Atomic-Structure-Based Modeling

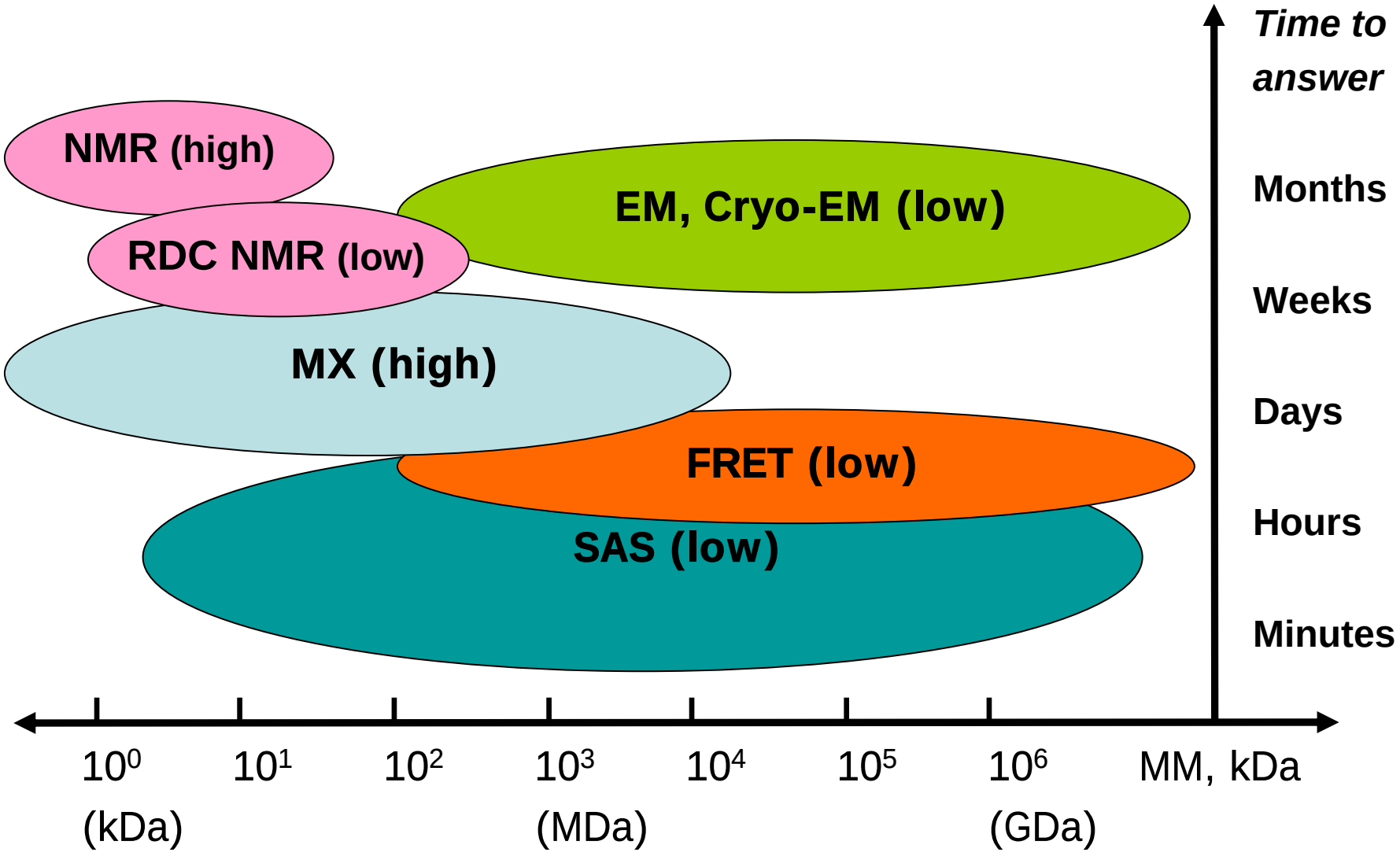
Maxim Petoukhov
EMBL, Hamburg Outstation

EMBL

Outline

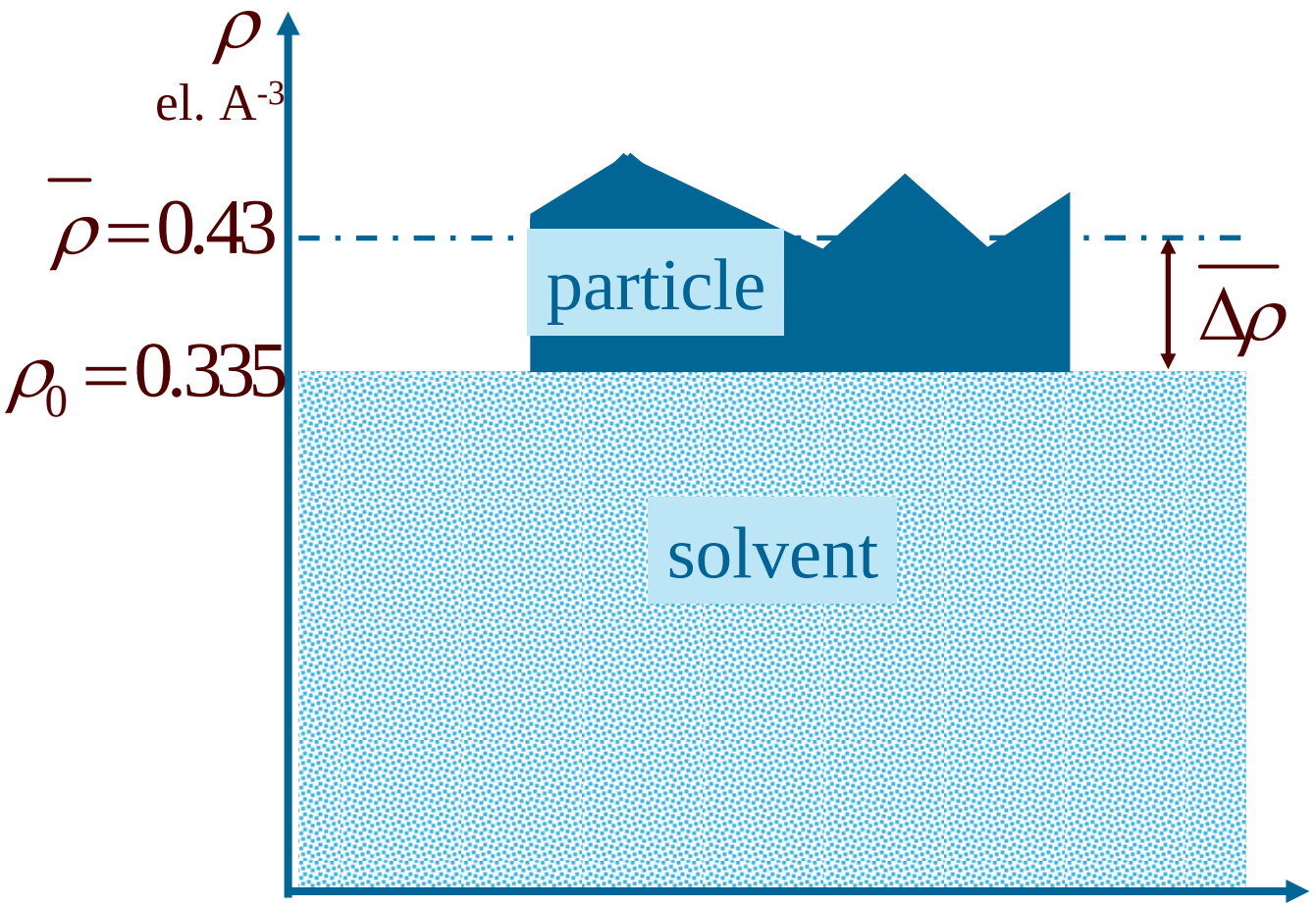
- Introduction
- Computation of SAS patterns from atomic models
- Incorporation of structural information from other methods
- Rigid body modelling of macromolecular complexes
- Hybrid modelling of multidomain proteins
- *Examples & questions*

Structural methods: resolution, accessible size and speed of experiment/analysis

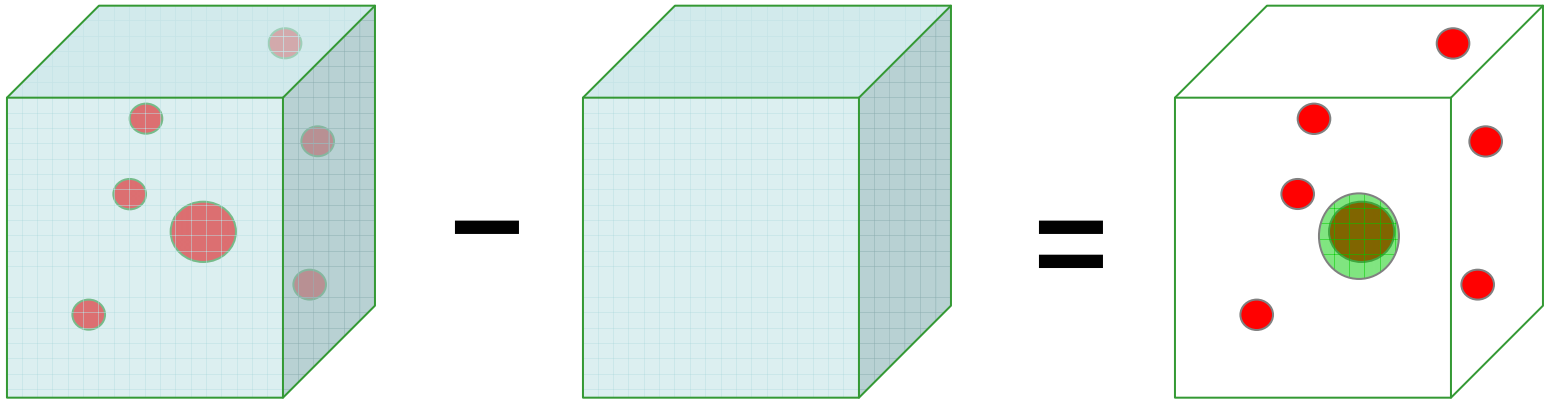


SAS Curve From Atomic Model – Is Your Structure Correct ?

Contrast of electron density



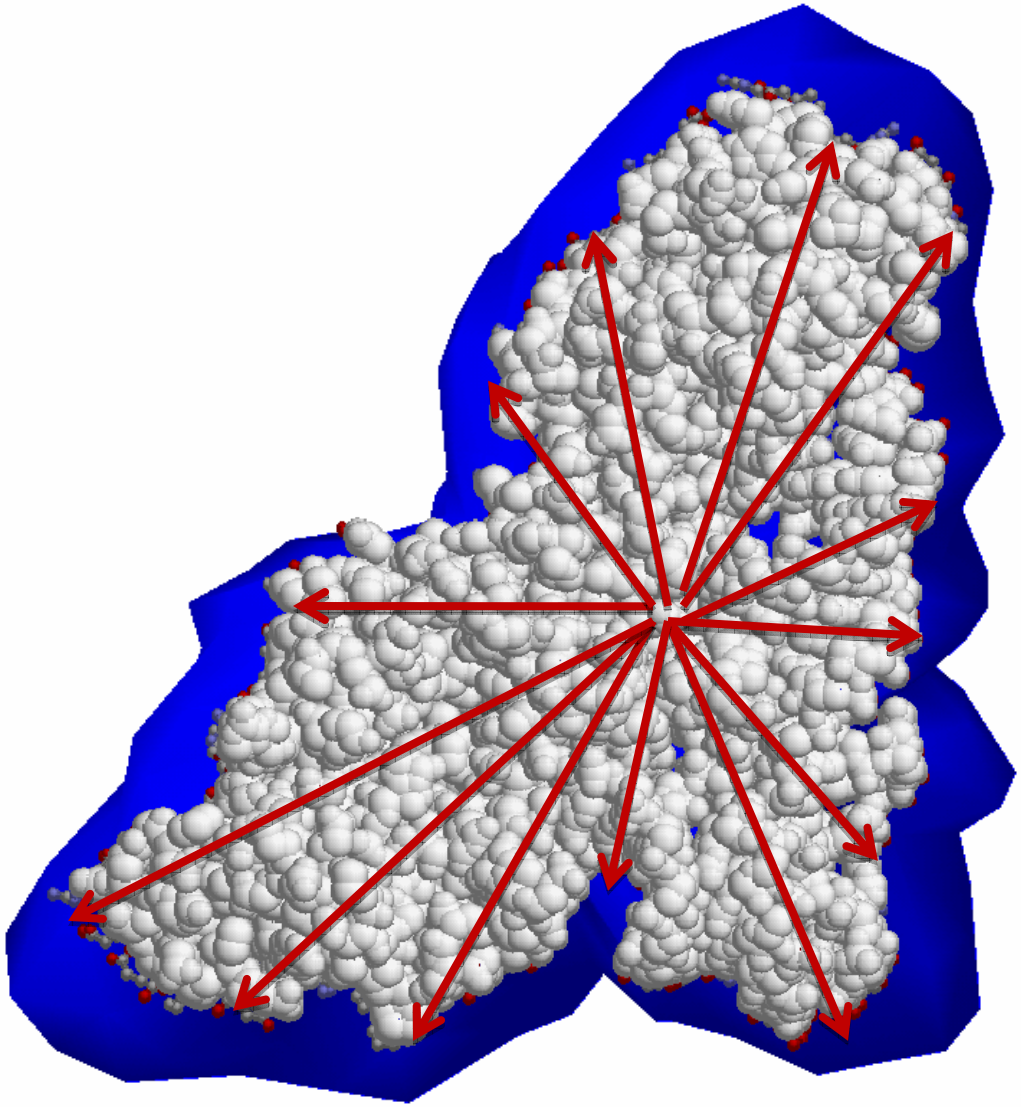
How to Compute SAS from Atomic Model



$I_{solution}(s)$ $I_{solvent}(s)$ $I_{particle}(s)$

- To obtain scattering from the particles, solvent scattering must be subtracted to yield effective density distribution $\Delta\rho = \langle\rho(\mathbf{r}) - \rho_s\rangle$, where ρ_s is the scattering density of the solvent
- Further, the bound solvent density may differ from that of the bulk

Scattering from a Macromolecule in Solution



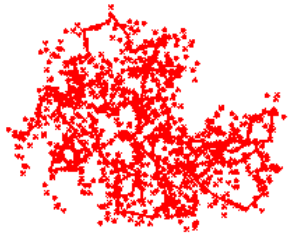
Atomic scattering

- Excluded volume

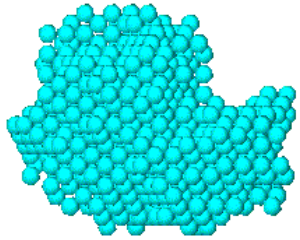
+ Shell scattering

Scattering Intensity via Amplitudes

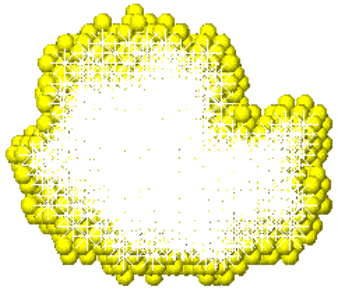
$$I(s) = \left\langle \left| A(\mathbf{s}) \right|^2 \right\rangle_\Omega = \left\langle \left| A_a(\mathbf{s}) - \rho_s A_s(\mathbf{s}) + \delta\rho_b A_b(\mathbf{s}) \right|^2 \right\rangle_\Omega$$



- $A_a(\mathbf{s})$: atomic scattering in vacuum



- $A_s(\mathbf{s})$: scattering from the excluded volume

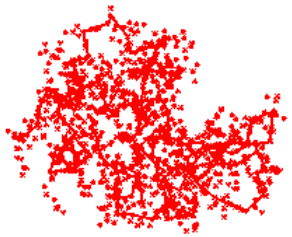


- $A_b(\mathbf{s})$: scattering from the hydration shell

***CRYSOL** (X-rays):* Svergun et al. (1995). *J. Appl. Cryst.* **28**, 768

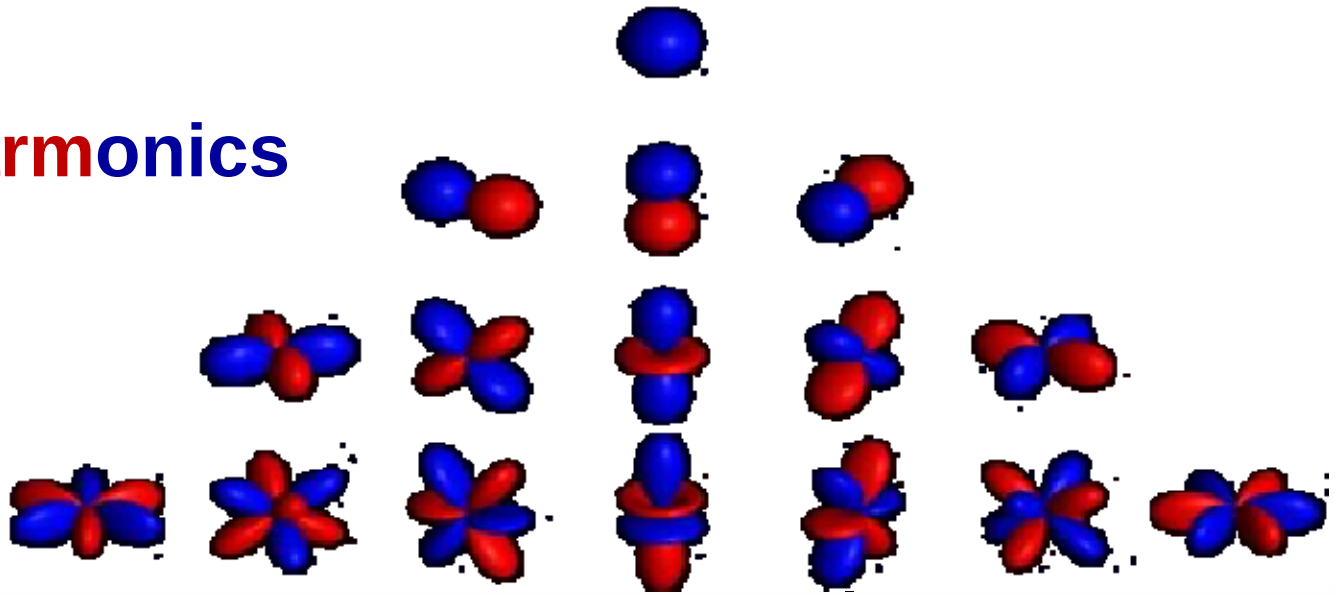
***CRYSON** (**neutrons**):* Svergun et al. (1998) *P.N.A.S. USA*, **95**, 2267

Use of Multipole Expansion



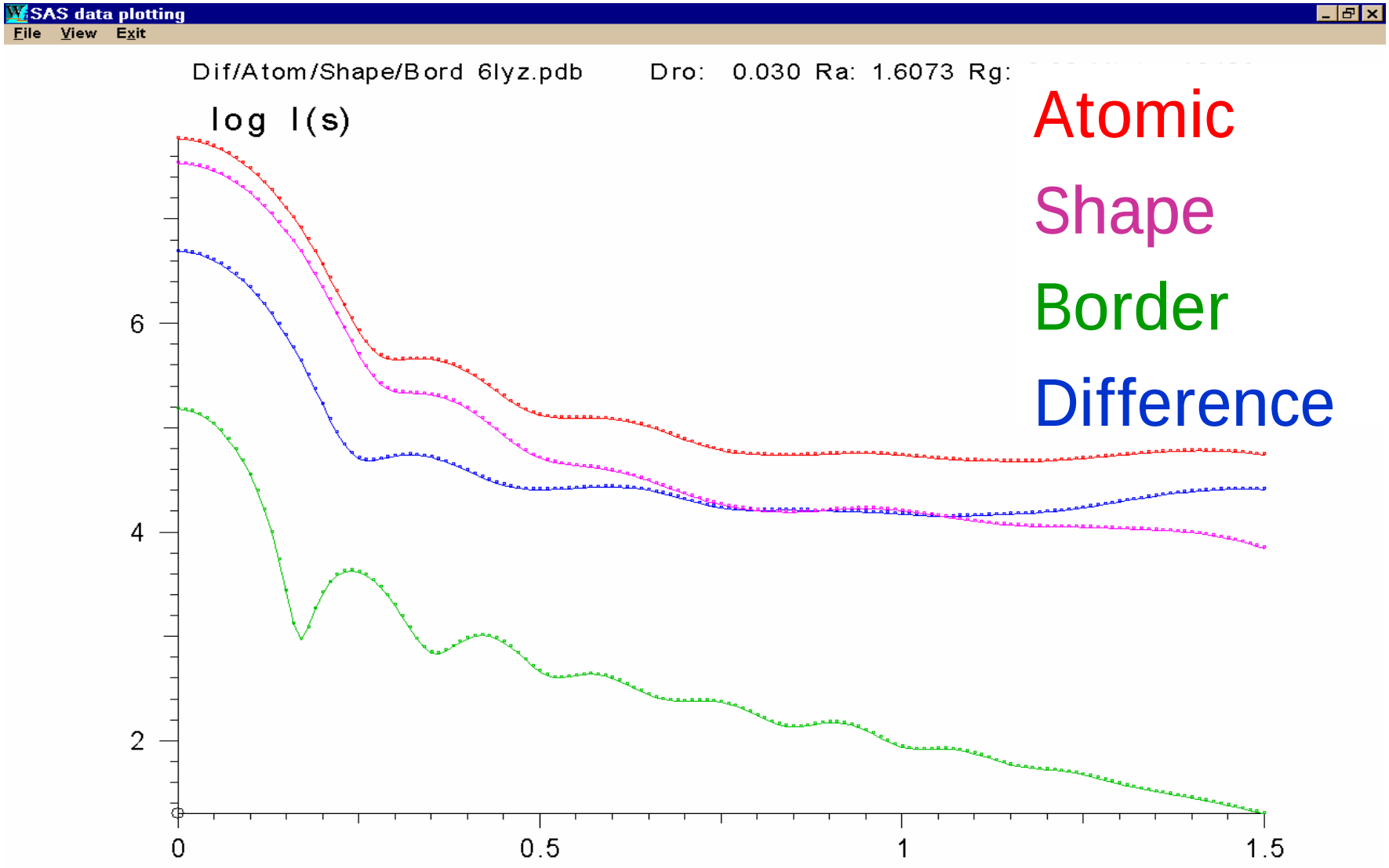
$$_a(\mathbf{s}) = \sum_{j=1}^{N} f_j(s) \exp(i\mathbf{s}\mathbf{r}_j)$$

spherical harmonics expansion



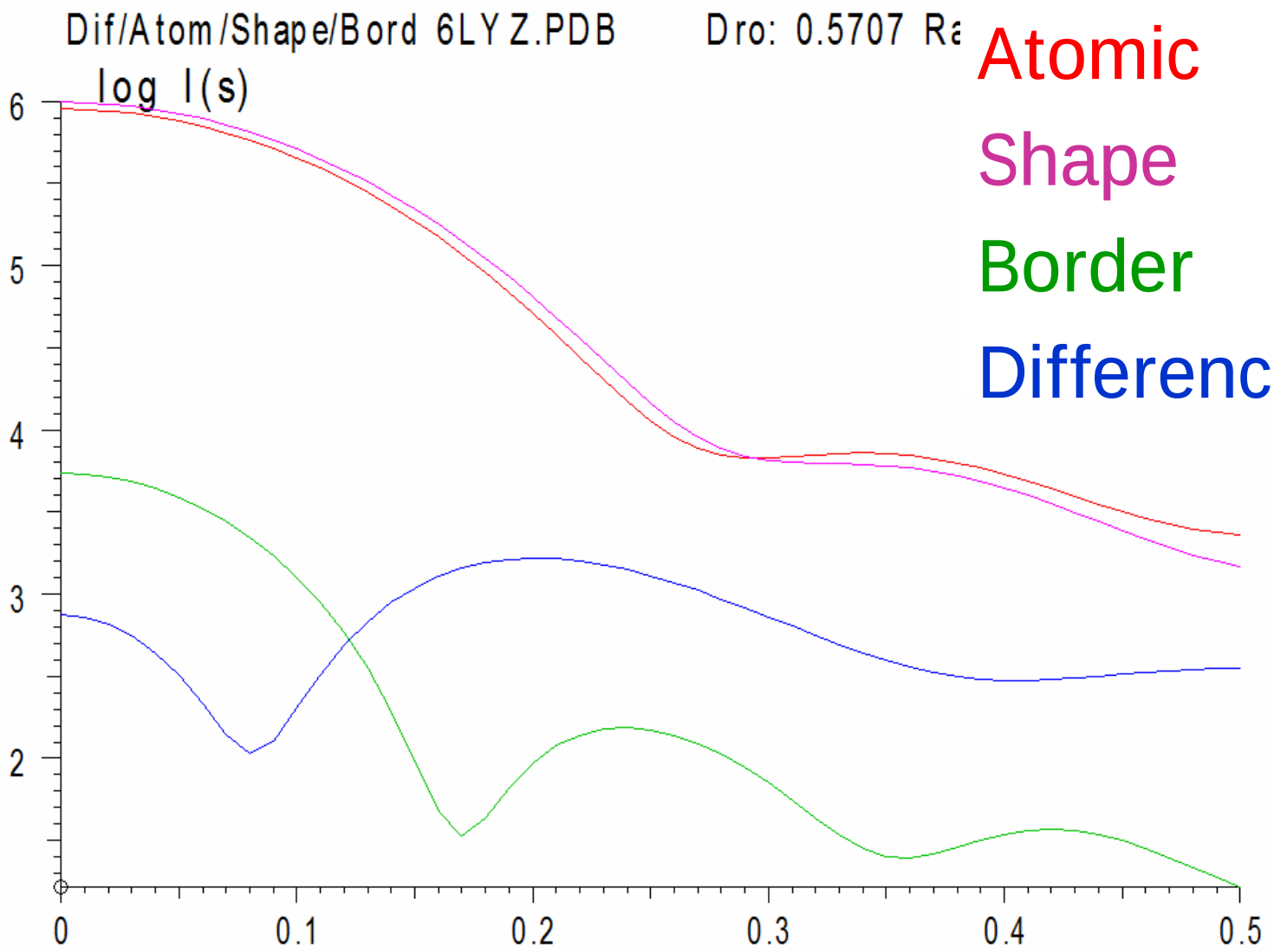
$$\exp(i\mathbf{s}\mathbf{r}) = 4\pi \sum_{l=0}^{\infty} \sum_{m=-l}^{l} i^l j_l(sr) Y^*_{lm}(\omega) Y_{lm}(\Omega)$$

Scattering components (lysozyme)



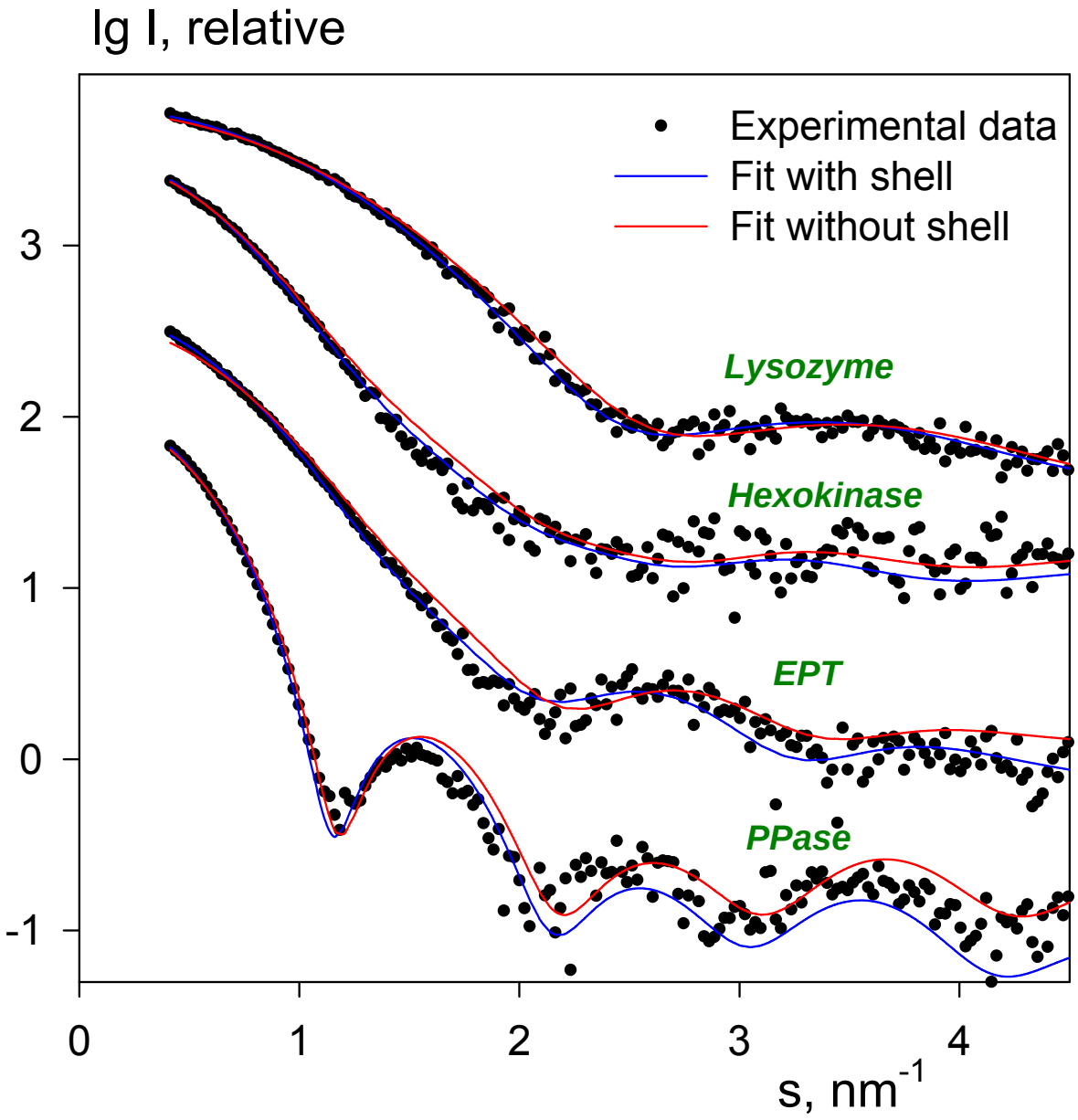
SAXS case

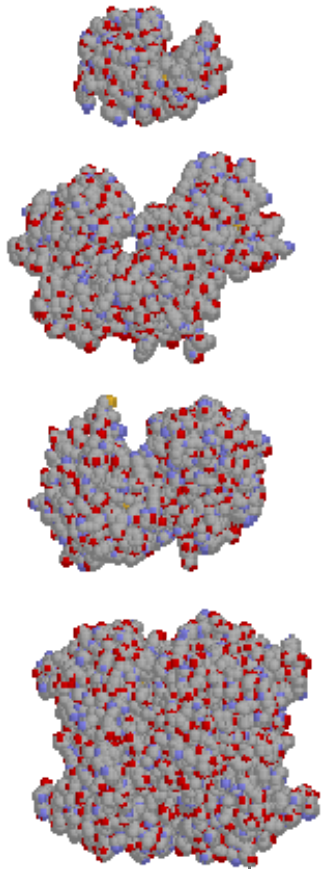
Scattering components (lysozyme)



SANS case:
50% perdeuterated lysozyme in 90% D2O

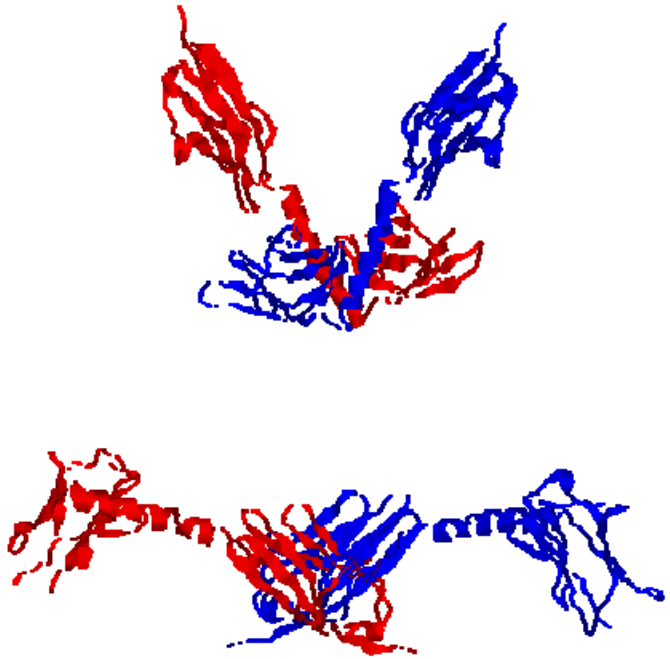
Effect of the hydration shell, X-rays

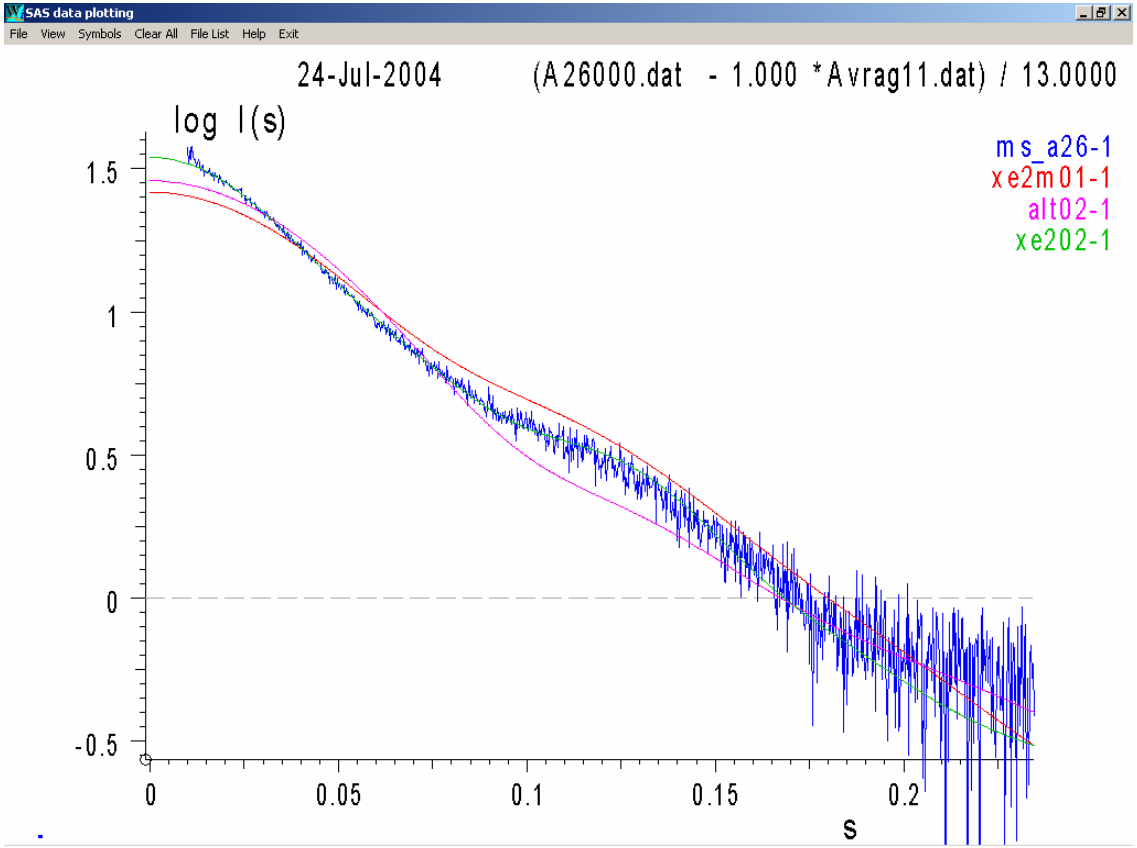




Identification of Biologically Active Oligomers

Biologically active dimer of myomesin-1

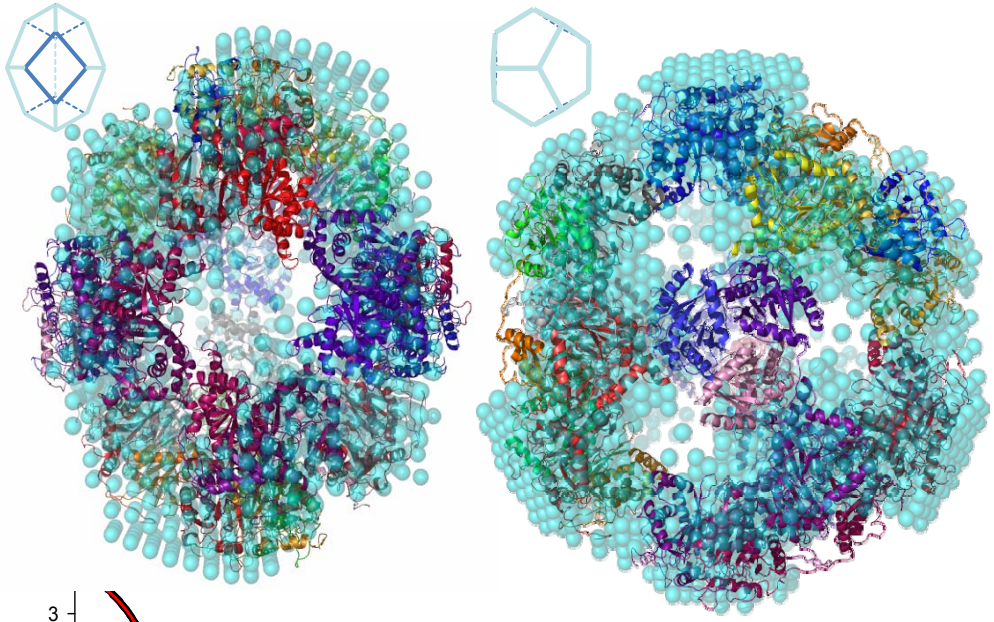


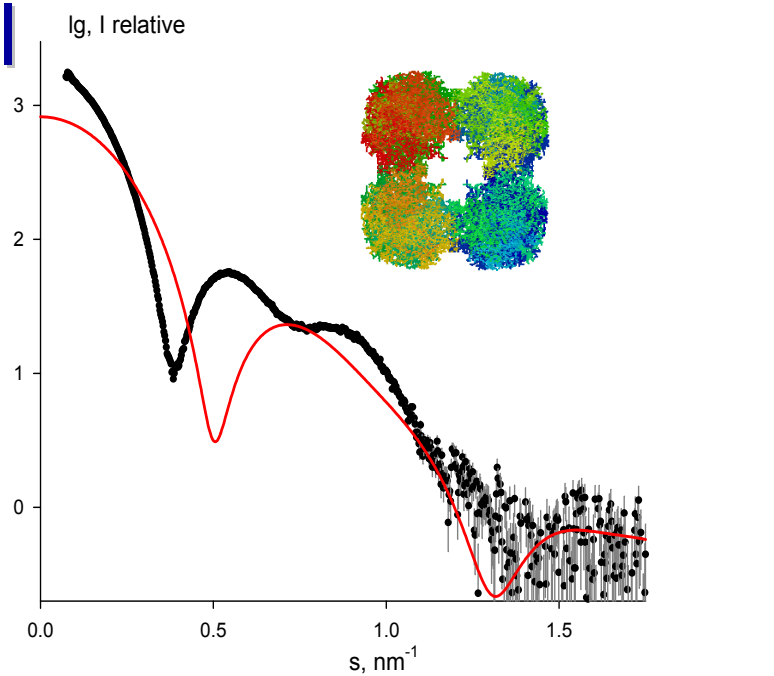


What if none of the models fits the data ?

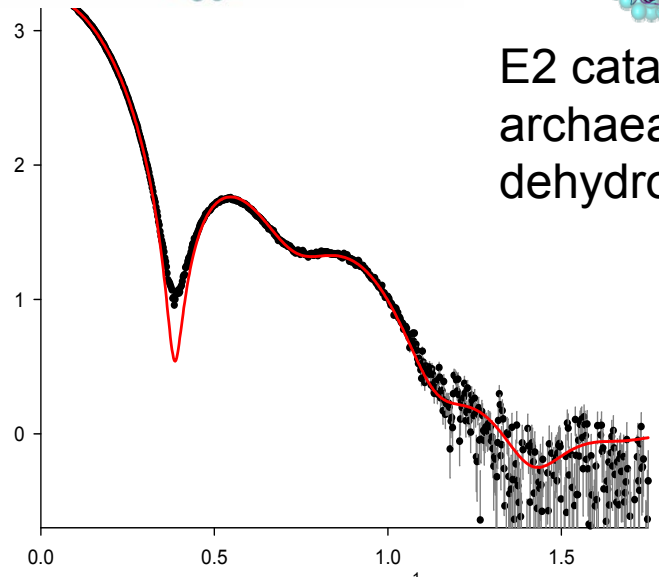
Collaboration: N. Pinotsis, S. Lange (2004)

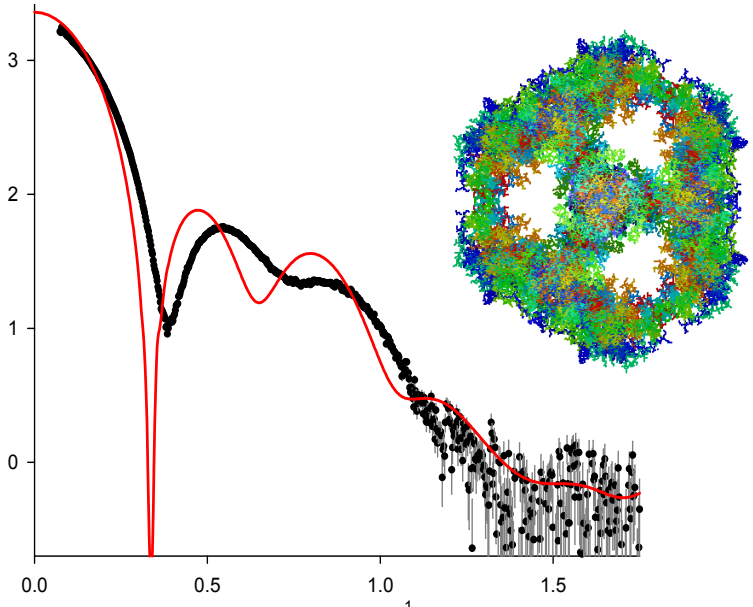
Validation of an Unusual Oligomeric Assembly in the Crystal





E2 catalytic core of an archaeal 2-oxoacid dehydrogenase

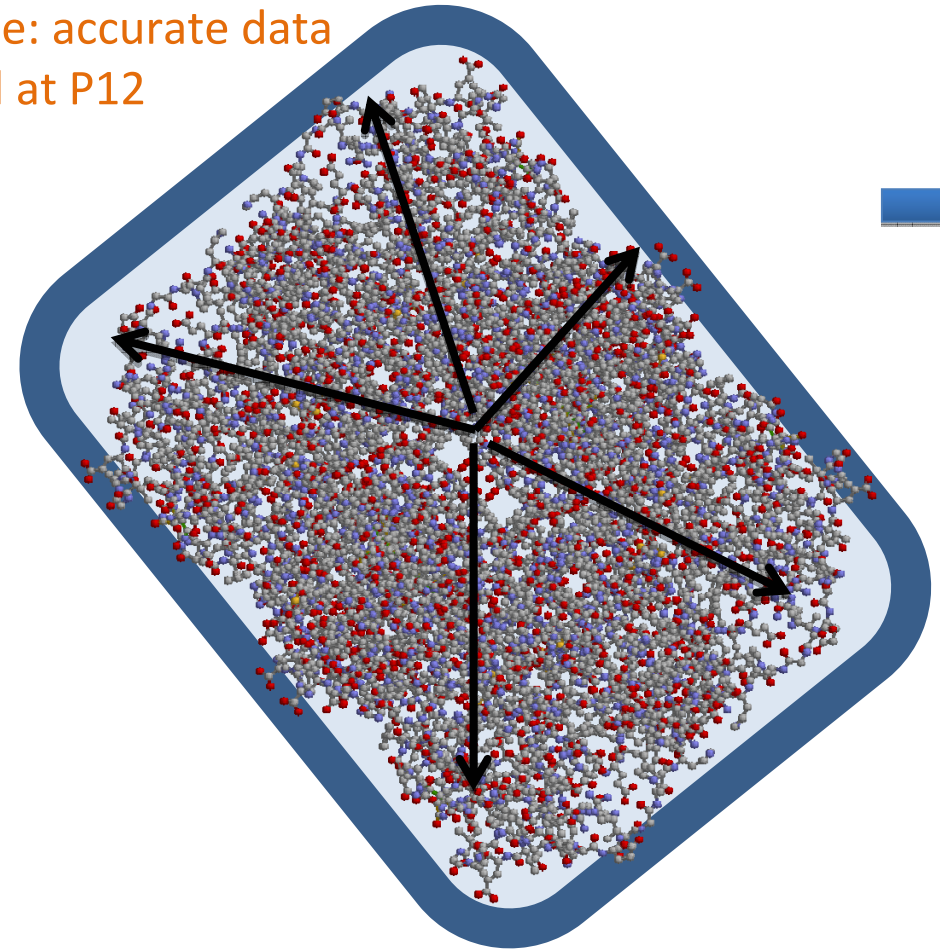




Marrott, N. L., Marshall, J. J., Svergun, D. I., Crennell, S. J., Hough, D. W., Danson, M. J. & van den Elsen, J. M. (2012). *FEBS J, 279*, 713-723

Updating CRYSOL

- Challenge: accurate data acquired at P12



Shell as envelope function

Atomic structure in vacuum

- Excluded volume

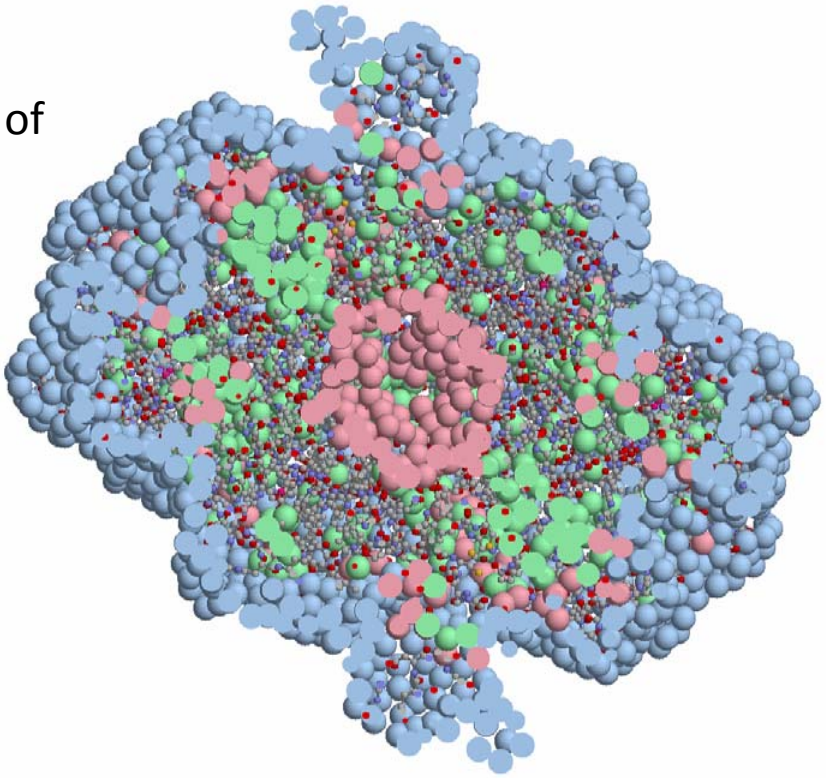
+ Hydration shell

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

CRYSOL 3.0

An essential prerequisite for reliable hybrid modeling:

Accurate computation of theoretical scattering patterns from atomic structures



Outer shell

Internal cavities

Extra excluded volume

- Test set of about 20 well-characterized proteins measured at X33 and P12 with and without HPLC (M.Graewert & D.Ruskule)
- MD simulations by A.Tuukkanen

When the atomic structure is solved… …it is still not the end of the story

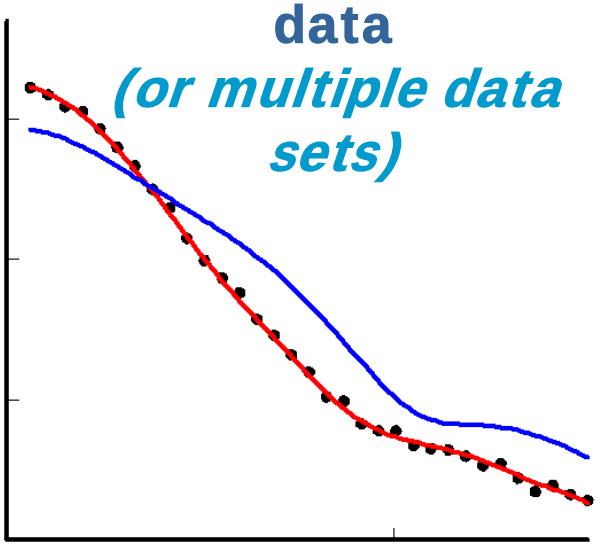
- SAXS can be applied for
  - Identification of biologically active oligomers
  - Structure validation in solution
  - Screening of multiple MD conformations

3D modelling against SAS data

Monodispersity and ideality of solution are required

General principle of SAS modelling

1D scattering data ***(or multiple data sets)***

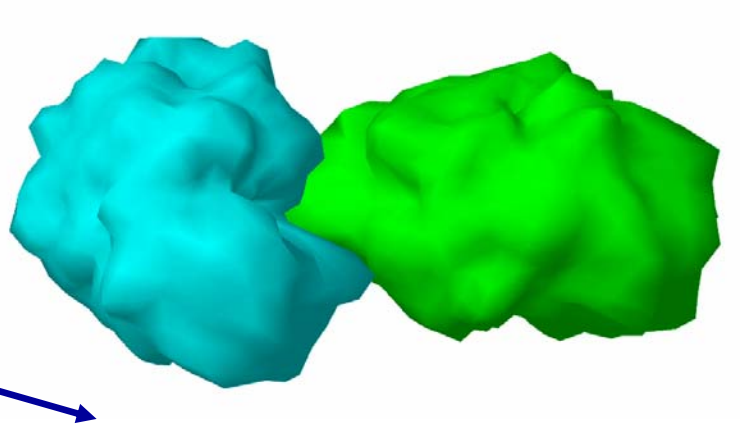


Trial-and-error

3D search model

$\boldsymbol{X} = \{X\} = \{X_1 \ldots X_M\}$

M parameters

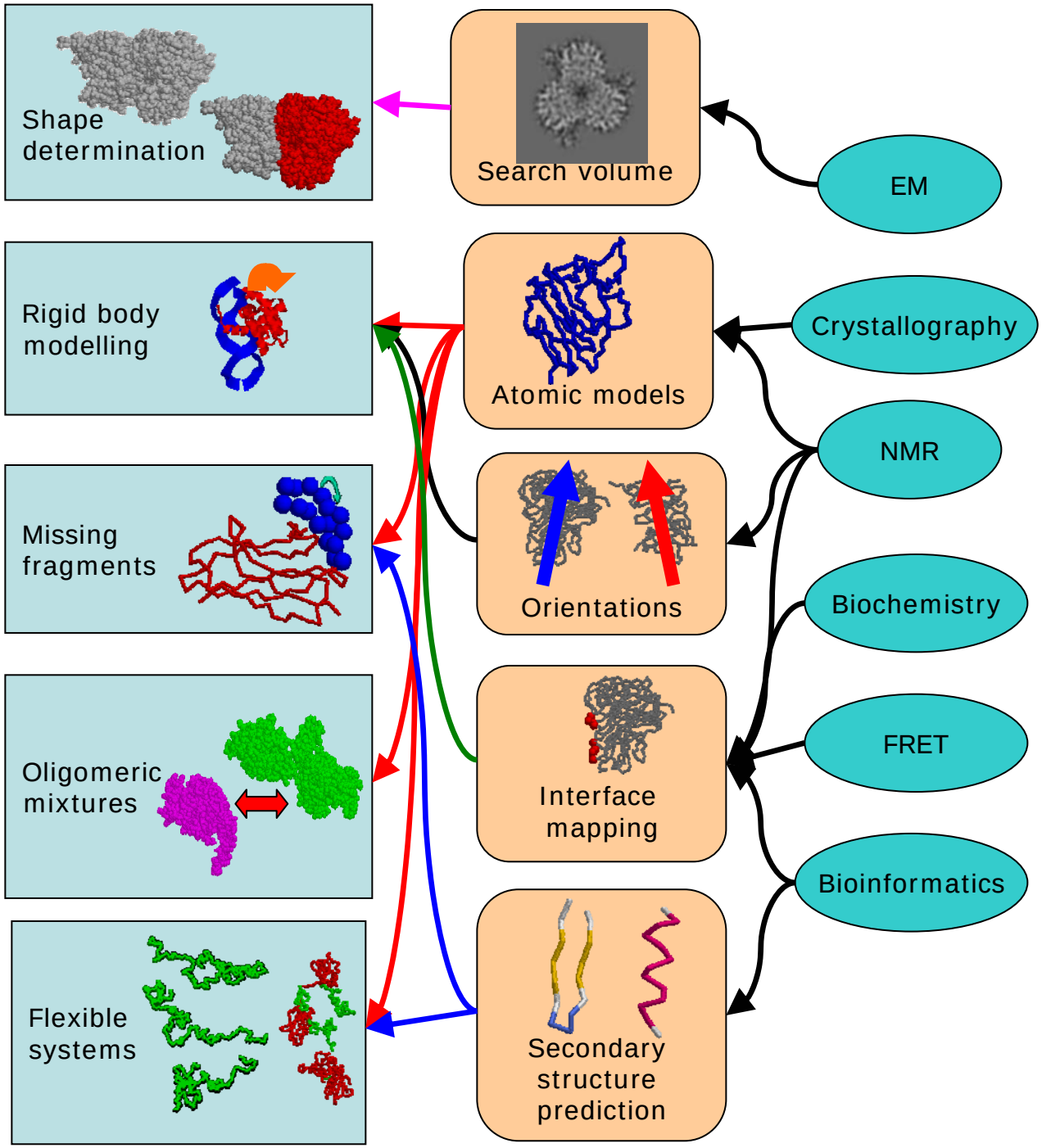


discrepancy: $\chi^2 = \frac{1}{N-1}\sum_j \left[\frac{I_{\exp}(s_j) - cI(s_j)}{\sigma(s_j)}\right]^2$

Non-linear search

Additional information is ALWAYS required to resolve or reduce ambiguity of interpretation at given resolution

Constraints & Restraints



Use of contrast variation for SAS modelling

X-rays:
Addition of sucrose or salts

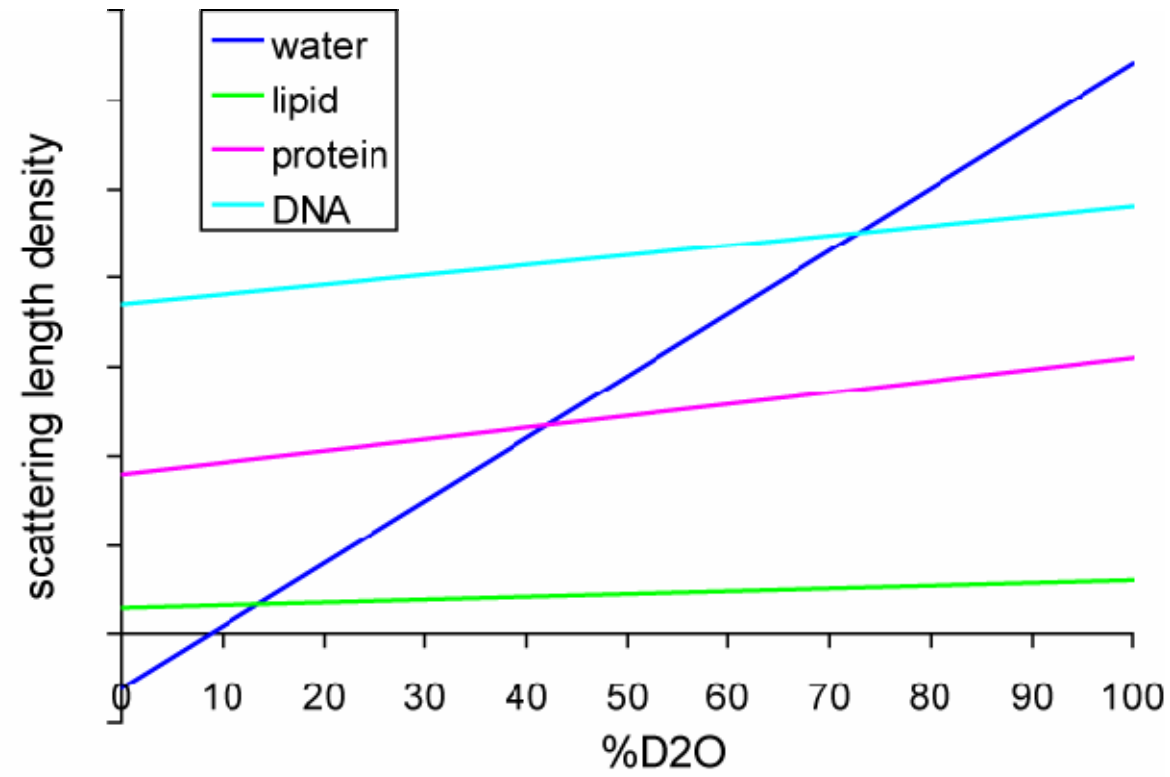
RNA, 550 e/nm^3

60% sucrose, 430 e/nm^3

Protein, 410 e/nm^3

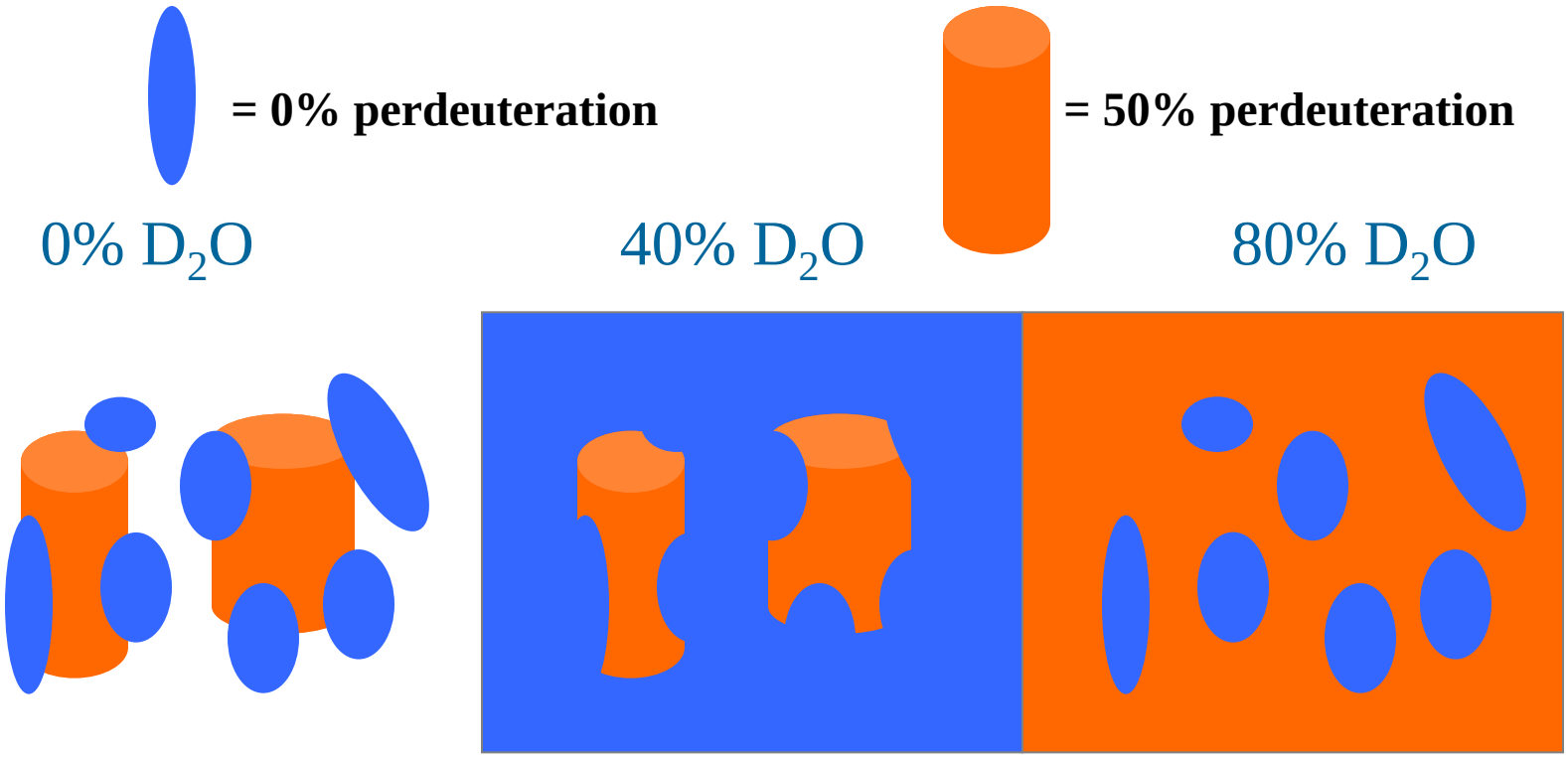
H_2O, 344 e/nm^3

Neutrons:
Isotopic H/D substitution



Use of contrast variation for SAS modelling

- Contrast variation possibilities in SANS
  - Perdeuteration of subunits
  - Different % of D_2O in the solvent



Target Function

- To reduce the ambiguity of data analysis

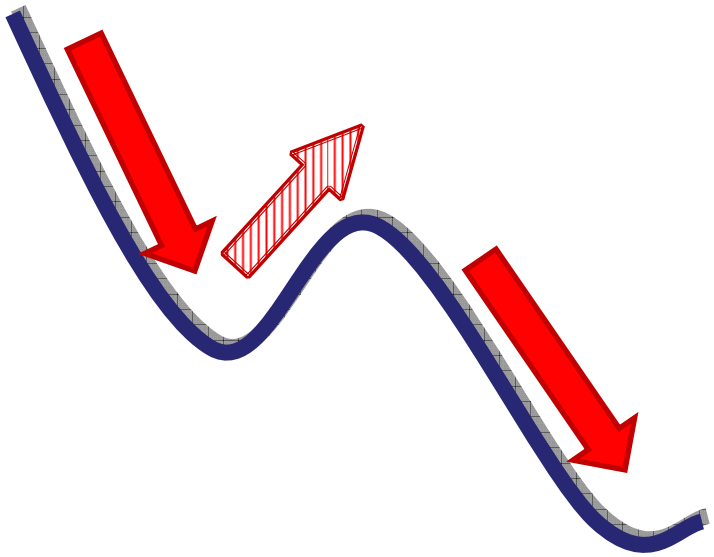
$$E(\{X\}) = \chi^2[(I(s), I_{\exp}(s)] + \sum_i \alpha_i P_i$$

  is minimized

- Penalties describe model-based restraints and/or introduce the available additional information from other methods: MX, NMR, EM etc)
- A brute force (grid) search is applied if the number of free parameters is small
- Otherwise a Monte-Carlo based technique (e.g. simulated annealing) is employed to perform the minimization of *E({X})*

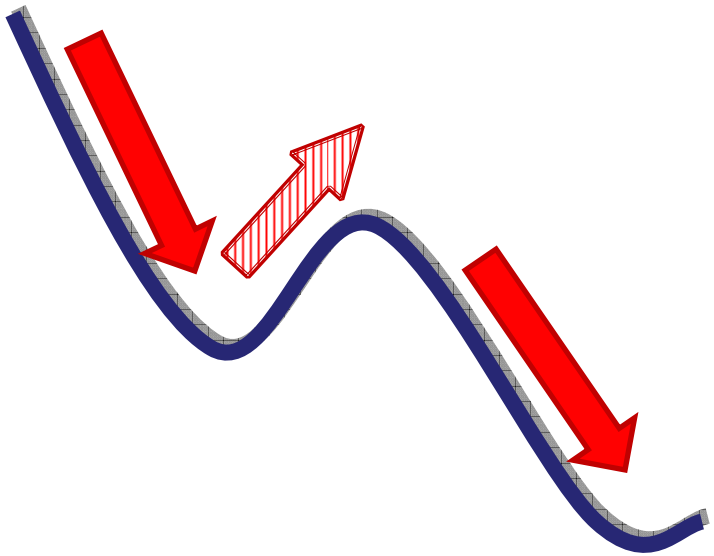
Simulated Annealing Protocol

- **Main idea:** Minimization of the target function $E(\boldsymbol{X})$ by random modifications of the system always moving to configurations that decrease $E(\boldsymbol{X})$ but to also occasionally move to configurations that increase the scoring function.



Simulated Annealing Protocol

- The probability of accepting “unprofitable” moves decreases in the course of the minimization (the system is cooled).
- At the beginning, the temperature is high and the changes almost random, whereas at the end a configuration with nearly minimum energy is reached.



Simulated Annealing Protocol

Start from some initial (e.g. random) configuration X at a high temperature T

X : $\{X_1 \ldots X_i \ldots X_j \ldots X_M\}$

X': $\{X_1 \ldots X_i' \ldots X_j' \ldots X_M\}$

$\Delta E = E(X') - E(X)$

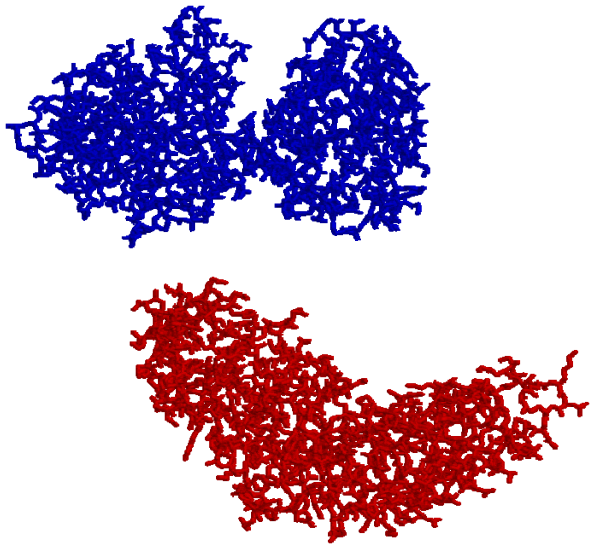
If $\Delta E<0$ move to X' else **$p = \exp (- \Delta E / T)$**

The system is cooled until no improvement is observed

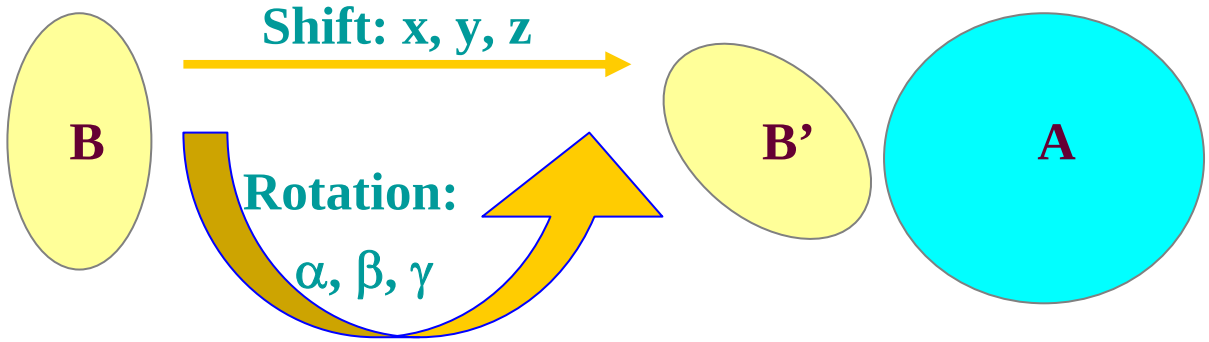
Rigid Body Modelling of Quaternary Structure: Playing with Molecular Building Blocks

Idea of rigid body modelling

- The atomic structures of the components (subunits or domains) are known.
- Assuming the tertiary structure is not changed by complex formation.
- Arbitrary complex can be constructed by moving and rotating the subunits.
- For each subunit this operation depends on three orientational and three translational parameters.



Scattering from a complex particle



The partial amplitudes of a rotated and displaced subunit are expressed *via* the initial amplitudes, three Euler rotation angles and three Cartesian shifts):

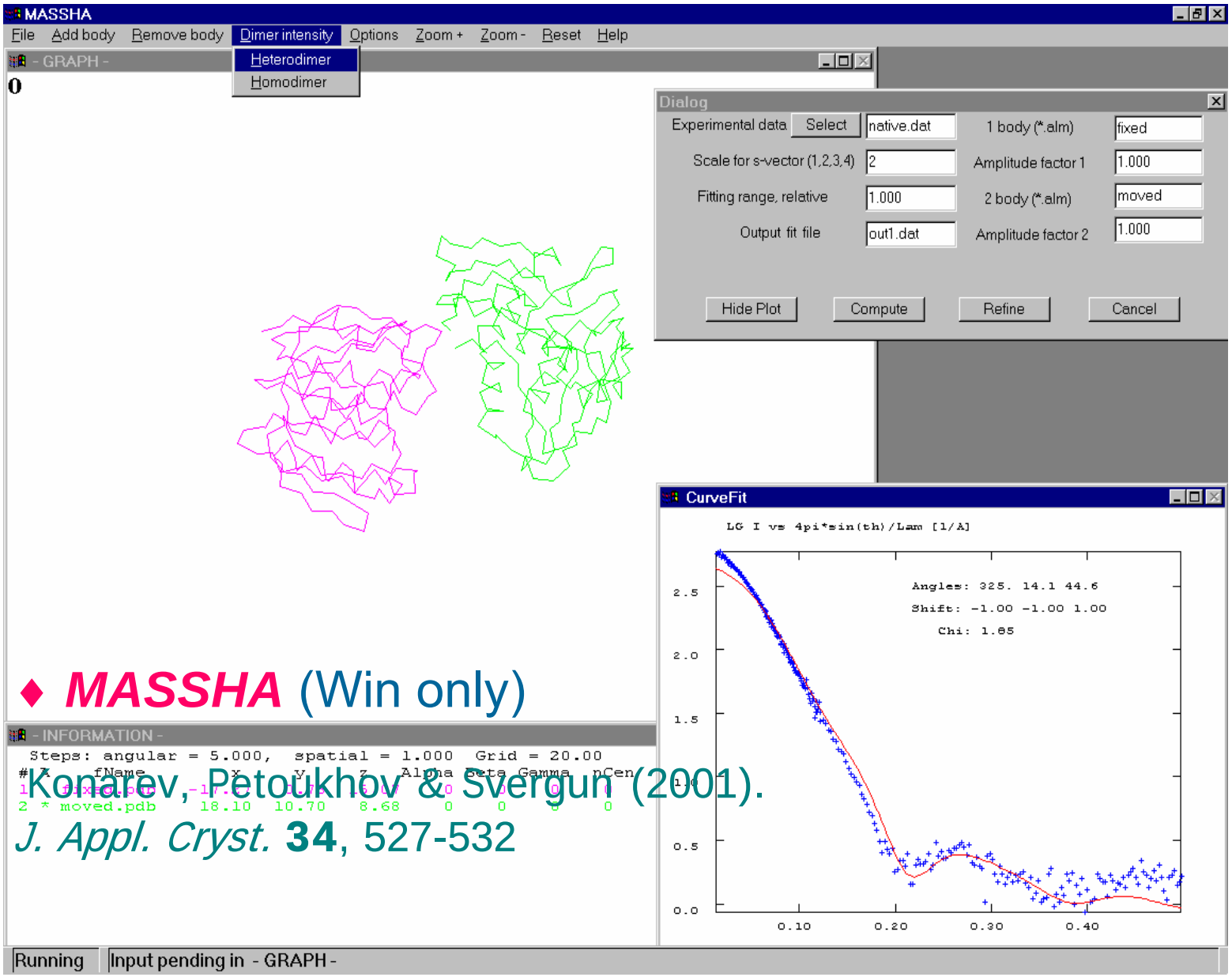
$$A^{(i)}_{lm}(s) = A^{(i)}_{lm}(s)\,\{\,A_0{}^{(i)}_{lm}(s),\ \alpha^{(i)},\ \beta^{(i)},\ \gamma^{(i)},\ x^{(i)},\ y^{(i)},\ z^{(i)}\,\}.$$

$$I(s) = 2\pi^2 \sum_{l=0}^{L} \sum_{m=-l}^{l} \left| \sum_{n} A^{n}{}_{lm}(s) \right|^2$$

For symmetric particles, there are fewer parameters and the calculations are faster

Svergun, D.I. (1991). *J. Appl. Cryst.* **24**, 485-492

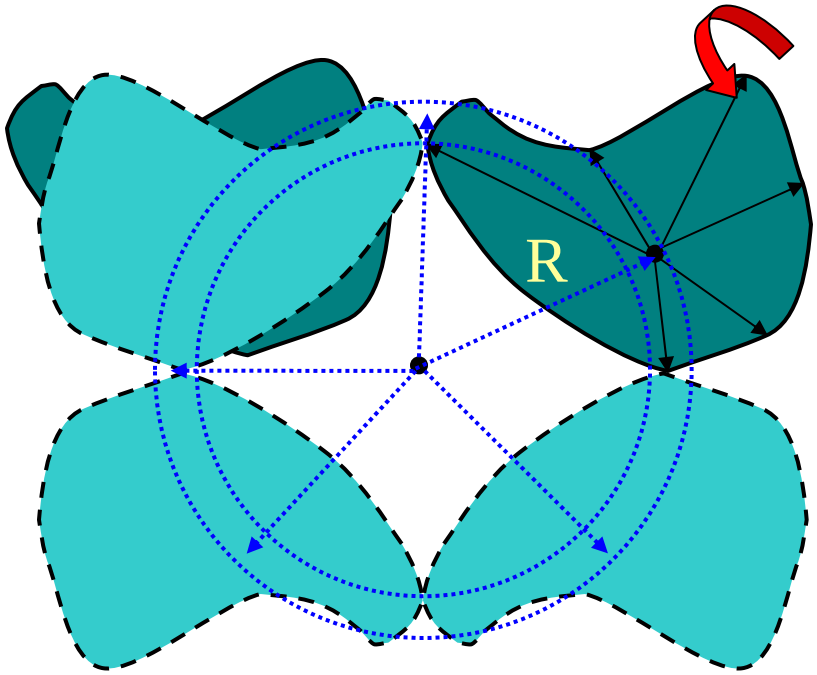
Interactive and automated local refinement (grid search)



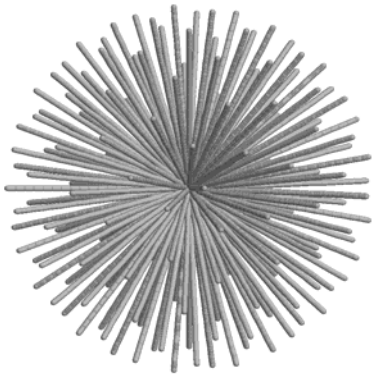
♦ ***MASSHA*** (Win only)

Konarev, Petoukhov & Svergun (2001).
J. Appl. Cryst. **34**, 527-532

GLOBSYMM: rigid body modelling of symmetric oligomers



Position / orientation of a single monomer defines the quaternary structure of oligomer, which therefore described by 4 or 6 parameters



quasi-uniform angular grids for positioning and rotations

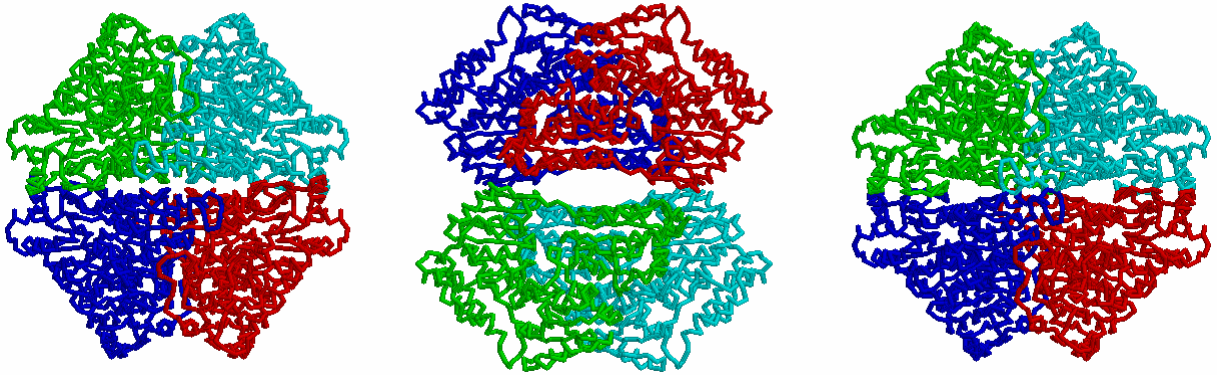
$$< R > \approx \sqrt{(R_g^{\exp})^2 - (R_g^{mon})^2}$$

estimate for the shift of monomer center from the origin

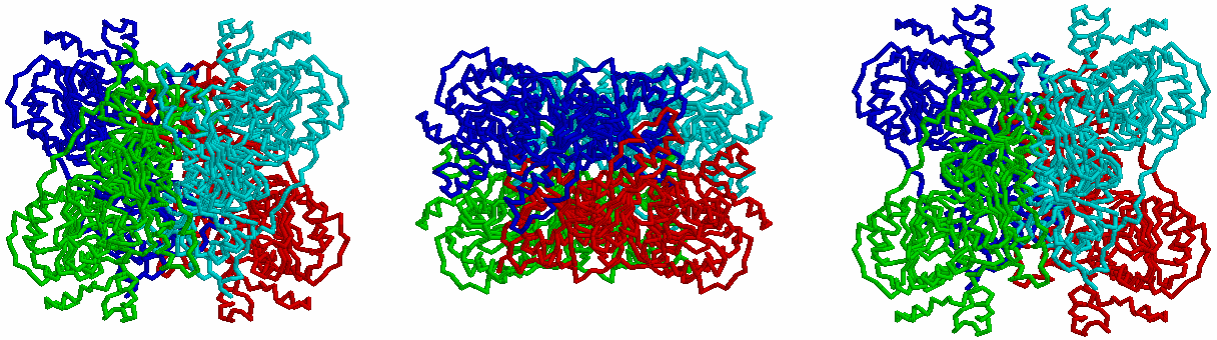
Petoukhov M.V., Svergun, D.I. (2005). *Biophys. J.* **89**, 1237-1250

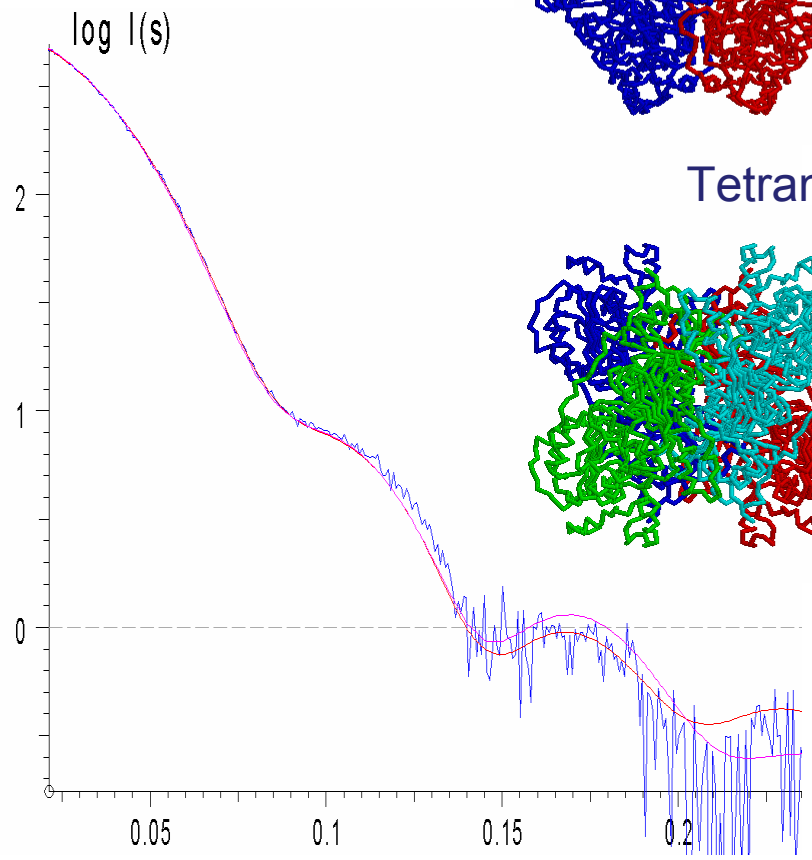
Limitations of the Grid Search

Xtal



Tetrameric pyruvate oxidase from yeast

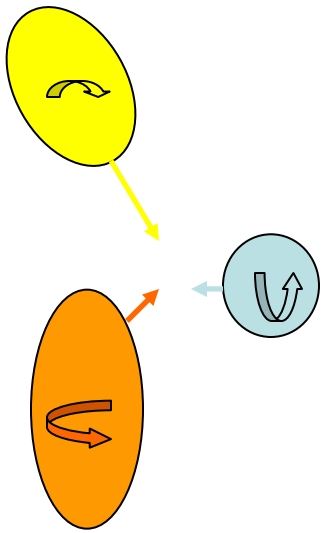




	GLOBSYMM, P222	
Grid order	12	13
χ	1.7	1.6
r.m.s.d.	31Å	5Å
CPU	~3.5 h	~9 h

Collaboration: S. König
(Martin-Luther-Universität, Halle)

Rigid Body Modelling of Multisubunit Complexes



- Start from arbitrary initial positions and orientations of the subunits
- Simulated annealing is employed
- Search of interconnected spatial arrangement of the subunits without clashes
- Random movement/rotation at one SA step
- Fitting the scattering data by minimizing the target function

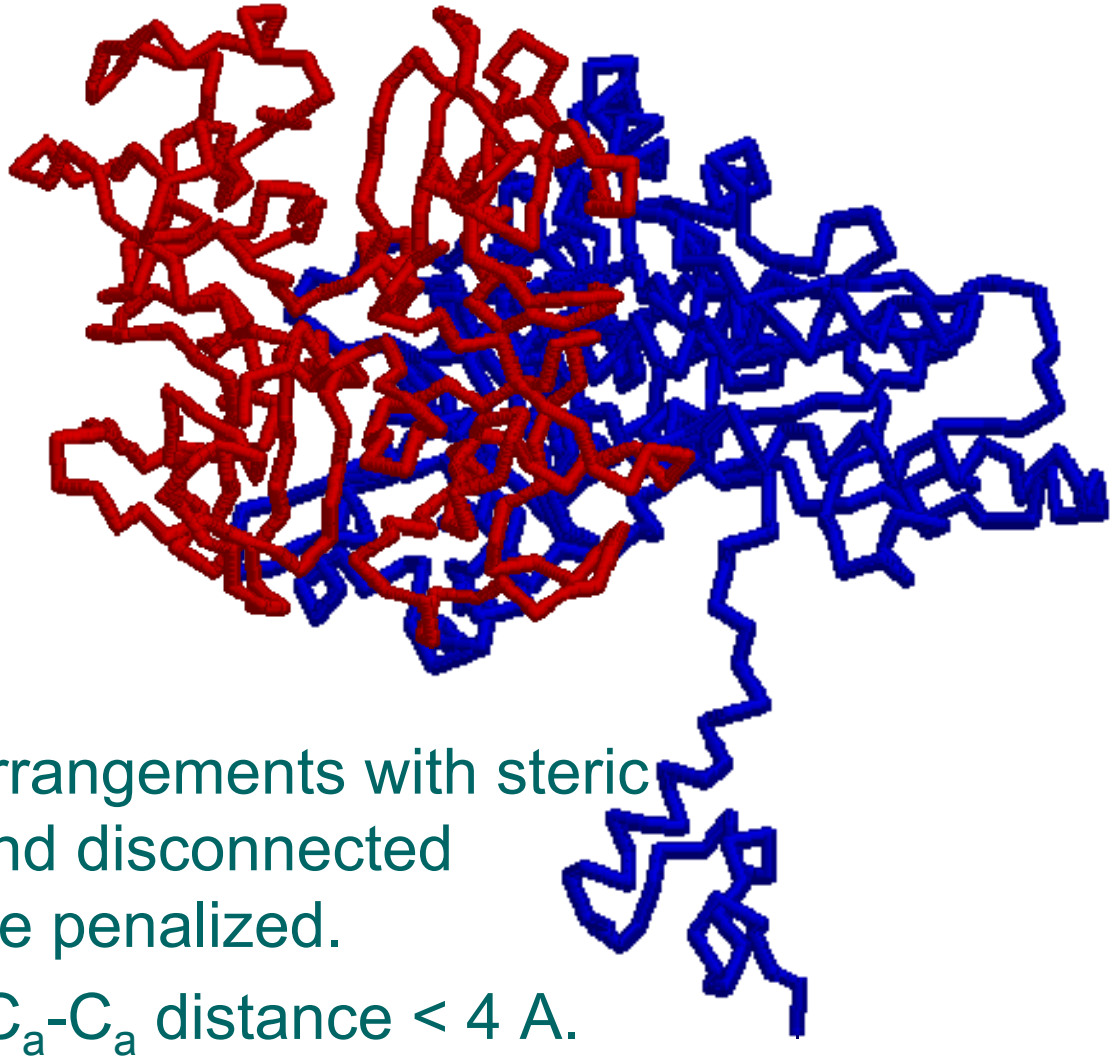
$$E(X) = \Sigma\chi^2[I_{exp}(s), I(X,s)] + \Sigma\alpha_i P_i(X)$$

- Additional restraints may be applied

Petoukhov, M. V., and Svergun, D. I. (2006). ***Eur Biophys J.***, 35, 567-576

SASREF restraints

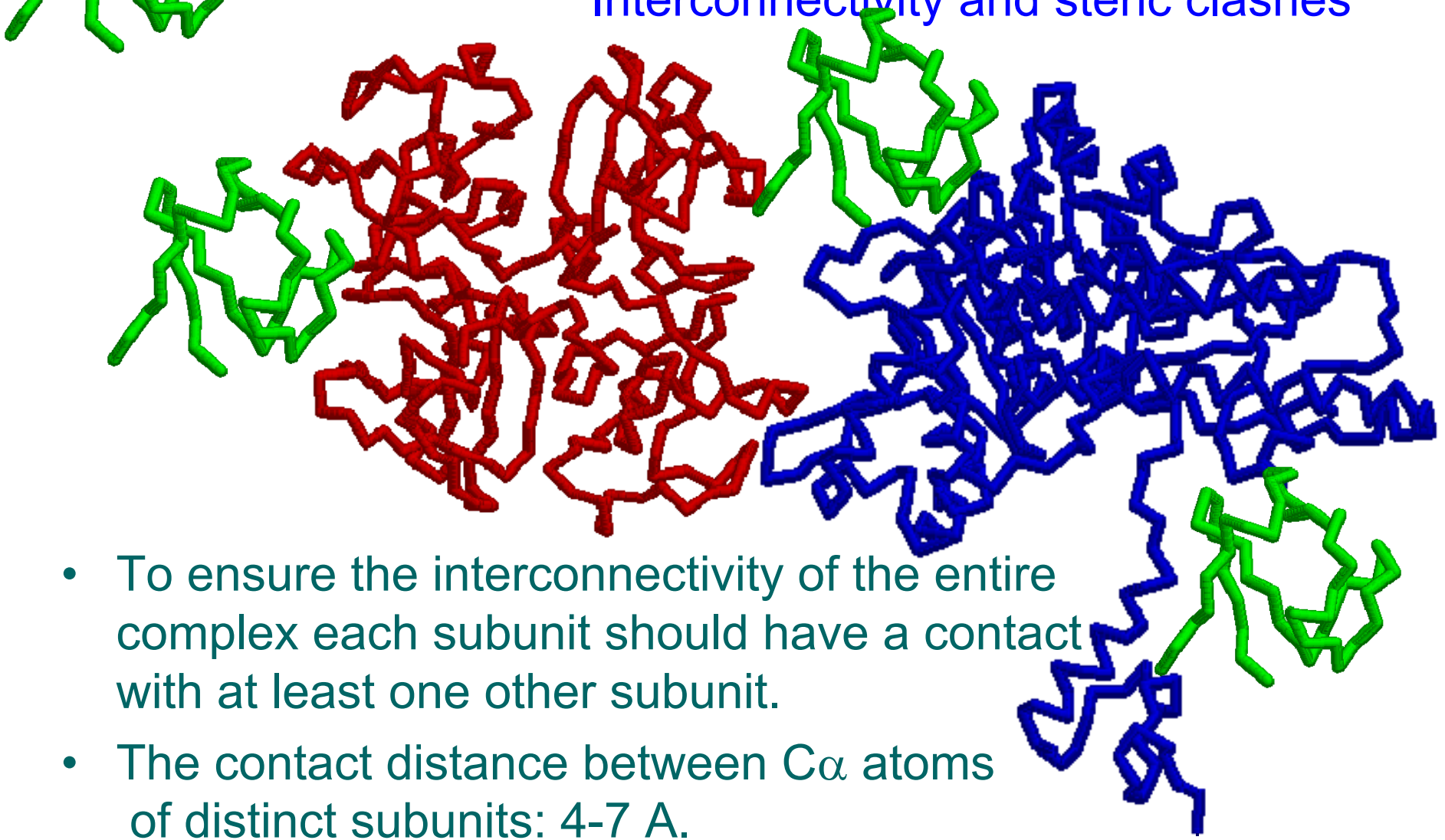
Interconnectivity and steric clashes



- Subunit arrangements with steric clashes and disconnected models are penalized.
- Overlap: C_a-C_a distance < 4 A.

SASREF restraints

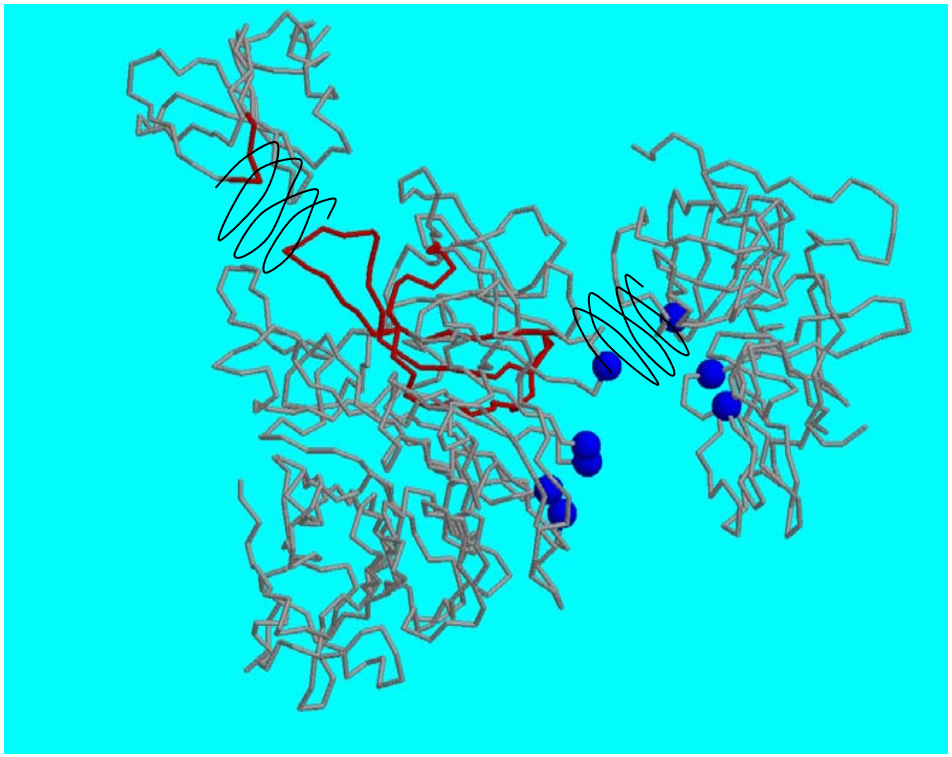
Interconnectivity and steric clashes



- To ensure the interconnectivity of the entire complex each subunit should have a contact with at least one other subunit.
- The contact distance between Cα atoms of distinct subunits: 4-7 A.

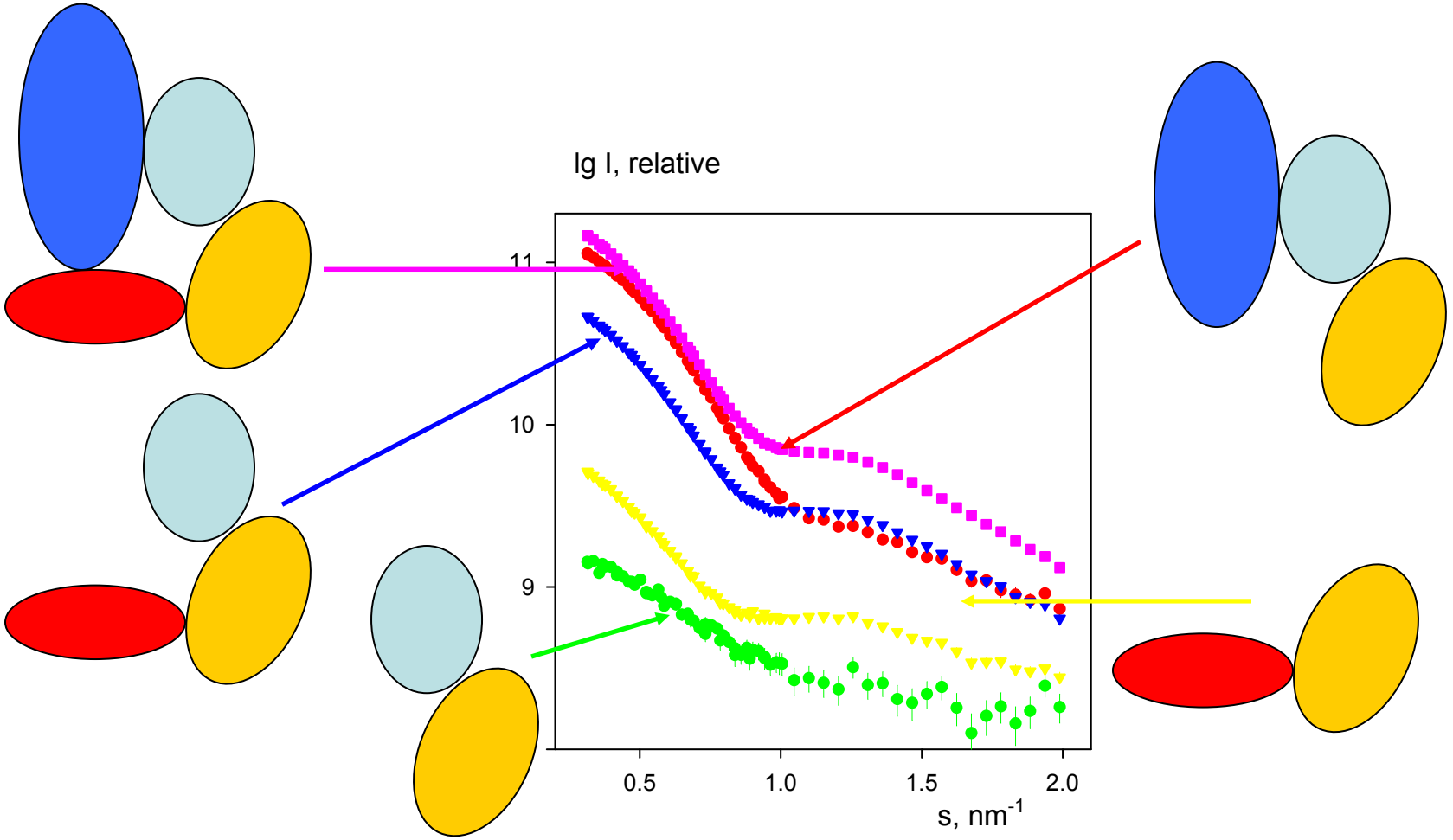
Contacts restraints

- From binding affinity studies or from mutagenesis data the information on contacting subunits and even individual residues can be available.
- Such information is accounted for by specifying the ranges of residues or nucleotides which can be involved in interactions between the partners.
- Spring force potentials are added as penalties



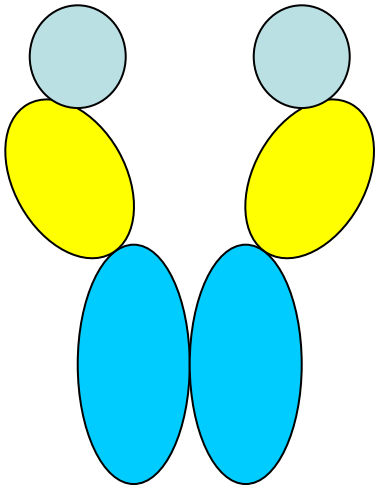
Use of multiple data sets from partial constructs

Simultaneous fitting of multiple scattering curves



Further SASREF options

Symmetry constraint



Groups Pn / Pn2 (n=1..6), P23, P432 and icosahedral symmetry can be taken into account.

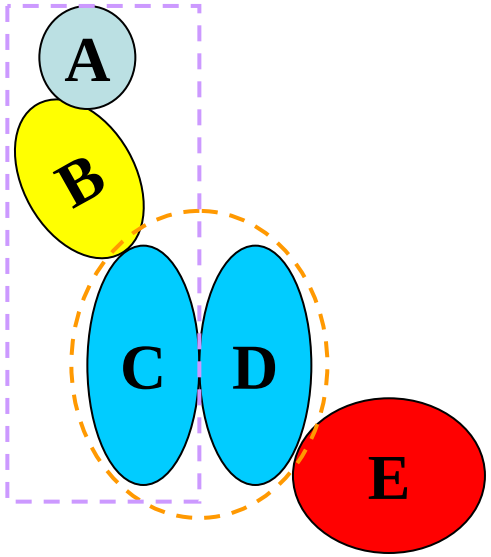
- fewer spatial parameters to describe the model
- selection rules for the partial amplitudes:

  m equal to 0 or multiples of n,

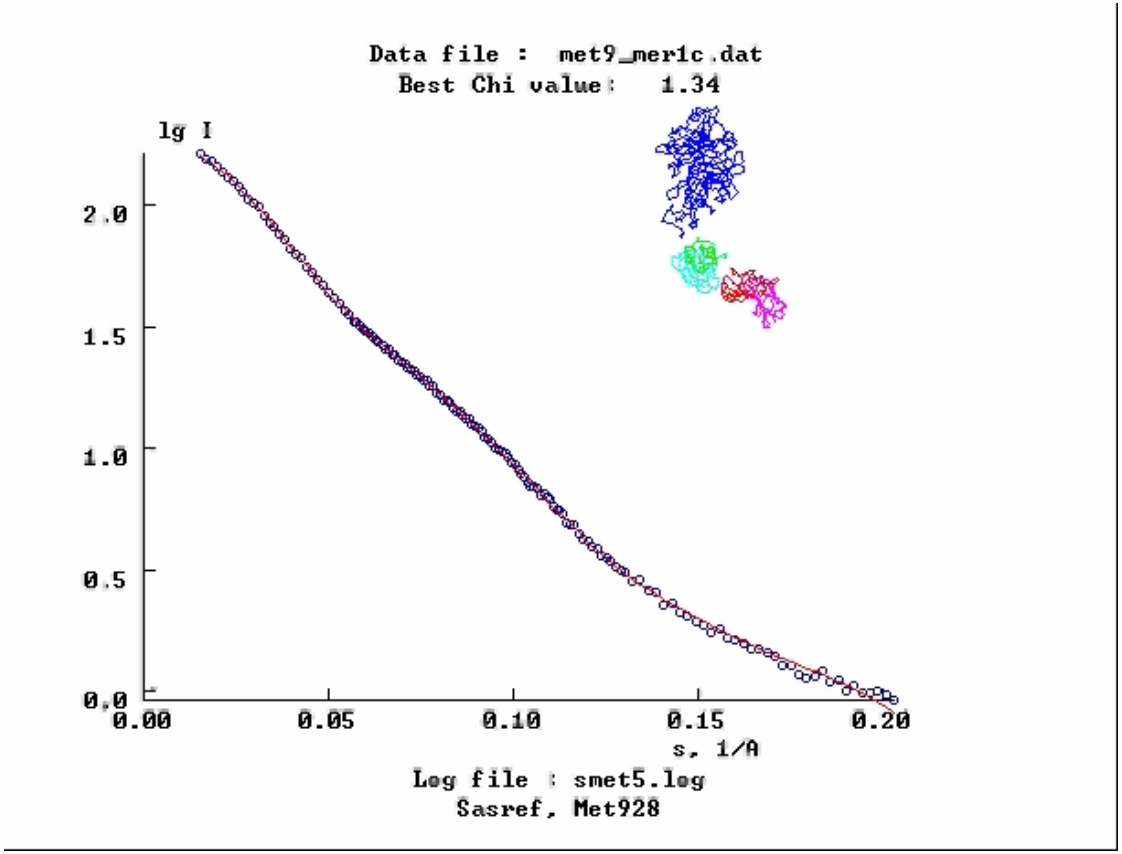
  for Pn2, terms of order *l0* with odd *l* and all imaginary parts vanish

Fixation of subset

Some subunits can be fixed at the initial positions and orientations to keep their mutual arrangement



SASREF run


Data file : met9_mer1c.dat
Best Chi value: 1.34
lg I
2.0
1.5
1.0
0.5
0.0
0.00
0.05
0.10
0.15
0.20
s, 1/A
Log file : smet5.log
Sasref, Met928

ATSAS-online

Web interface for **SASREF** --- Page 1 of 3

Main parameters of the project	
Total number of **curves**	6
Total number of **subunits**	3
Overall **symmetry**	P1

SUBMIT

Web interface for **SASREF** --- Page 2 of 3
no. of **subunits:** 3 - no. of **curves:** 6 - overall symmetry: **P1**

Curve	File	D2O fraction	Symmetry	Angular units $4\pi sin(\theta)/\lambda$	Fraction to fit	Setting	Weight	Use a constant?
1	C:\project1\xcomp.dat Browse...	-1.00	P1	Å-1	1.00	0	1.00	No
2	C:\project1\x-12.dat Browse...	-1.00	P1	Å-1	1.00	0	1.00	No
3	C:\project1\x-23.dat Browse...	-1.00	P1	Å-1	1.00	0	1.00	No
4	C:\project1\nc_0.dat Browse...	0	P1	Å-1	1.00	1	1.00	Yes
5	C:\project1\nc_p50_0.d Browse...	0	P1	Å-1	1.00	1	1.00	Yes
6	C:\project1\nc_p50_10(Browse...	1.00	P1	Å-1	1.00	2	1.00	Yes

Subunit	File	Shift?	Fix?	Symmetry
1	C:\project1\sub1.pdb Browse...	Yes	No	P1
2	C:\project1\sub2.pdb Browse...	Yes	Yes	P1
3	C:\project1\sub3.pdb Browse...	Yes	No	P1

Optional files:

C:\project1\smear.res Browse... Smearing parameters (*.res)
C:\project1\contacts.cn(Browse... Contacts conditions (*.cnd)

SUBMIT

Web interface for **SASREF** --- Page 3 of 3

Perdeuterations of the subunits in each construct
(specify "**-1.0**" if the subunit does not appear in the construct)

	sub1.pdb	sub2.pdb	sub3.pdb
xcomp.dat	0.00	0.00	0.00
x-12.dat	0.00	0.00	-1.0
x-23.dat	-1.0	0.00	0.00
nc_0.dat	0.00	0.00	0.00
nc_p50_0.dat	0.00	0.5	0.00
nc_p50_100.dat	0.00	0.5	0.00

SUBMIT

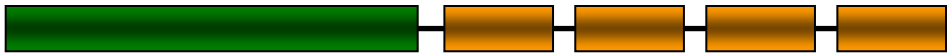
http://www.embl-hamburg.de/ExternalInfo/Research/Sax/atsas-online

Combining SAXS & SANS

Combining SAXS and SANS: complex of MET and the bacterial ligand InlB

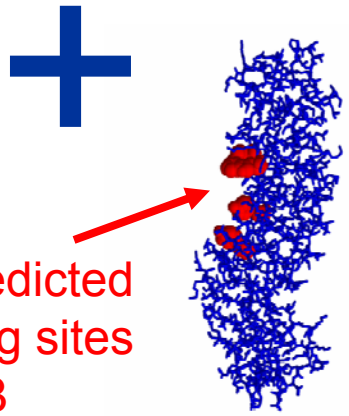
3 constructs of receptor tyrosine kinase Met

- Met9



- Met8

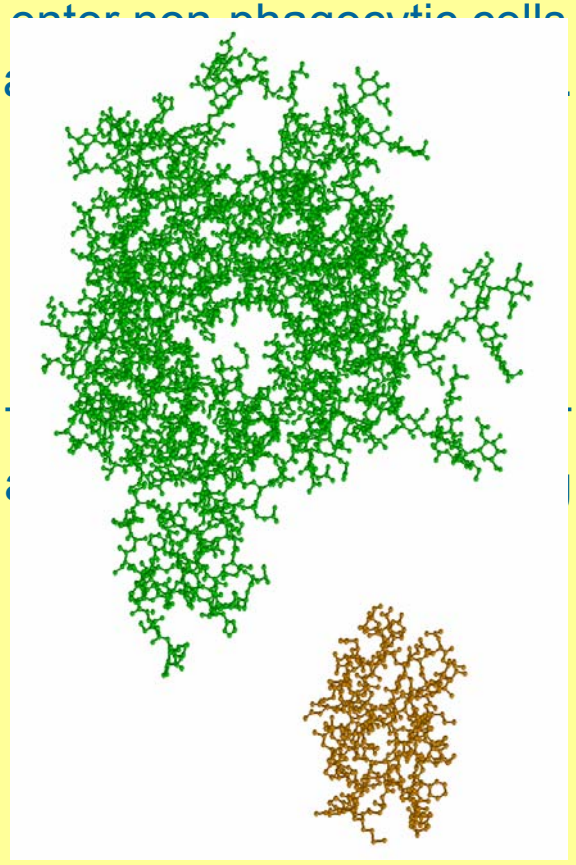
- Met5



30 kDa

Human pathogen *Listeria monocytogenes* is able to

60 kDa



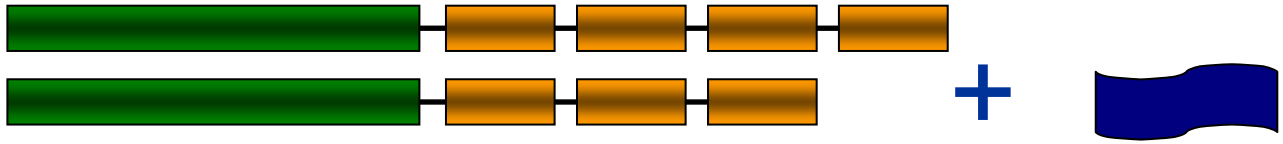
10 kDa

Collaboration:
H.Niemann (Braunschweig) and P.Timmins (Grenoble)

Scattering data used for the modelling

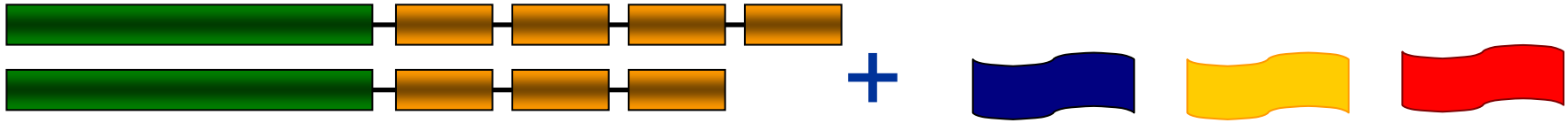
X-ray scattering

- Met8, Met9 and complexes with InlB — 4 curves



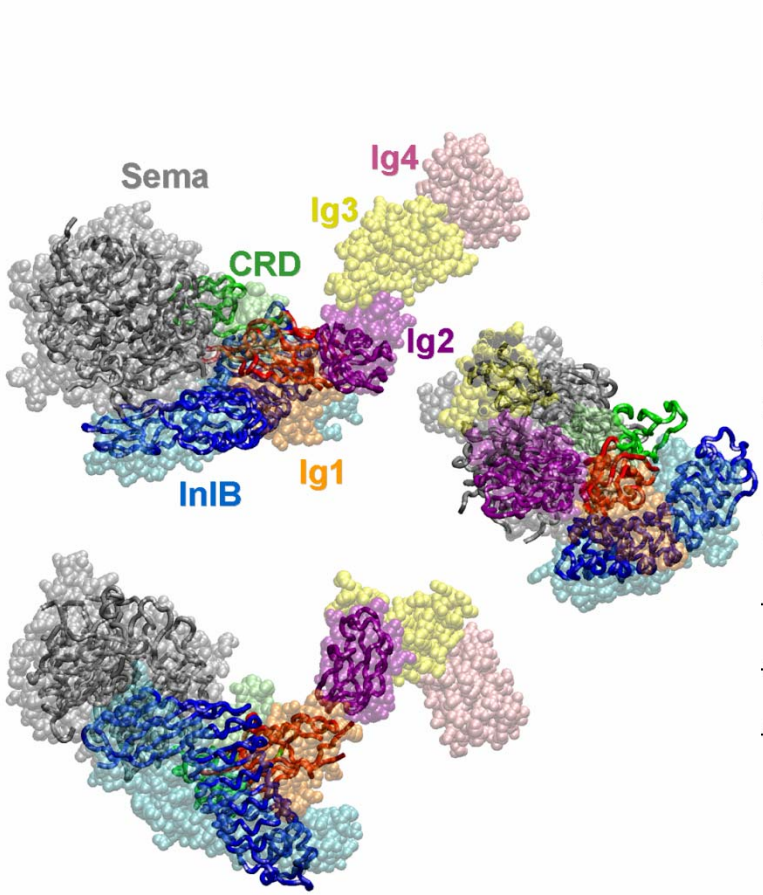
Neutron scattering

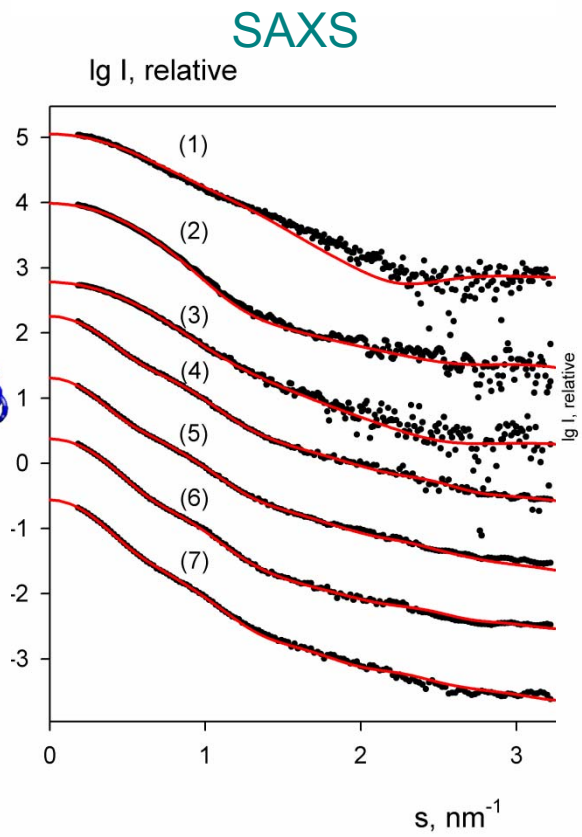
- Met8, Met9 and complexes with fully protonated and 50 and 100 % deuterated InlB in 0, 35, 50, 60, 80, 100% D_2O — 27 curves

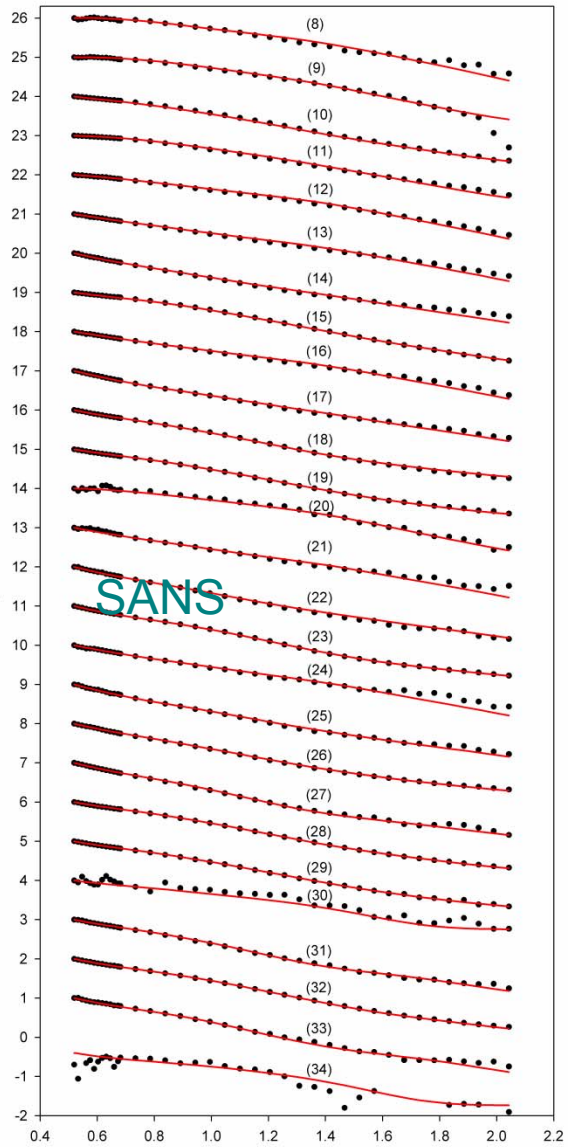


TOTAL (fitted simultaneously) — ***31 curves***

SAS rigid body & Xtal models of MET–InlB complex



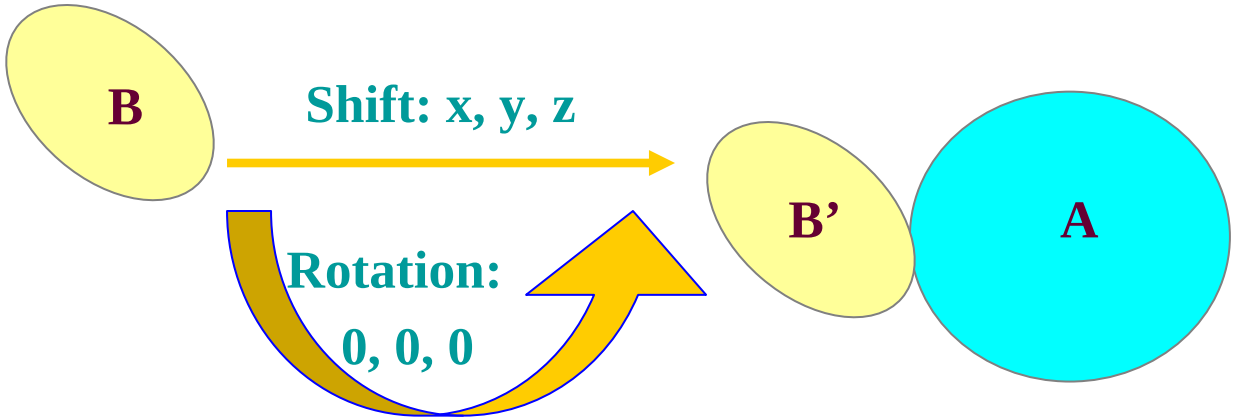




Niemann, H., Petoukhov, M.V., Härtlein, M., Moulin, M., Gherardi, E., Timmins, P., Heinz, D.W. & Svergun, D.I. (2008). ***JMB*** *377*, 489-500.

Combining SAXS & NMR

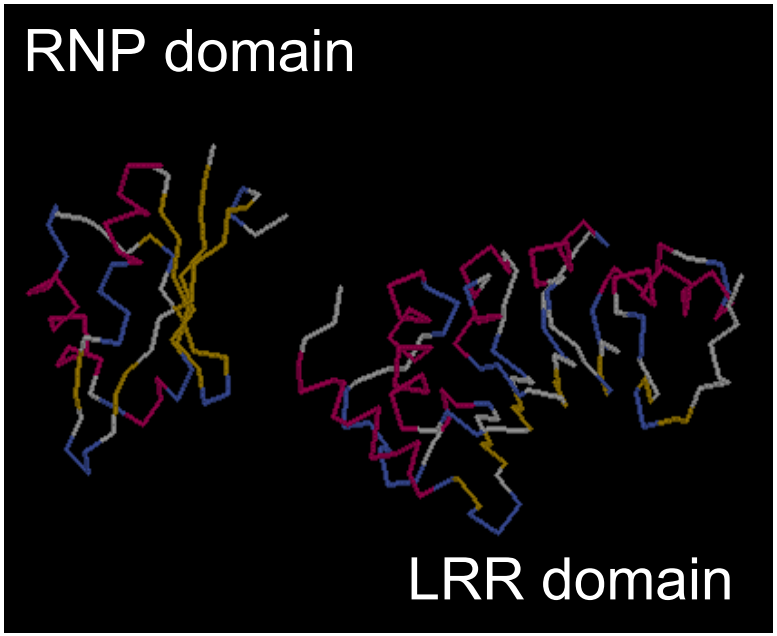
The use of RDC's to reduce orientational ambiguity of rigid body modelling

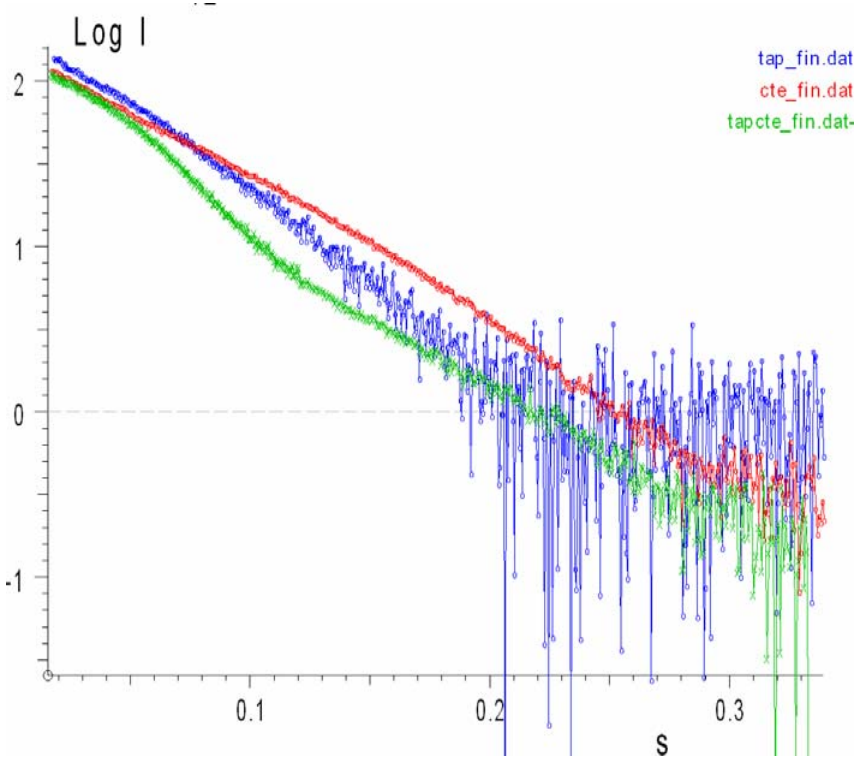


- Relative orientations of subunits derived from RDCs are kept unchanged accounting for four-fold orientation degeneracy
- Other subunits may rotate and move arbitrarily

TAP-CTE complex

Complex with 67 nucleotides CTE RNA





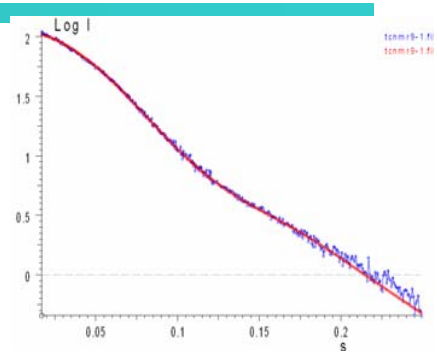
Crystal structure of the RNA-binding domain of the mRNA export factor TAP (PDB entry 1ft8)

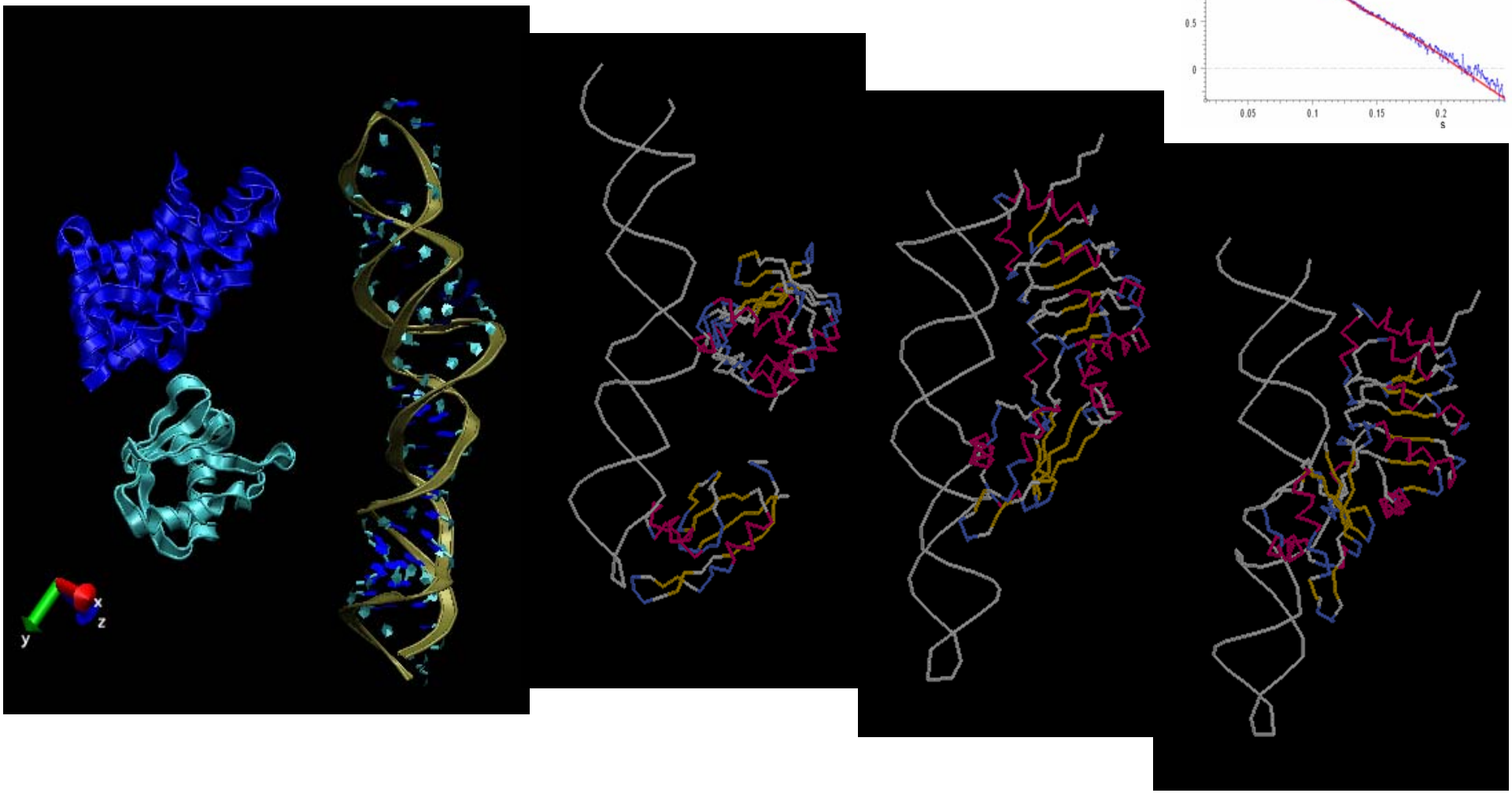
SAXS patterns from TAP, CTE and complex

Liker, E., Fernandez, E., Izaurralde, E. & Conti, E. (2000) *EMBO J*, **19**, 5587

Collaboration: F. Gabel (IBS, Grenoble)

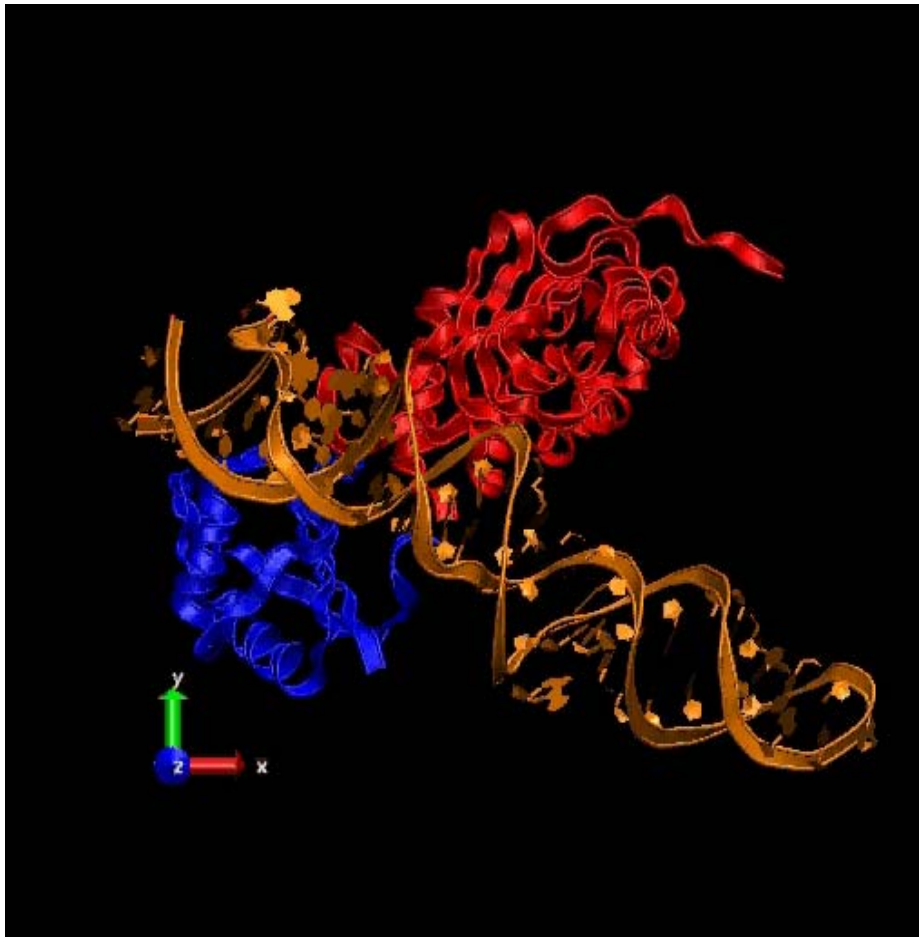
TAP-CTE complex



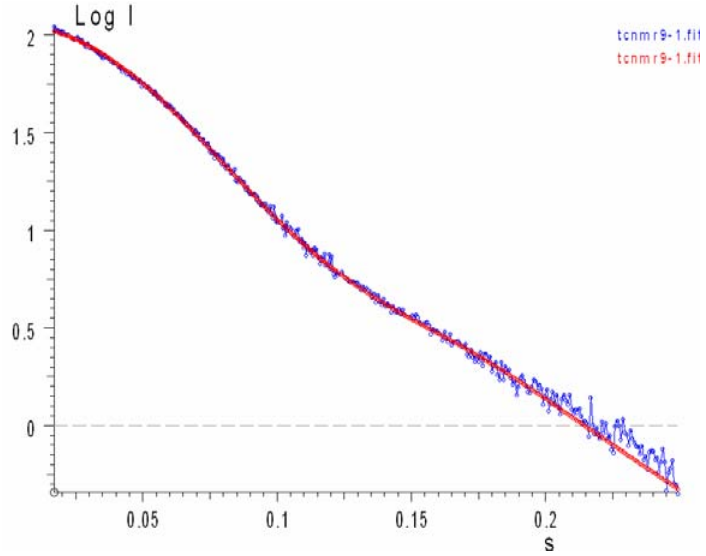


Unconstrained rigid body modeling using three bodies (separate RNP and LRR domains of TAP plus tentative model of CTE) yields ambiguous results, all well fitting the scattering from the complex

TAP-CTE complex



Only one of the four possible RDC configurations of TAP (namely, the one closest to the Xtal structure) allows one to fit the data without steric clashes



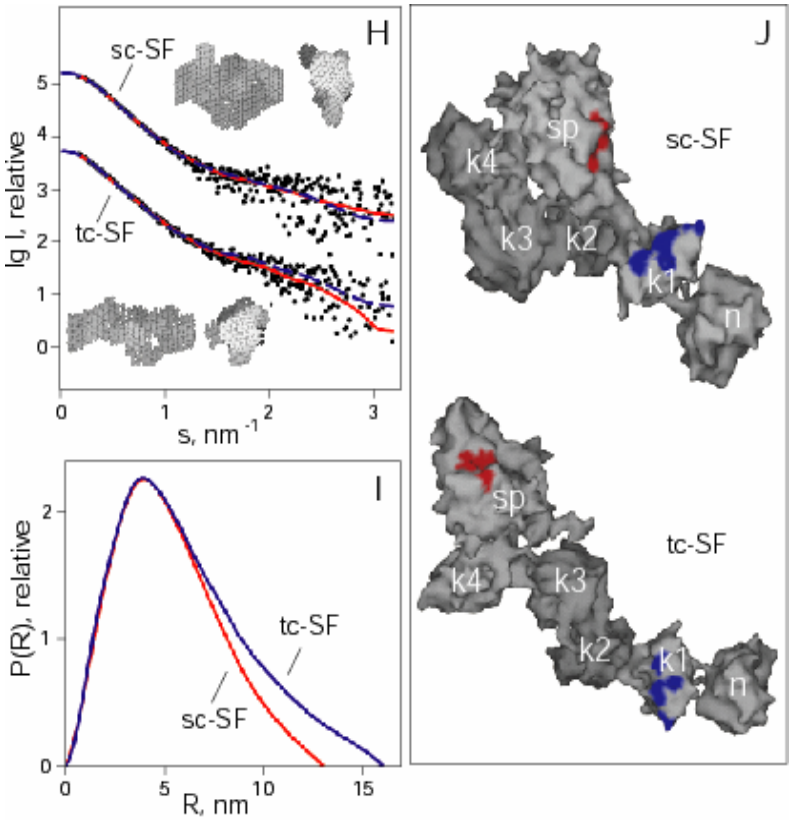
Constrained rigid body modeling using the relative orientation of RNP and LRR domains of TAP obtained by RDC and contact information between TAP and CTE from chemical shift perturbations and mutagenesis data

Combining SAXS & EM

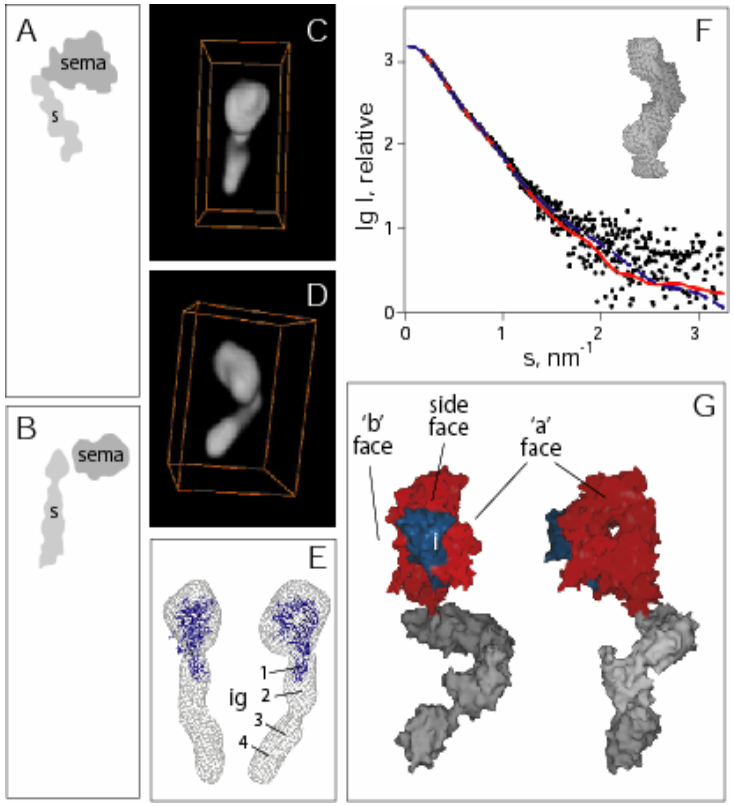
Hepatocyte growth factor/scatter factor and MET signalling

HGF/SF (6 structural domains) controls the growth of epithelial cells through the receptor tyrosine kinase MET (5 structural domains).

Conversion of pro-HGF/SF into the active two-chain form involves transition from a closed to open conformation



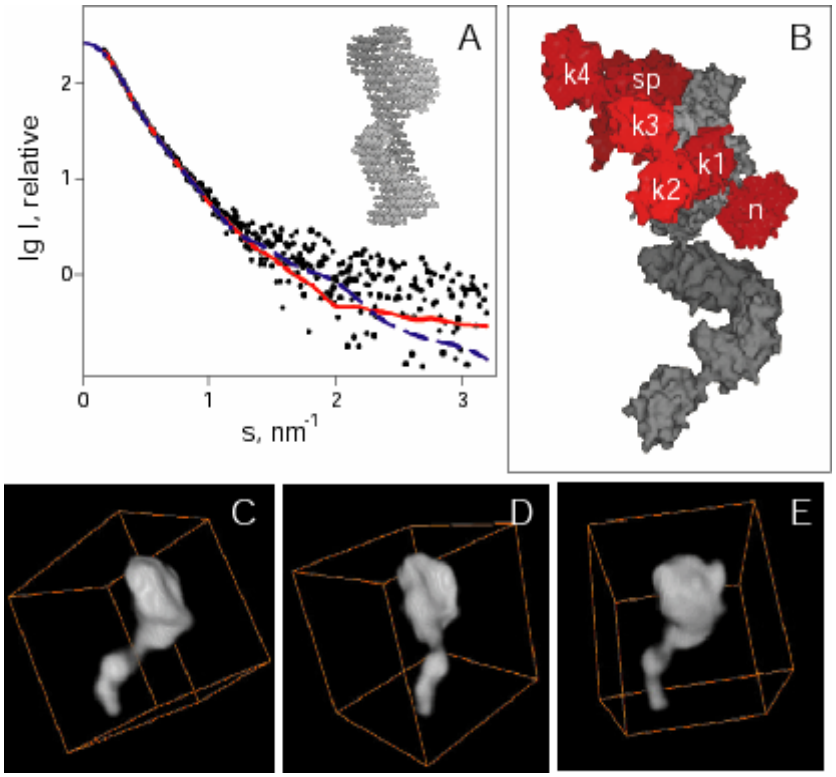
Structure of MET ectodomain (MET928) determined by SAXS and cryo-electron tomography



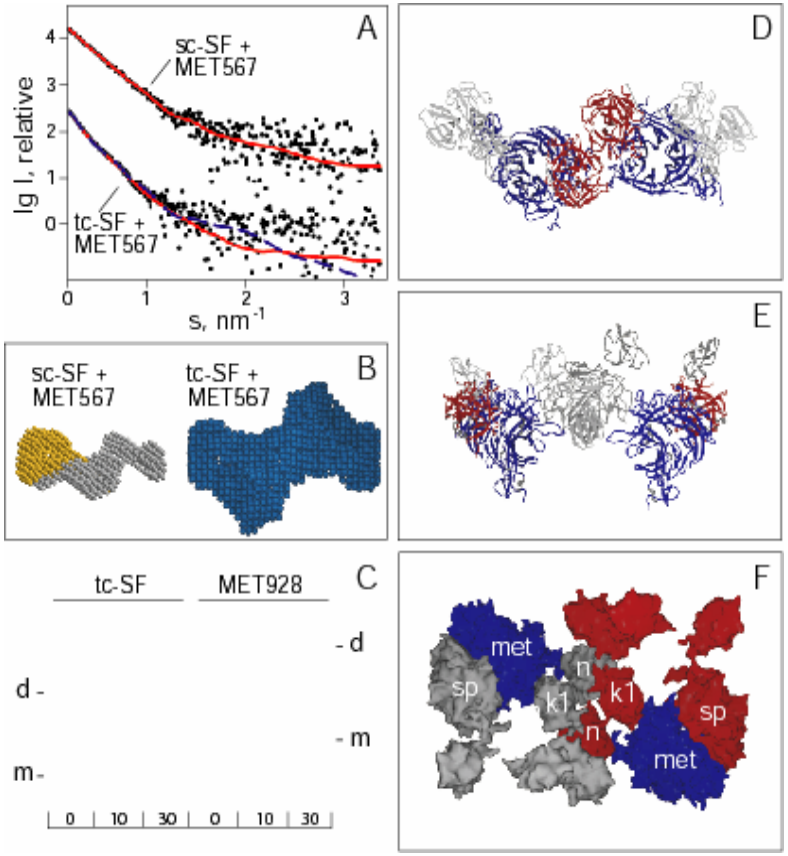
Gherardi, E., Sandin, S., Petoukhov, M.V., Finch, J., Öfverstedt L.-G., Miguel, R.N., Blundell, T.L. Woude, G.V., Skoglund, U. & Svergun, D. I. (2006) P.N.A.S. USA (accepted)

Hepatocyte growth factor/scatter factor and MET signalling

Two-chain HGF/SF forms a 1:1 complex with MET928 where the former whaps around the 7-blade β-propeller domain of MET928



Single- and two-chain HGF/SF form a 1:1 and a functional 2:2 complex with truncated MET567, respectively



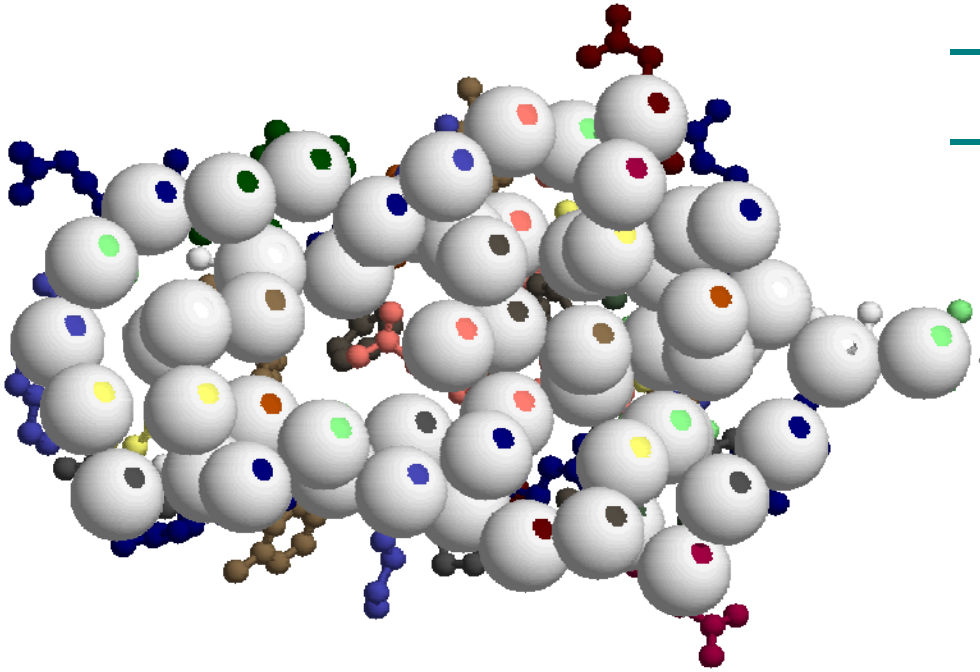
Gherardi, E., Sandin, S., Petoukhov, M.V., Finch, J., Öfverstedt L.-G., Miguel, R.N., Blundell, T.L. Woude, G.V., Skoglund, U. & Svergun, D. I. (2006) P.N.A.S. USA (accepted)

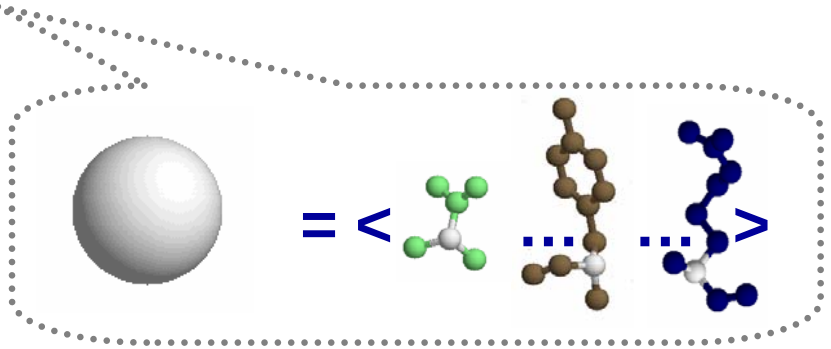
Modelling of multidomain proteins

MRGSHHHHHH GSGVPSRVIH IRKLPIDVTE GEVISLGLPF GKVTNLLMLK
GKNQAFIEMN TEEAANTMVN YYTSVTPVLR GQPIYIQFSN HKELKTDSSP
NQARAQAALQ AVNSVQSGNL ALAASAAAVD AGMAMAGQSP VLRIIVENLF
YPVTLDVLHQ IFSKFGTVLK IITFTKNNQF QALLQYADPV SAQHAKLSLD
GQNIYNACCT LRIDFSKLTS LNVKYNNDKS RDYTRPDLPS GDSQPSLDQT
MAAAFGLSVP NVHGALAPLA IPSAAAAAAA AGRIAIPGLA GAGNSVLLVS
NLNPERVTPQ SLFILFGVYG DVQRVKILFN KKENALVQMA DGNQAQLAMS
HLNGHKLHGK PIRITLSKHQ NVQLPREGQE DQGLTKDYGN SPLHRFKKPG

Concept of Dummy Residues

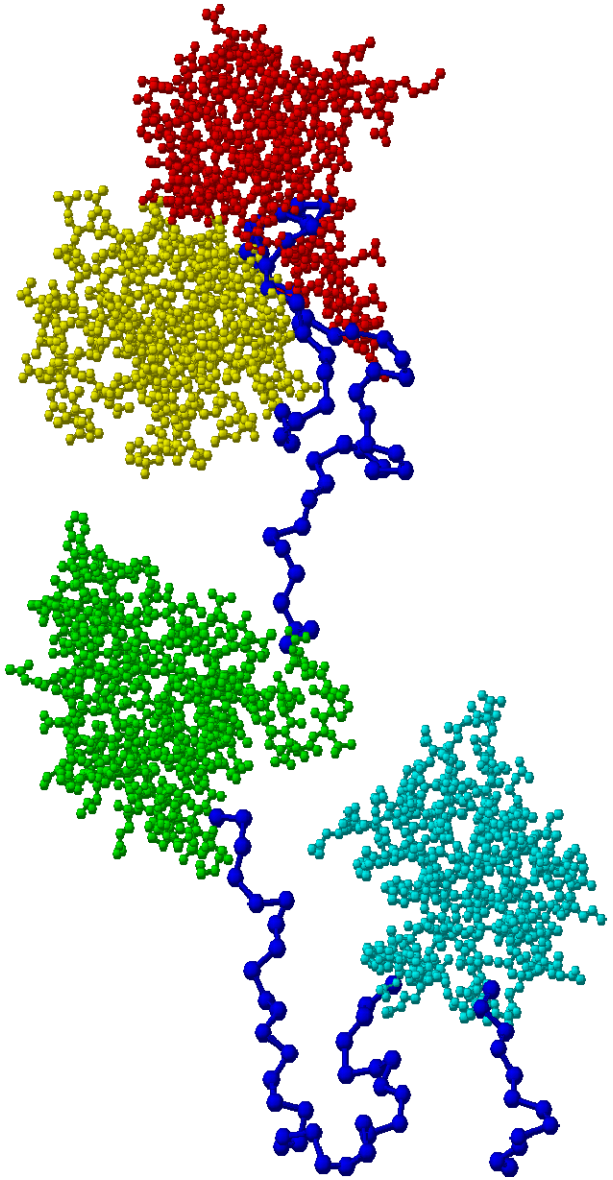
- Proteins are (folded) polypeptide chains composed of amino acids
- At a resolution of ~1 nm each amino acid can be represented as one entity (dummy residue)
- For simplicity DRs are
  - Identical
  - Centered at the Cα positions





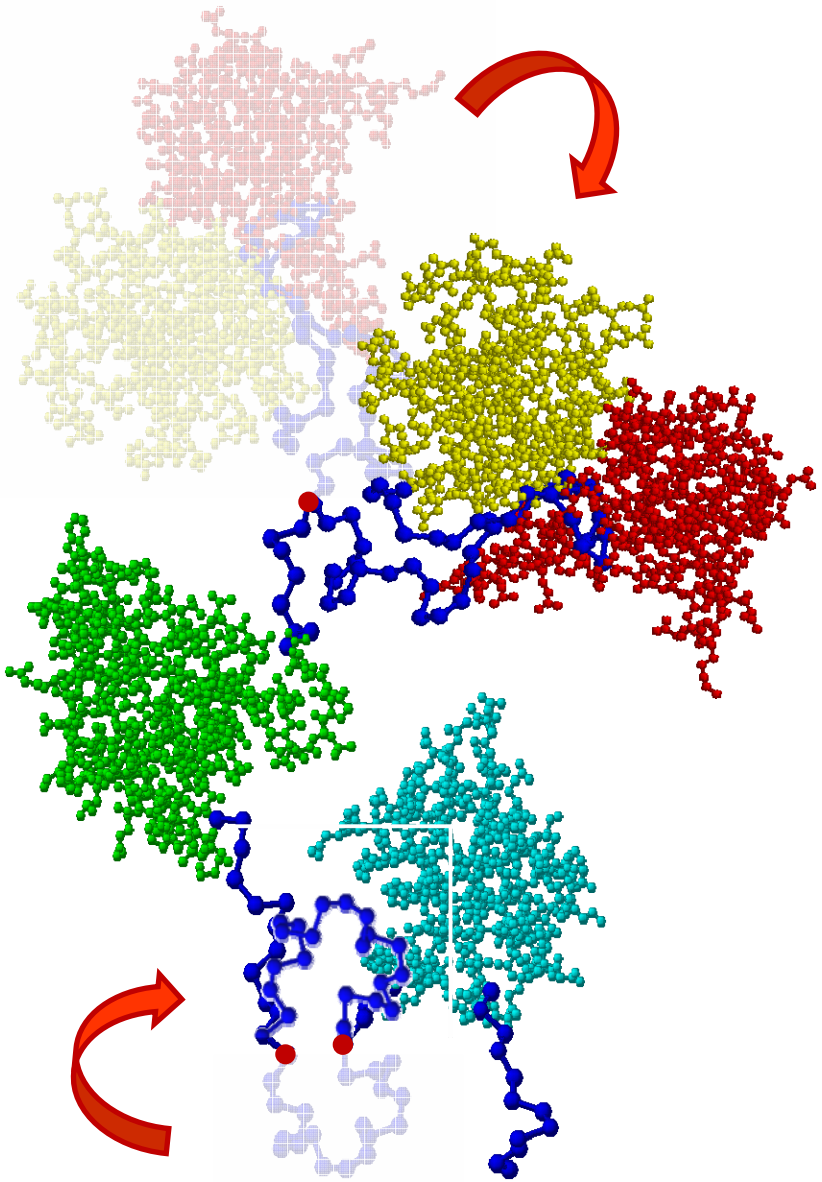
D.I. Svergun, M.V. Petoukhov, & M.H.J. Koch (2001) *Biophys. J.* 80, 2946-53

Modelling of multidomain proteins

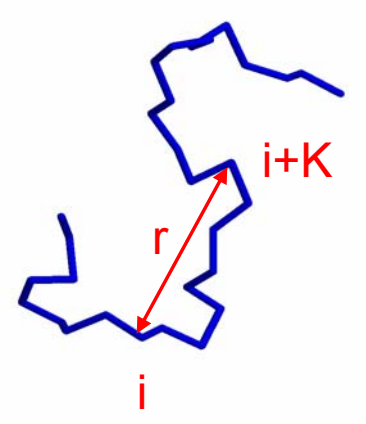


- A combined approach is proposed to built the models of multidomain proteins with large and flexible interdomain linkers
- The latter are represented as DR chains which are attached to the appropriate terminals in rigid domains.
- A single modification of a model is a rotation about one or two randomly selected DR(s).

Modelling of multidomain proteins

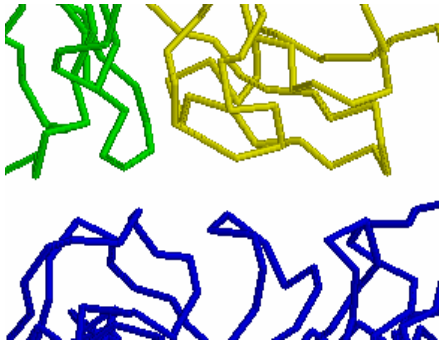


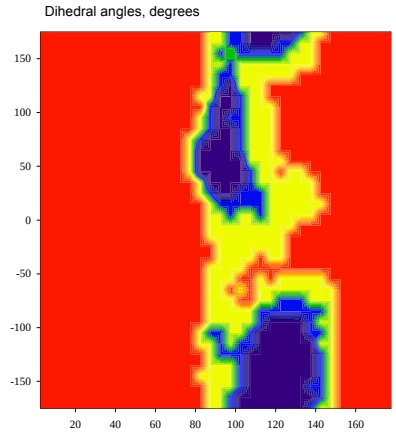
Building native-like folds of linkers



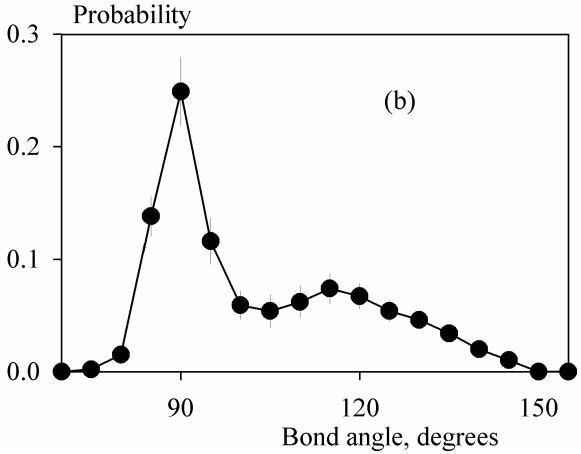
Neighbors distribution along the sequence

Absence of steric clashes



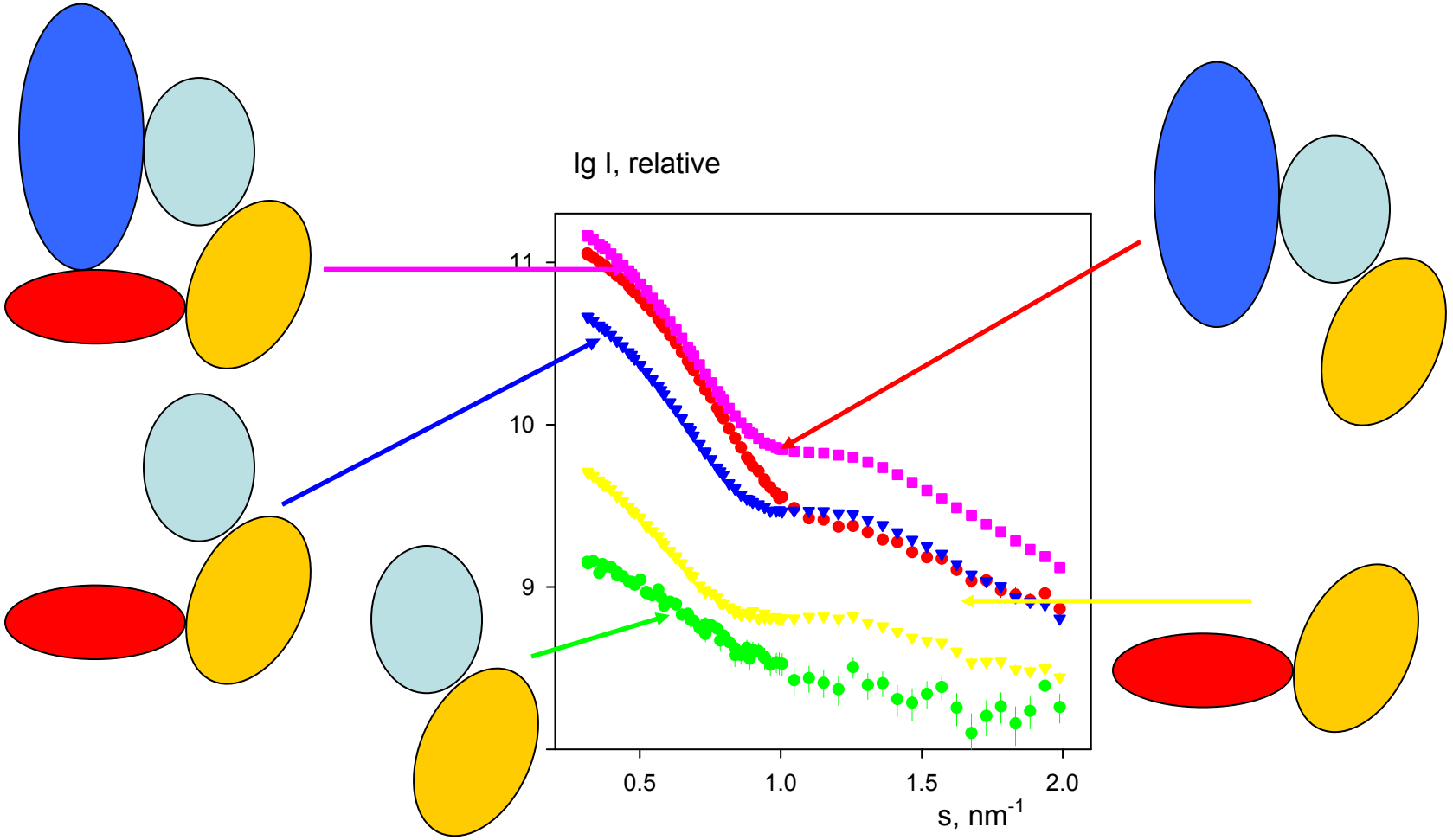


Bond angles & dihedrals distribution



Loop compactness may also be required $Rg_{id} = 3\sqrt[3]{n_l}$

Simultaneous fitting of multiple data sets from deletion mutants



BUNCH: Modelling of multidomain proteins

- Search of the optimal positions and orientations of rigid domains and probable conformations of DR linkers, those fit the SAXS data.
- Proper bond and dihedral angles in the DR chains are required together with the absence of overlaps.
- The scattering pattern is calculated from partial amplitudes of domains and form-factors of DR comprising the loops using spherical harmonics.

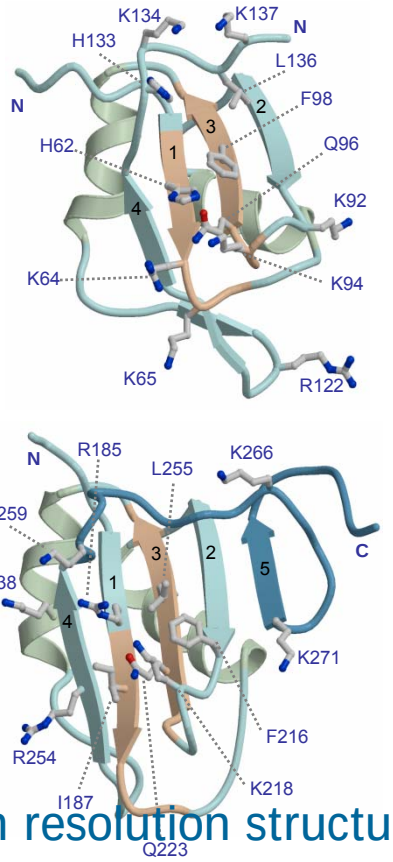
$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left| \sum_{k} A^{(k)}{}_{lm}(s) + \sum_{i} D^{(i)}{}_{lm}(s) \right|^2$$

- Multiple scattering curves fitting from deletion mutants

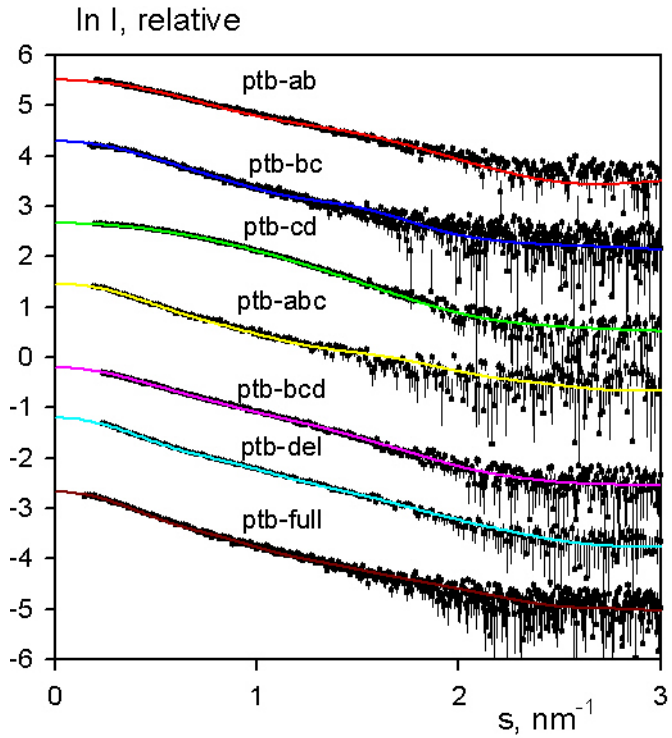
Petoukhov M.V., Svergun, D.I. (2005). *Biophys. J.* **89**, 1237-1250

Structure and RNA interactions of polypyrimidine tract binding protein

PTB is an important regulator of alternative splicing, which allows the production of multiple mRNA transcripts from a single pre-mRNA species. PTB contains four domains (RNA recognition motifs, RRMs), whose structure is solved by NMR.



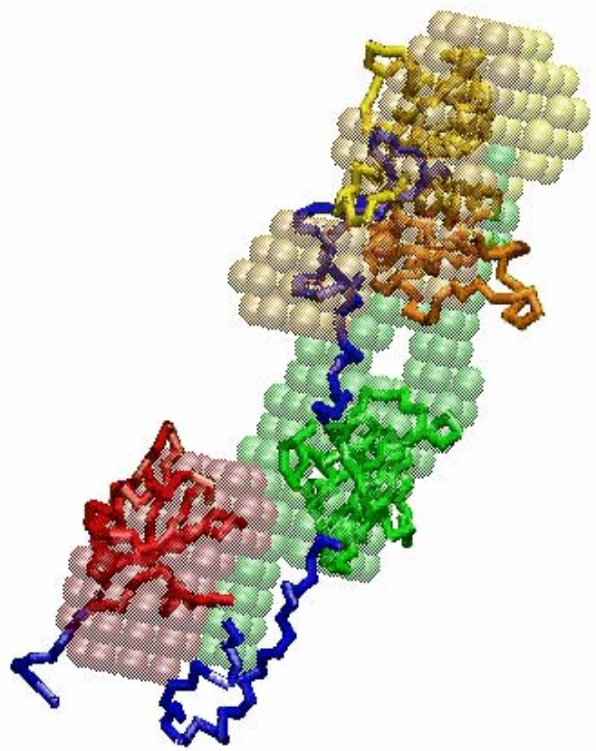
NMR: high resolution structures of RRM1 and RRM2

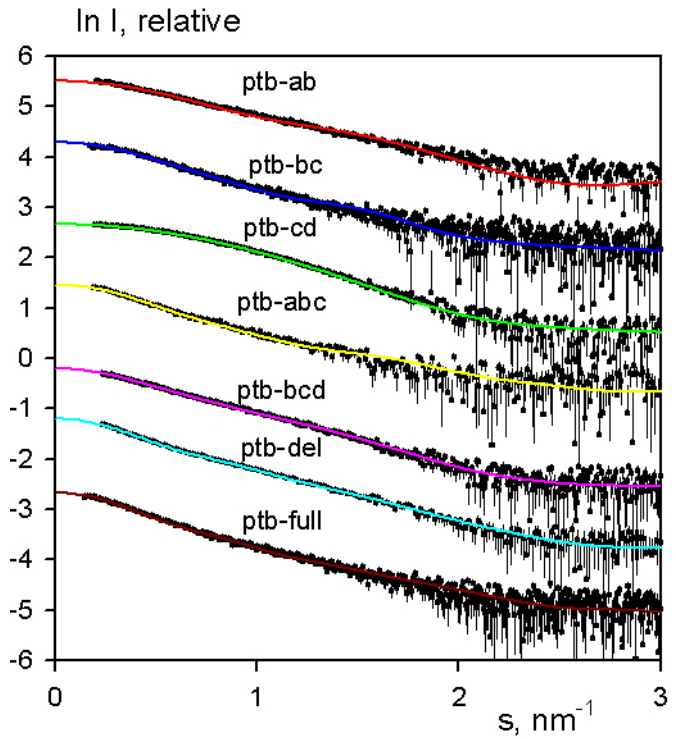


Multiple scattering curves from deletion mutants fitted simultaneously

Collaboration: S.Curry (London)

Structure and RNA interactions of polypyrimidine tract binding protein





Overlap of the typical *ab initio* and rigid body models

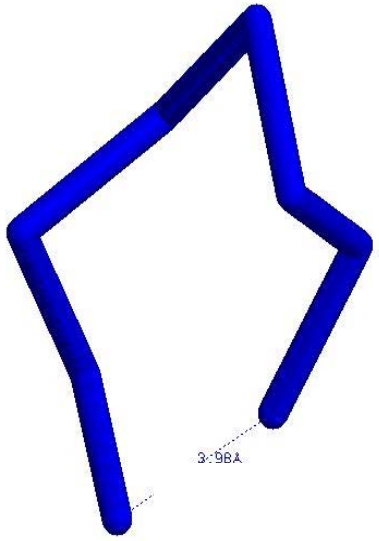
Further restraints (e.g. from NMR) are required to resolve the orientational ambiguity

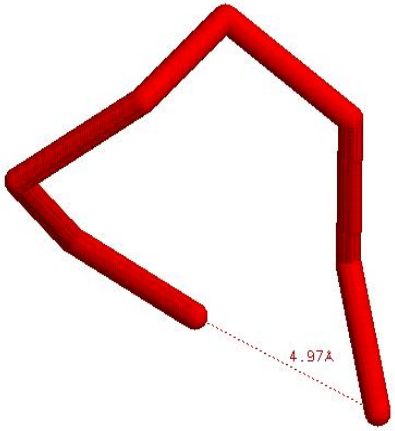
Petoukhov, M. V., Monie, T. P., Allain, F. H., Matthews, S., Curry, S., and Svergun, D. I. (2006). ***Structure*** *14*, 1021-1027.

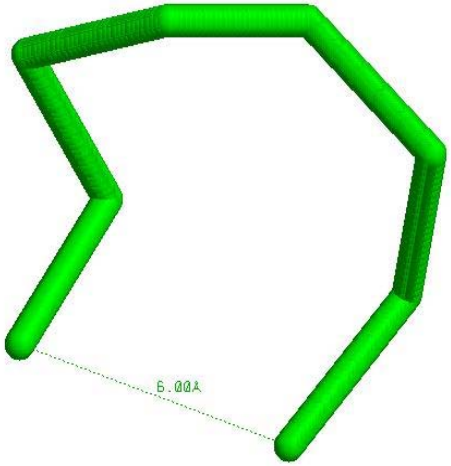
CORAL: Crossing SASREF & BUNCH

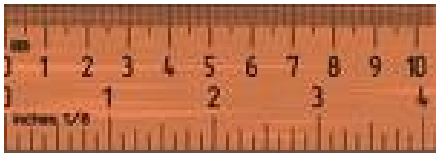
- Sasref:
  - Does not account for missing portions
- Bunch:
  - Single polypeptide chain

Random Loop Library for Combined Modelling





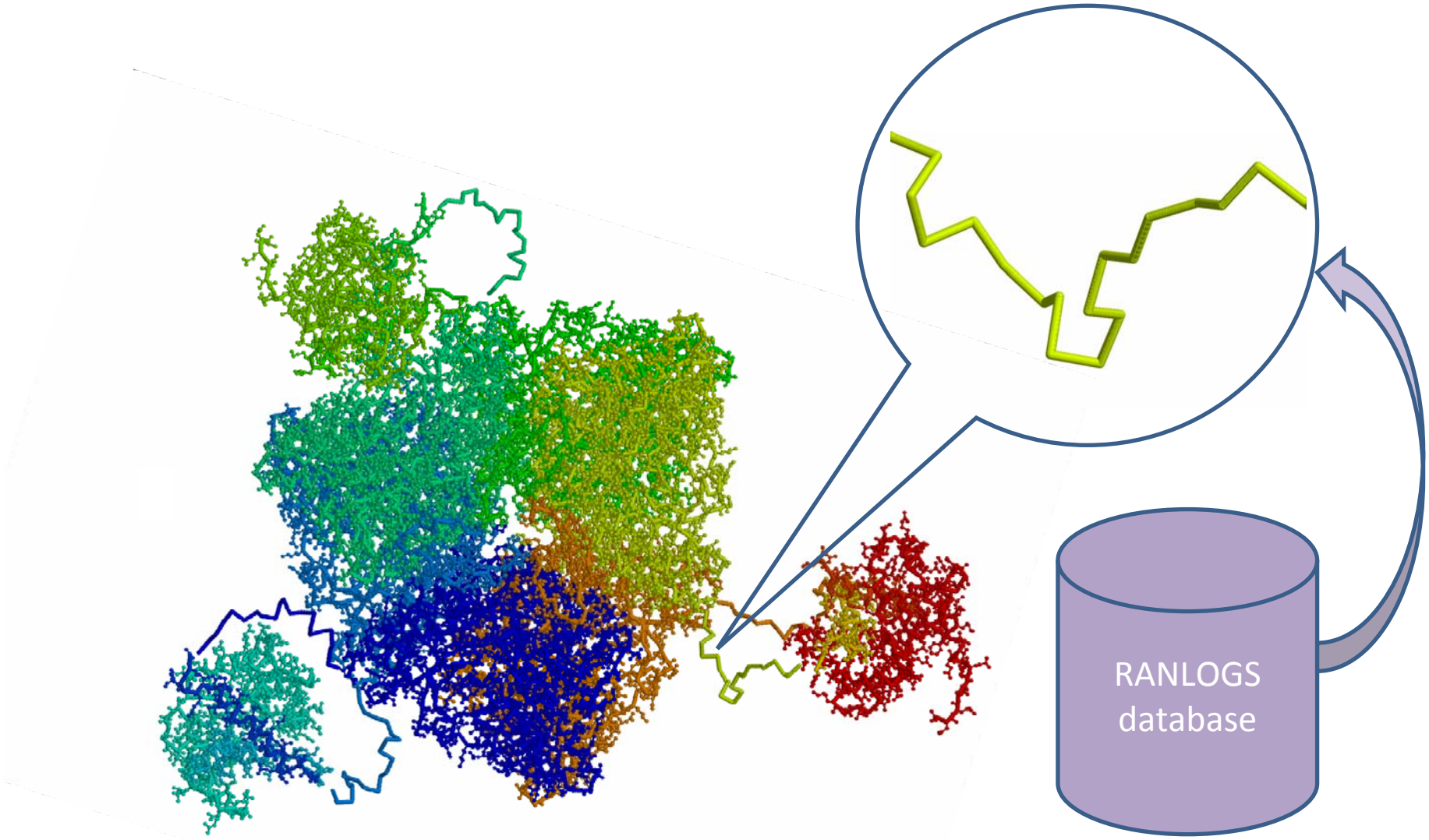




RANLOGS database

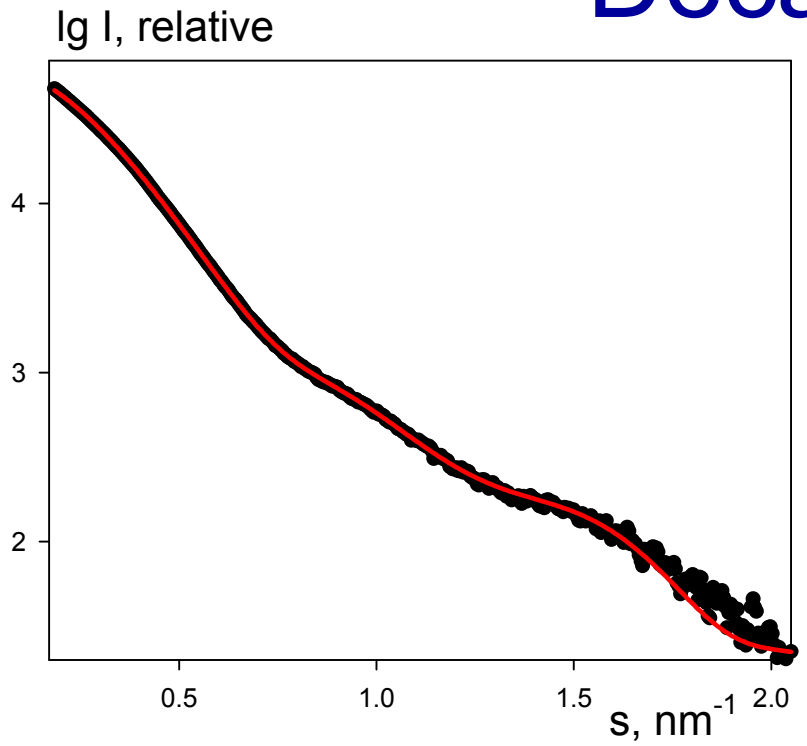
CORAL

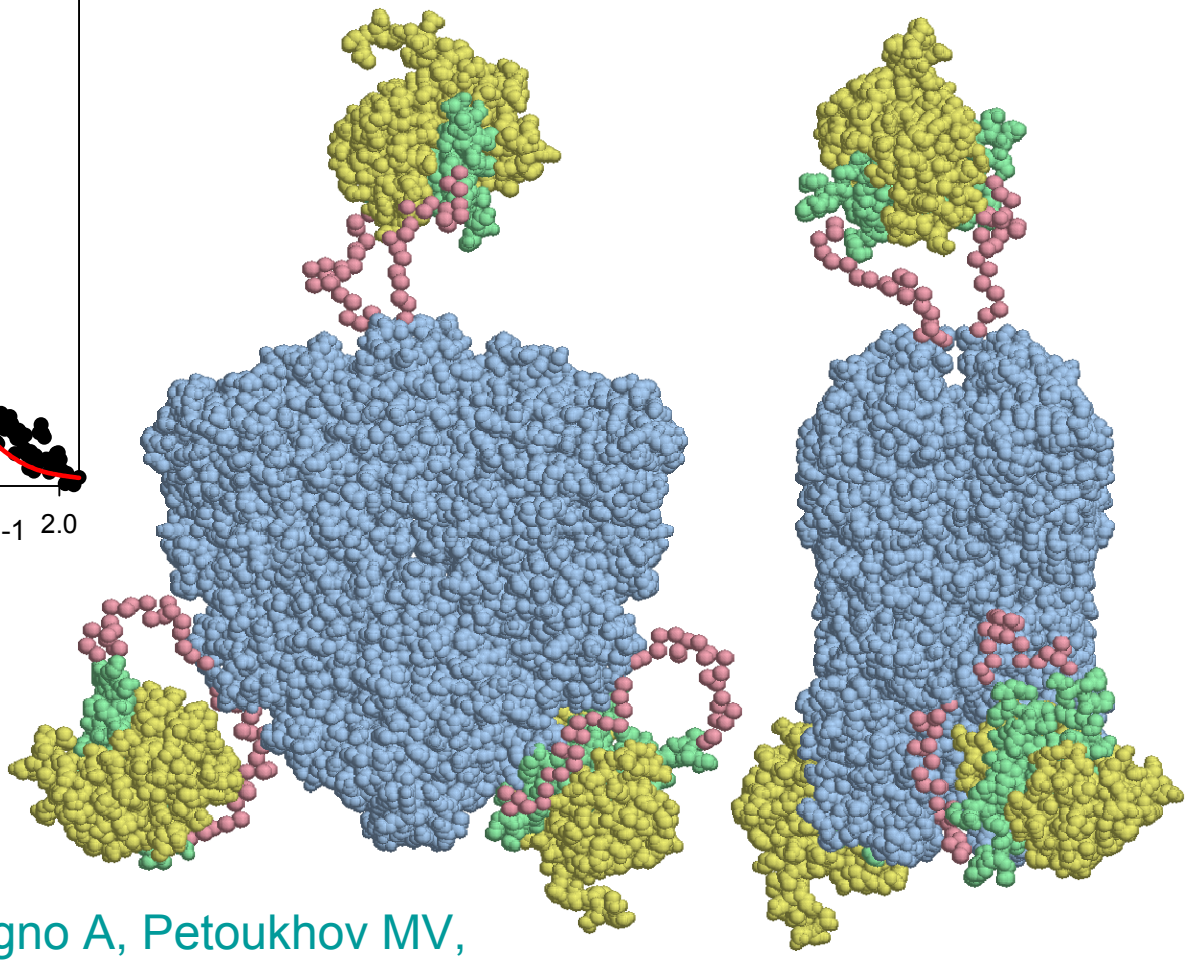
Hybrid Modelling in Coral



Novel feature: consorted movements

Calmodulin-Activated Glutamate Decarboxylase





- 6:3 Gad:CaM stoichiometry
- Planar arrangement of CaM
- In proximity of the three 2-fold axes of Gad

Gut H, Dominici P, Pilati S, Astegno A, Petoukhov MV, Svergun DI, Grütter MG and Capitani G. *J Mol Biol.* 2009 **392**:334-51

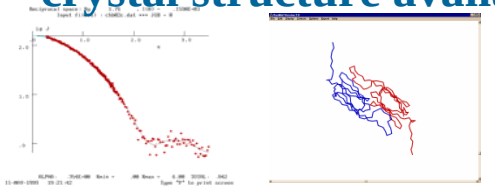
Summary of Hybrid Modelling

- Rigid body modelling allows quaternary structure analysis of macromolecular complexes and multidomain proteins
- The use of complementary data significantly reduces the ambiguity

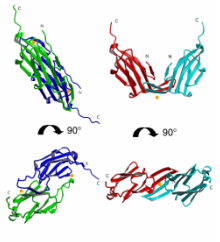
– The resulting models are still low resolution ones

Use of Atomic Models in SAS

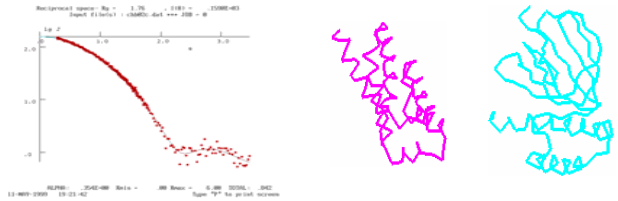
Theoretical model or complete crystal structure available



Validation/identification in solution

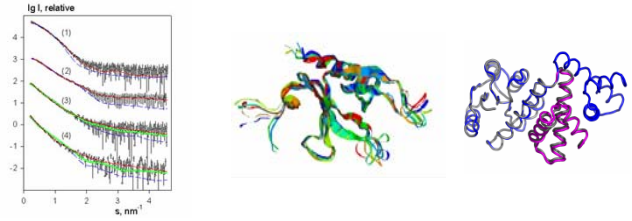


Structure of subunits available



Quaternary structure analysis by rigid body modelling

Structure of domains and multiple curves available



Hybrid modelling of the conformation