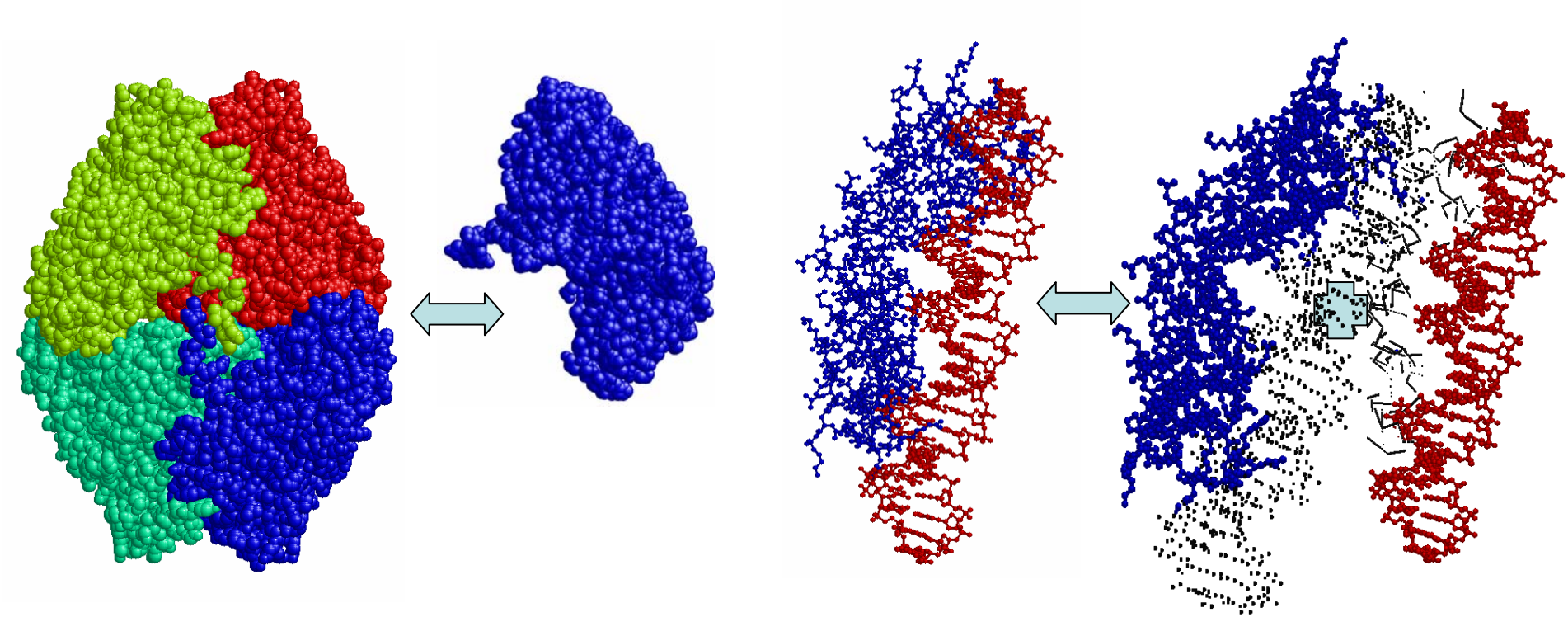


Flexible Systems Summary



- SAS van address protein flexibility, conformational changes and assembly/dissociation processes
- Ensemble Methods are appropriate tools to study (potentially) flexible molecules
- Unique structural information can be obtained based on distributions of descriptors whereas structures collected are simply a TOOL to describe the shape distributions

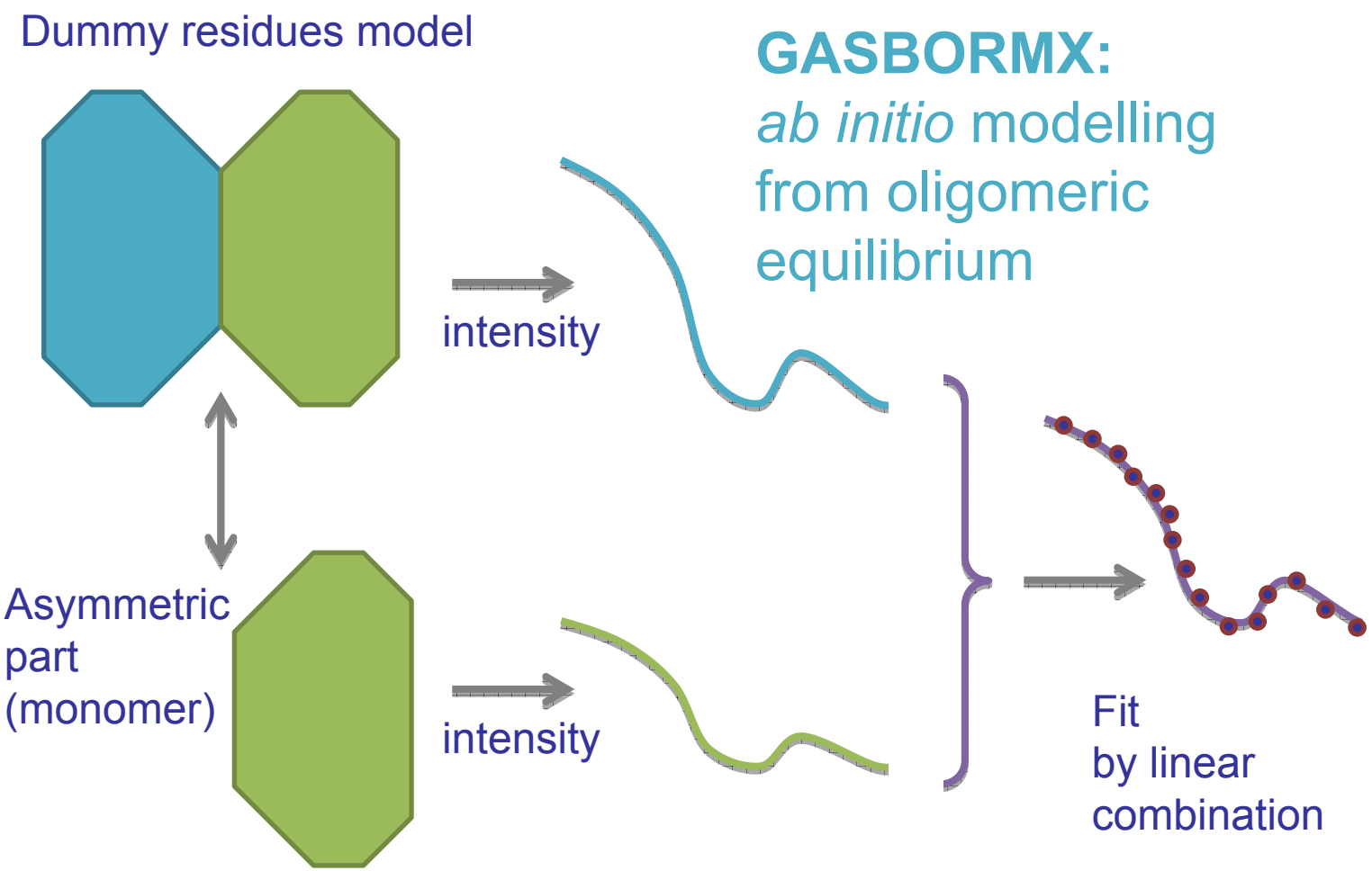
3D Reconstruction from Polydisperse Data?



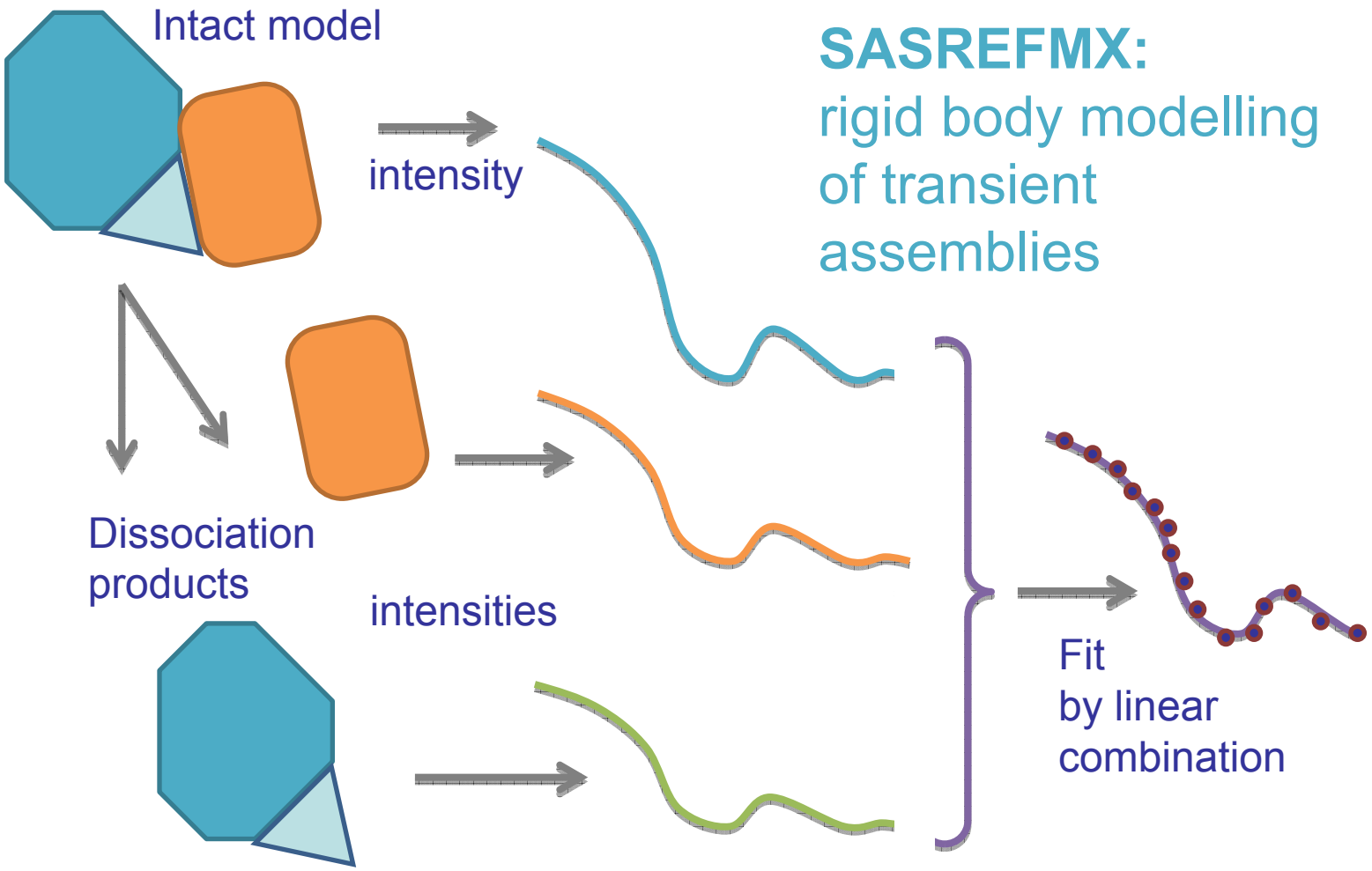
GASBOR_MX: quaternary structure of weak (symmetric) oligomers

SASREF_MX: structural analysis of transient complexes

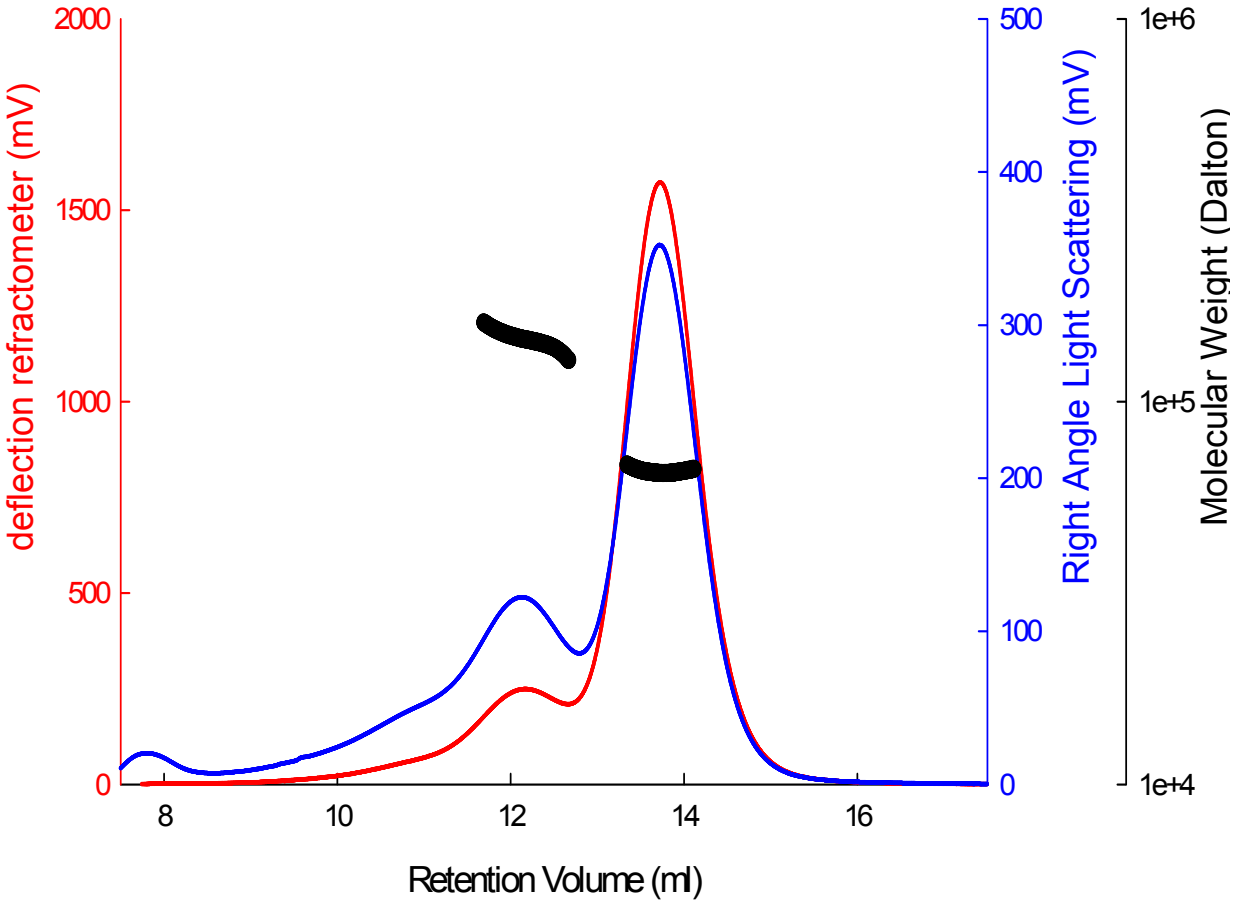
Ab initio Approach for Oligomeric Mixtures



Rigid Body Modelling from Mixtures

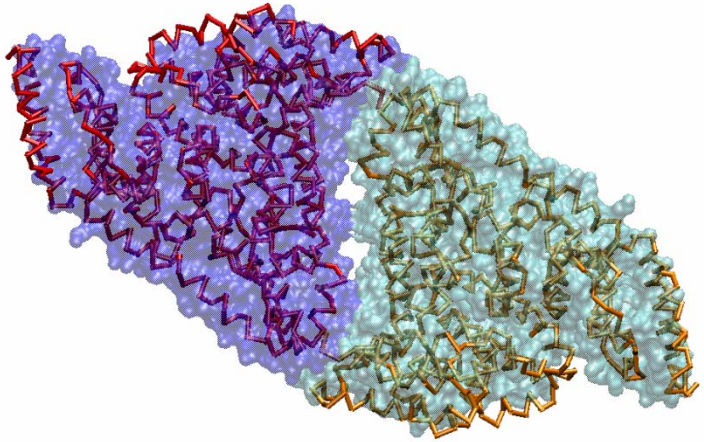


Proof of Principle: BSA Sample

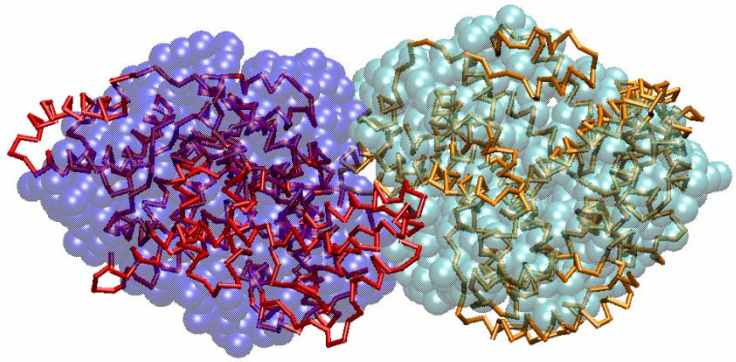


Proof of Principle: BSA Sample

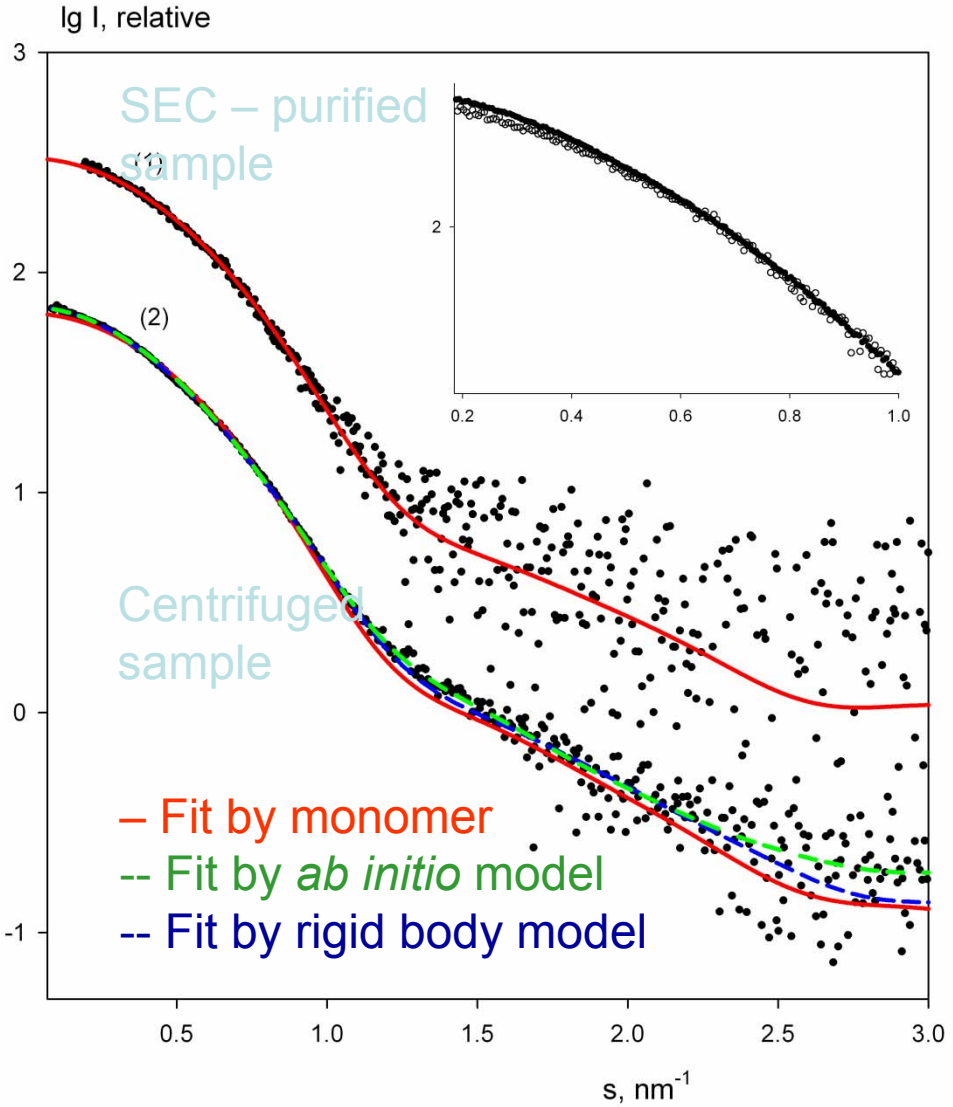
SASREFMX reconstruction



Closest interfaces from BSA crystal structure 3V03 shown as backbones

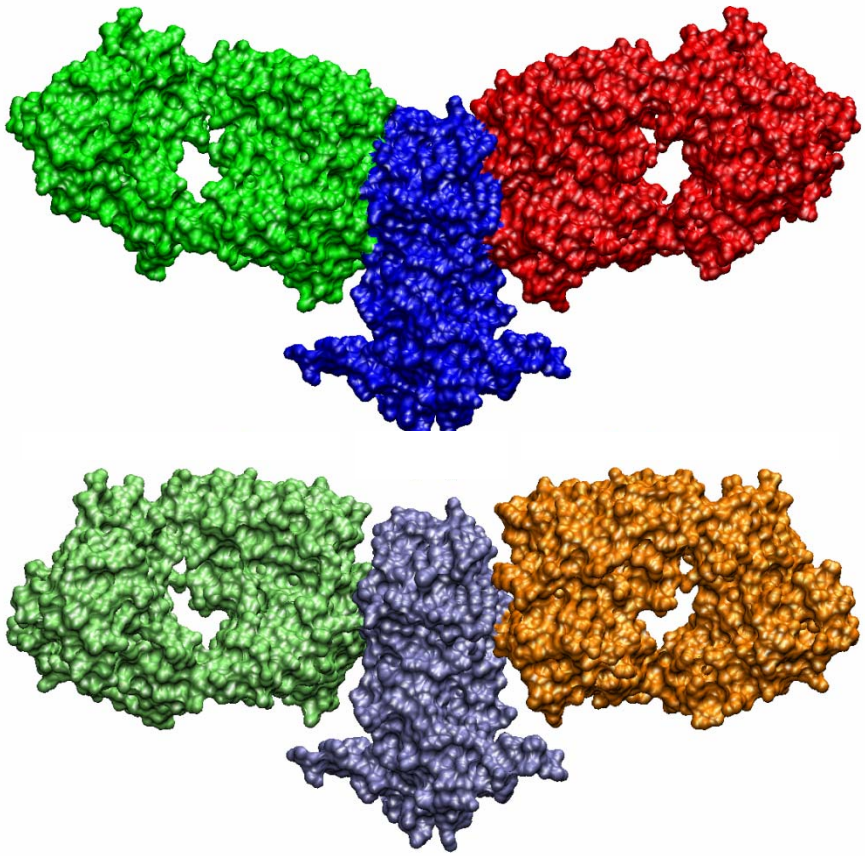


GASBORMX reconstruction

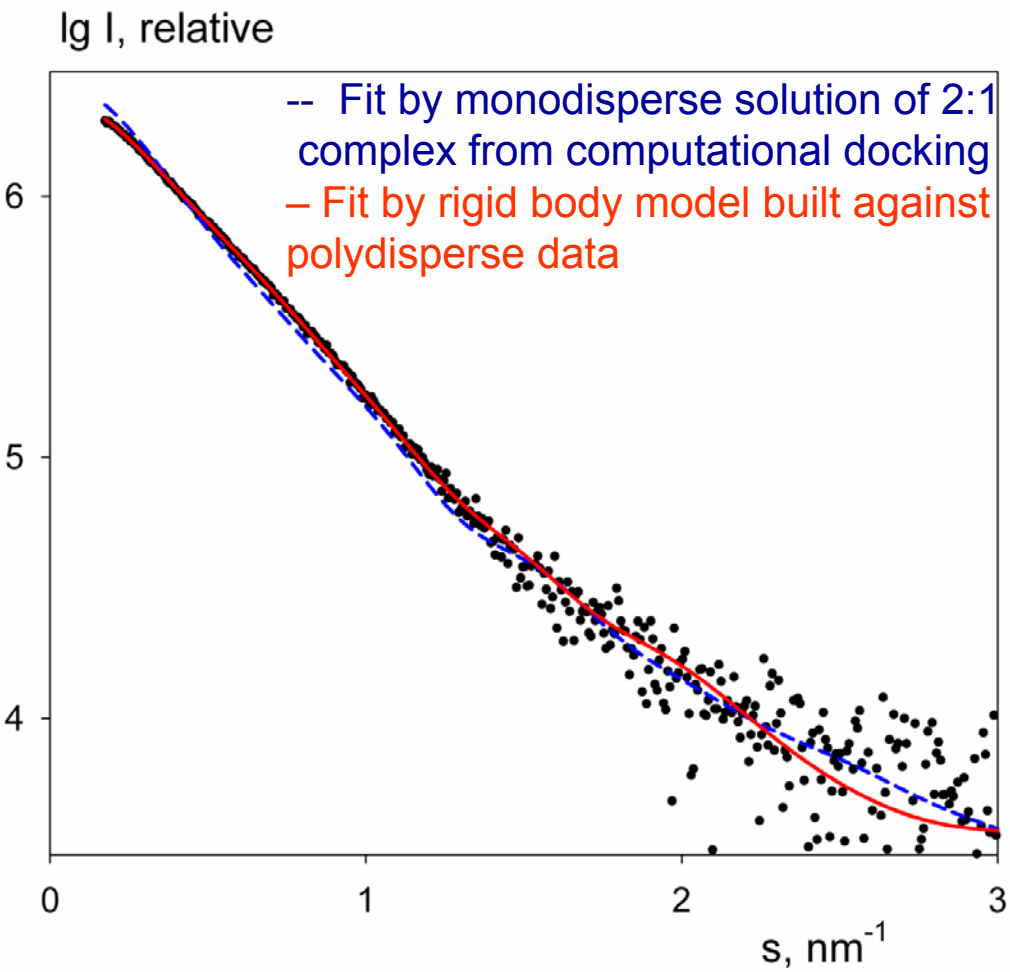


Partially Dissociated αD11 Fab–NGF Complex

Computational docking model



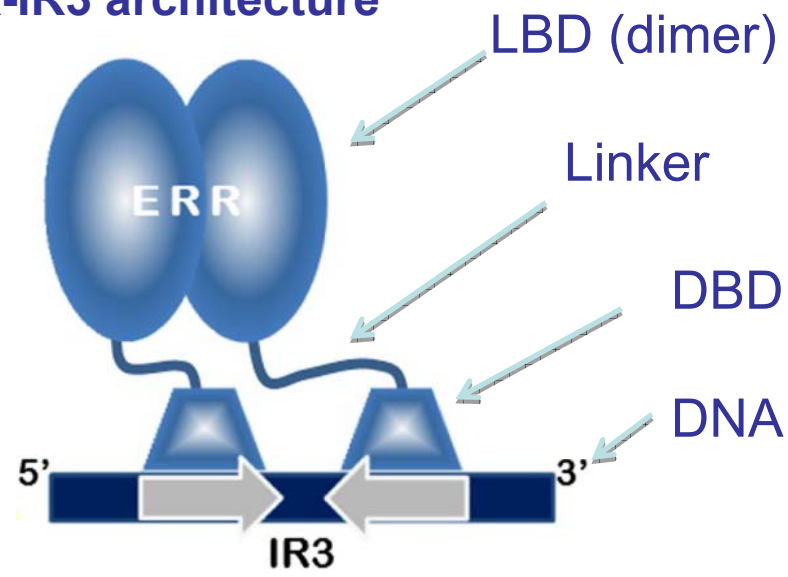
RB reconstruction from the mixture of 2:
Fab-NGF with its dissociation products



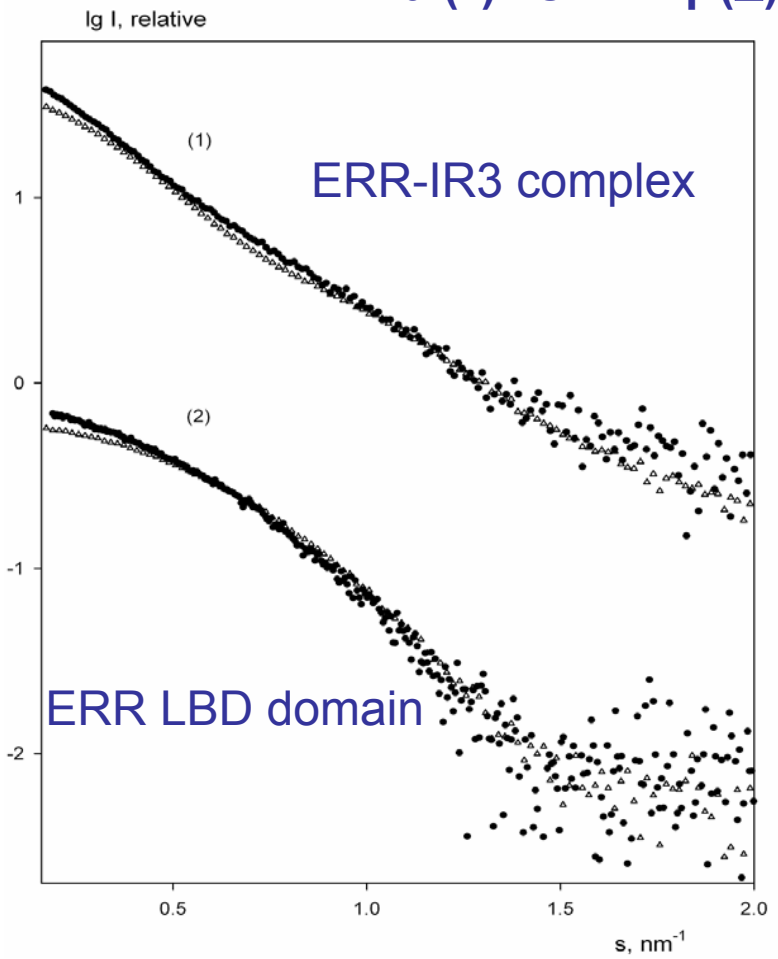
Covaceuszach, S., Cassetta, A., Konarev, P.V., Gonfloni, S., Rudolph, R., Svergun, D.I., Lamba, D. and Cattaneo, A. (2008), Journal of Molecular Biology 381, 881-896.

Study of ERR-IR3 Complex

ERR-IR3 architecture

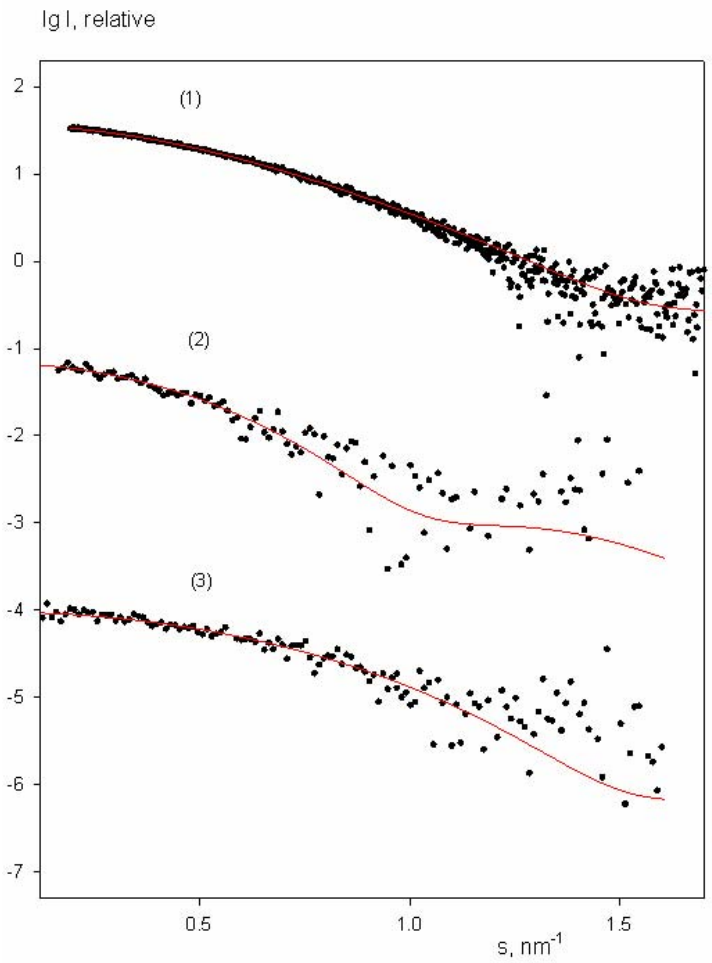


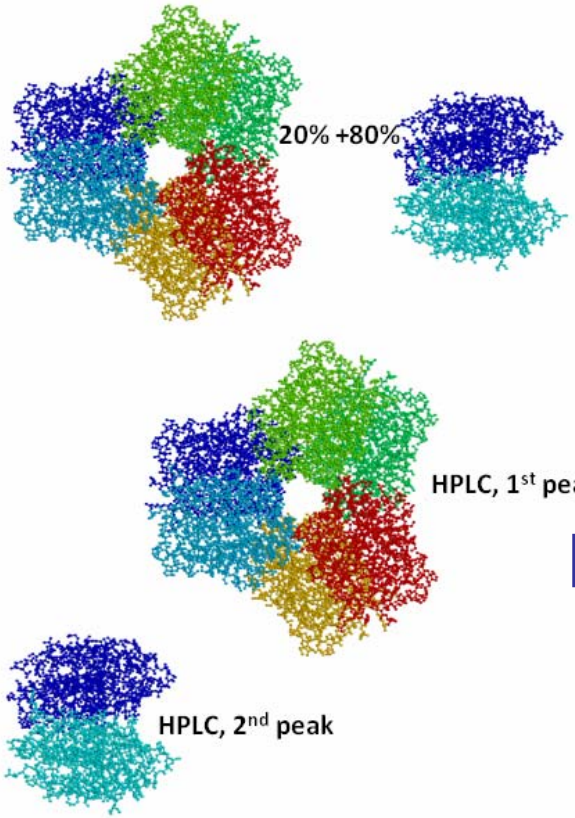
ERRα (•) *vs* ERRγ (Δ)



Petoukhov, M.V., Billas, I.M.L., Takacs, M., Graewert, M.A., Moras, D. & Svergun, D.I. (2013) Biochemistry

Study of ERR-IR3 Complex

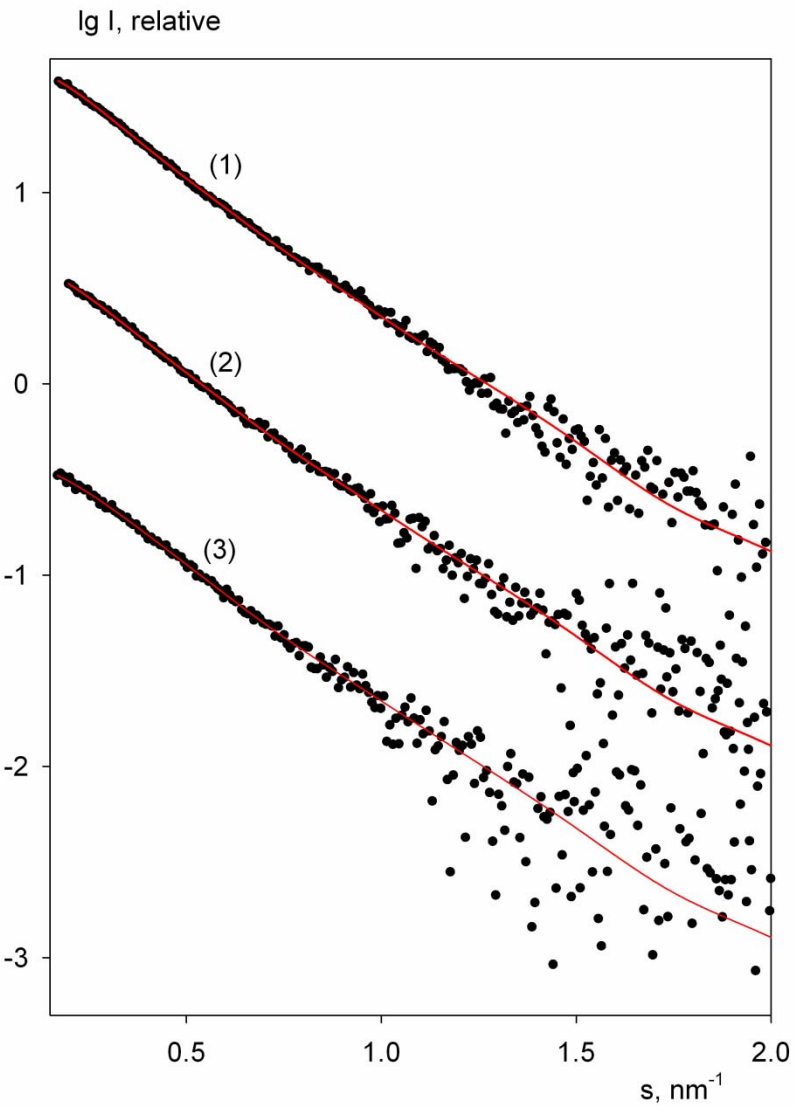


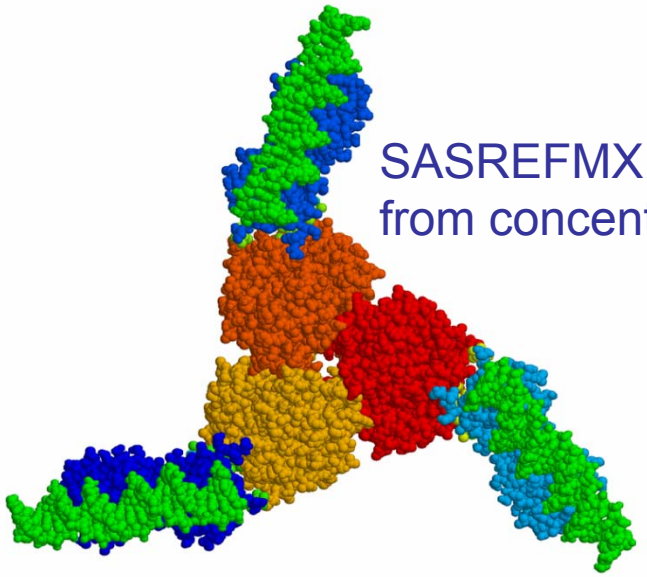


Differences between α and γ are attributed to partial hexamerization of ERRα LBD

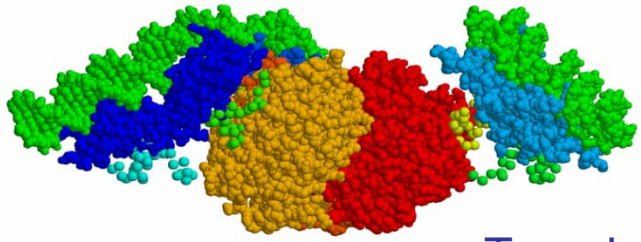
Petoukhov, M.V., Billas, I.M.L., Takacs, M., Graewert, M.A., Moras, D. & Svergun, D.I. (2013) Biochemistry

Study of ERR-IR3 Complex





SASREFMX reconstruction
from concentration series

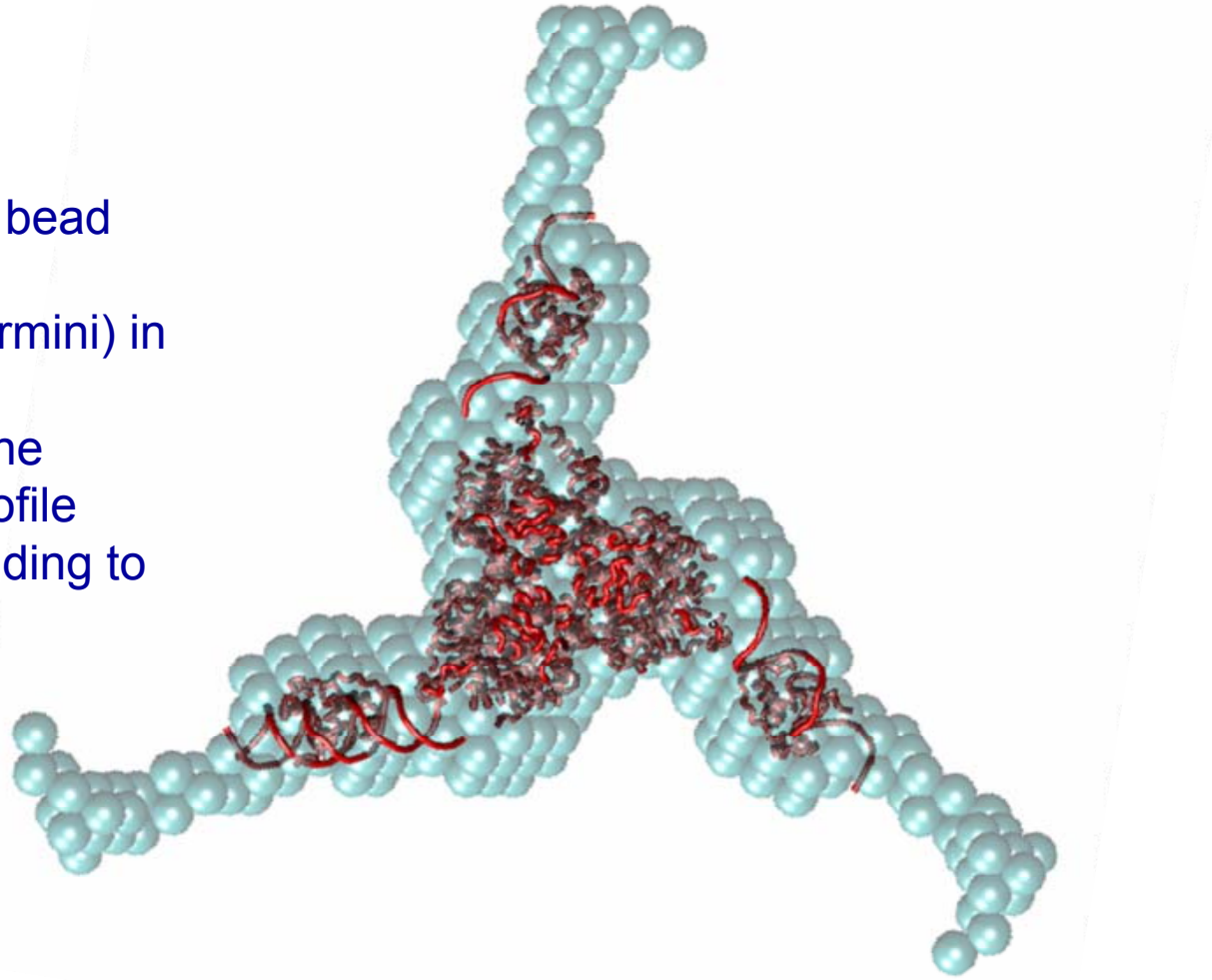


Transient 6:3 assembly
in equilibrium with 2:1 complex

Petoukhov, M.V., Billas, I.M.L., Takacs, M., Graewert, M.A., Moras, D. & Svergun, D.I. (2013) Biochemistry

Study of ERR-IR3 Complex

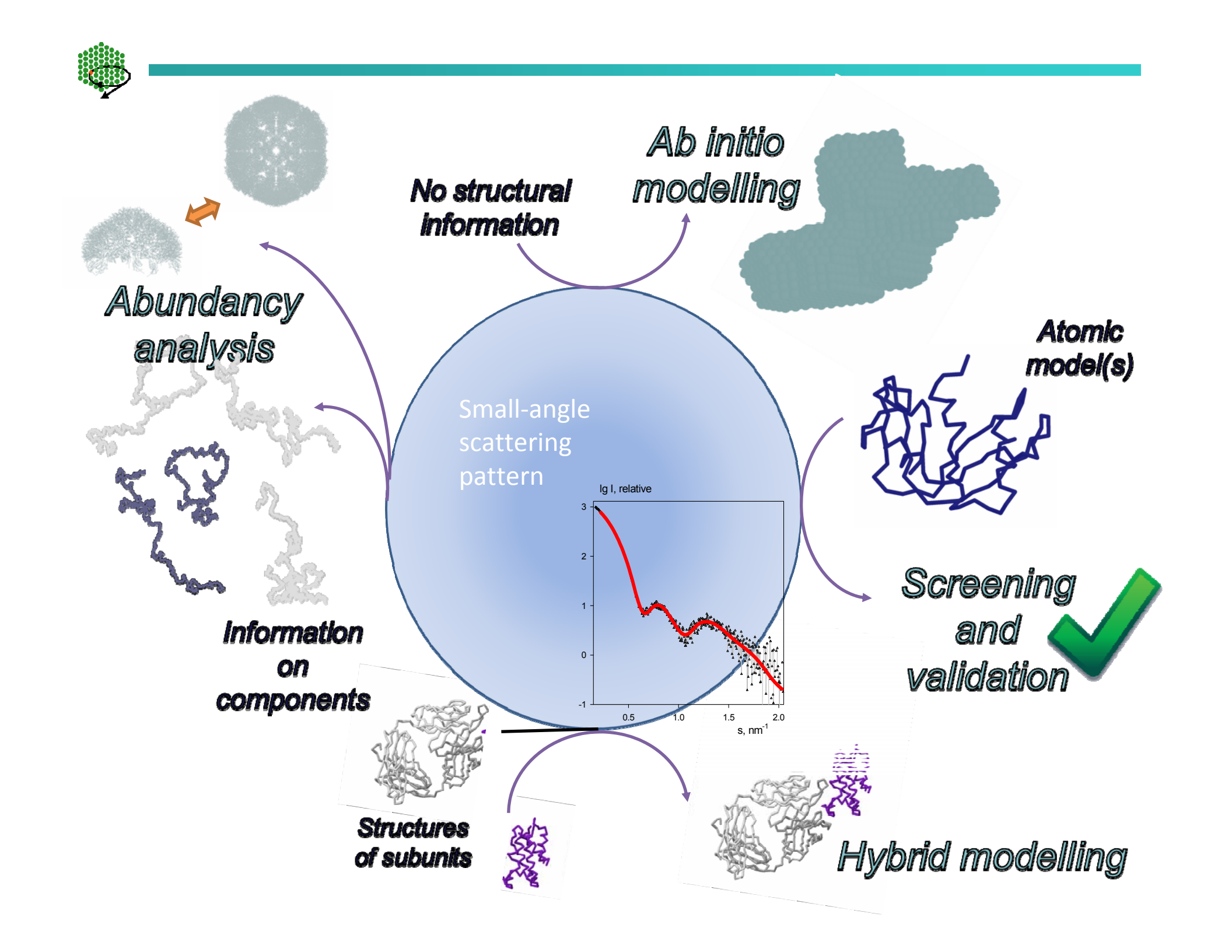
Cross validation:
Comparison with *ab initio* bead model of the full length ERRα (including N-termini) in complex with IR3, which is built against the monodisperse SAXS profile from HPLC peak corresponding to the 6:3 assembly



Petoukhov, M.V., Billas, I.M.L., Takacs, M., Graewert, M.A., Moras, D. & Svergun, D.I. (2013) Biochemistry

3D Modelling from Equilibrium Mixtures

- Transient complexes and weak oligomers might be modelled (with caution!) against SAXS data
- Volume fractions of the entire assembly and of dissociation products are additional minimization parameters
- *Ab initio* analysis of weak symmetric homo-oligomers is done by dummy residues approach
- Quaternary structure of dissociating multisubunit assemblies (including nucleoproteins) is reconstructed by rigid body modelling
- Multiple scattering profiles (e.g. from concentration series) can be fitted simultaneously


No structural information
Ab initio modelling
Abundancy analysis
Atomic model(s)
Small-angle scattering pattern
lg I, relative
3
2
1
0
-1
0.5
1.0
1.5
2.0
s, nm-1
Screening and validation
Information on components
Structures of subunits
Hybrid modelling