

# Rigid body refinement (basics)

D.Svergun, EMBL-Hamburg

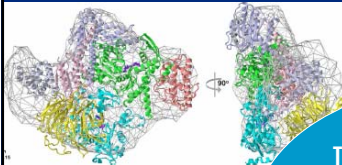


## Shapes from recent projects at EMBL-HH

Complexes and assemblies

Domain and quaternary structure

Filament nucleation complex Arp2/3



Boczkowska *et al*  
Structure (2008)

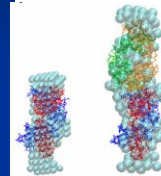
Insulin fibrillation



Glucoamylase



Fab-dye interactions

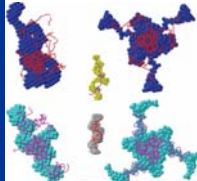


Hillig *et al*  
JMB (2008)

Flexible/transitions

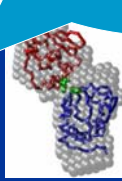
Structural transitions

(NC)-dUTPase



Nemeth-Pongrácz  
*et al* NAR (2007)

Cytochrome



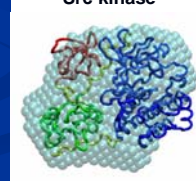
Xu *et al*  
JACS (2008)

Complex



She *et al*, Mol Cell (2008)

Src kinase



Bernado *et al*  
JMB (2008)

In most cases, high resolution models are drawn inside the shapes

## Using SAXS with MX/NMR: 'hybrid' modelling



Model building where high resolution portions are positioned to fit the low resolution SAXS data

## The use of high resolution models in SAS

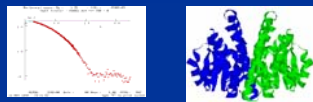
**Theoretical model or complete crystal structure available**



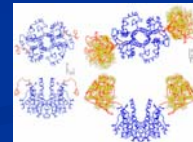
**Validation in solution**



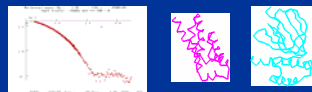
**Incomplete structure available**



**Addition of missing loops/domains**



**Structure of subunits available**



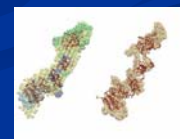
**Rigid body model of the complex**



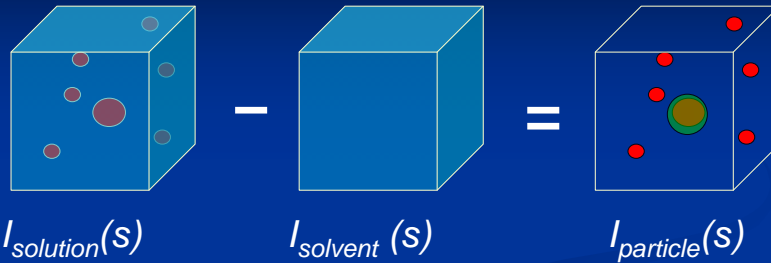
**Structure of domains and multiple curves available**



**Model of the domain structure**



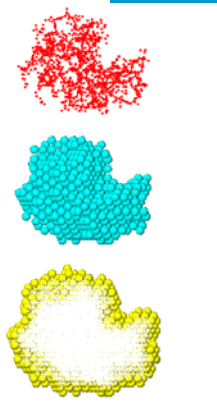
## How to compute SAS from atomic model



- ◆ 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  $\rho_s$  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

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



- ◆  $A_a(s)$ : atomic scattering in vacuum
- ◆  $A_s(s)$ : scattering from the excluded volume
- ◆  $A_b(s)$ : scattering from the hydration shell

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

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

## The use of multipole expansion

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

If the intensity of each contribution is represented using spherical harmonics

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^l |A_{lm}(s)|^2$$

the average is performed analytically:

$$I(s) = 2\pi^2 \sum_{l=0}^L \sum_{m=-l}^l |A_{lm}(s) - \rho_0 E_{lm}(s) + \delta\rho B_{lm}(s)|^2$$

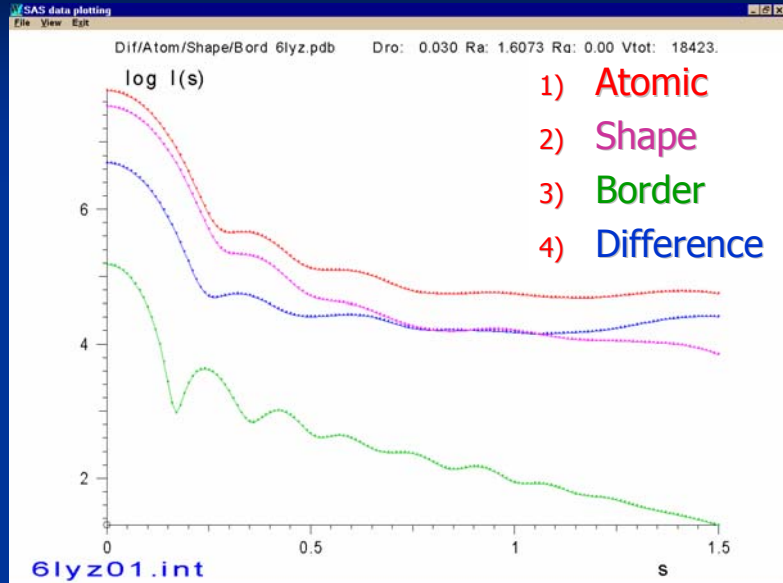
This approach permits to further use rapid algorithms for rigid body refinement

## **CRYSOL** and **CRYSON**: X-ray and neutron scattering from macromolecules

$$I(s) = 2\pi^2 \sum_{l=0}^L \sum_{m=-l}^l |A_{lm}(s) - \rho_0 E_{lm}(s) + \delta\rho B_{lm}(s)|^2$$

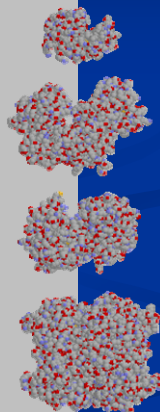
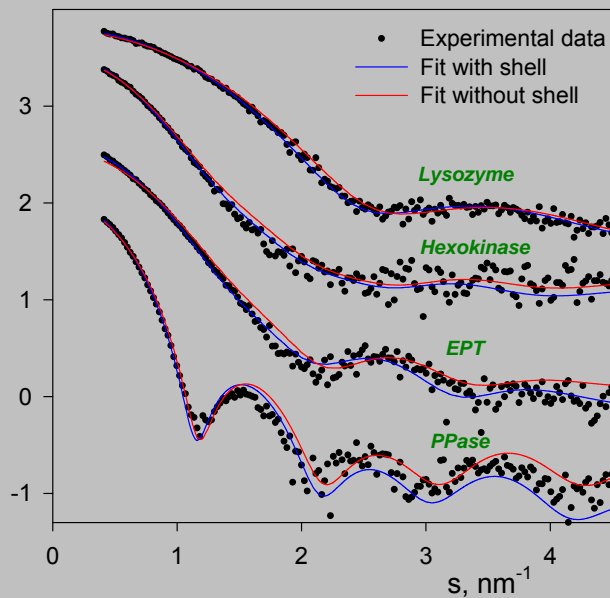
- The programs:
  - either fit the experimental data by varying the density of the hydration layer  $\delta\rho$  (affects the third term) and the total excluded volume (affects the second term)
  - or predict the scattering from the atomic structure using default parameters (theoretical excluded volume and bound solvent density of 1.1 g/cm<sup>3</sup>)
  - provide output files (scattering amplitudes) for rigid body refinement routines
  - compute particle envelope function  $F(\omega)$

## Scattering components (lysozyme)

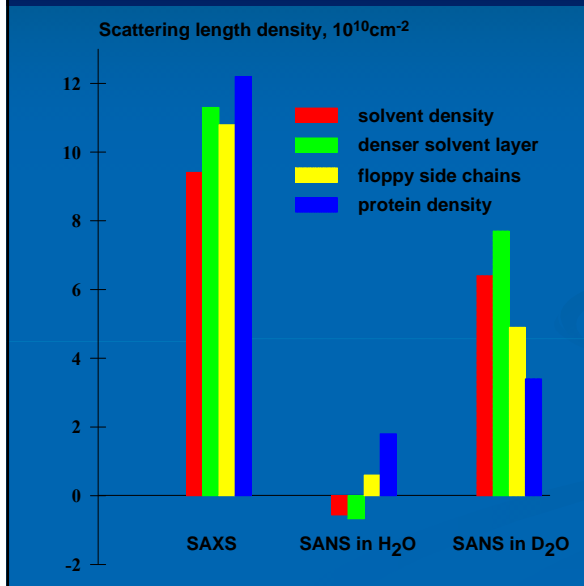


## Effect of the hydration shell, X-rays

lg I, relative

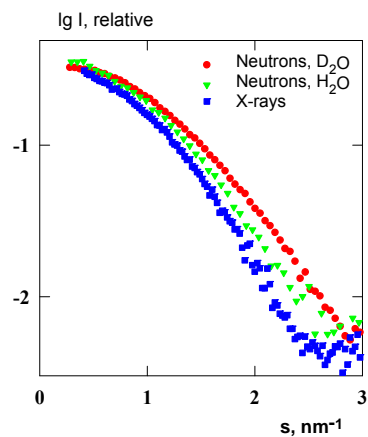


## Denser shell or floppy chains: X-rays *versus* neutrons

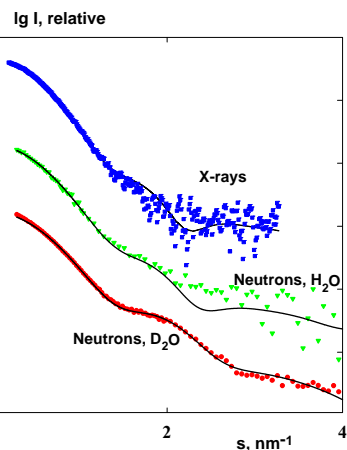


- ◆ For X-rays: both lead to similar effect (particle appears larger)
- ◆ Floppy chains would in all cases increase the apparent particle size
- ◆ Neutrons in H<sub>2</sub>O (shell): particle would appear nearly unchanged
- ◆ Neutrons in D<sub>2</sub>O (shell): particle would appear smaller than the atomic model

## X-rays *versus* neutrons: experiment



Lysozyme: appears larger for X-rays and smaller for neutrons in D<sub>2</sub>O



Thioredoxine reductase : CRYSON and CRYSON fits with denser shell

## CRY SOL and CRY SON: X-ray and neutron scattering from macromolecules

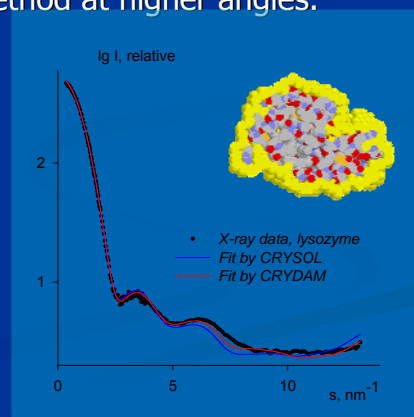
$$I(s) = 2\pi^2 \sum_{l=0}^L \sum_{m=-l}^l |A_{lm}(s) - \rho_0 E_{lm}(s) + \delta\rho B_{lm}(s)|^2$$

- The programs:
  - either fit the experimental data by varying the density of the hydration layer  $\delta\rho$  (affects the third term) and the total excluded volume (affects the second term)
  - or predict the scattering from the atomic structure using default parameters (theoretical excluded volume and bound solvent density of 1.1 g/cm<sup>3</sup>)
  - provide output files (scattering amplitudes) for rigid body refinement routines
  - compute particle envelope function  $F(\omega)$

## Other approaches/programs I

- The 'cube method' (Luzzati et al, 1972; Fedorov and Pavlov, 1983; Müller, 1983) ensures uniform filling of the excluded volume. Could/should/must be superior over the effective atomic factors method at higher angles.
- CRYDAM (still unpublished)
  - ◆ Represents hydration shell by dummy water atoms
  - ◆ Computes X-ray and neutron scattering profiles
  - ◆ Handles proteins, carbohydrates, nucleic acids and their complexes
  - ◆ Is applicable for wide angle scattering range

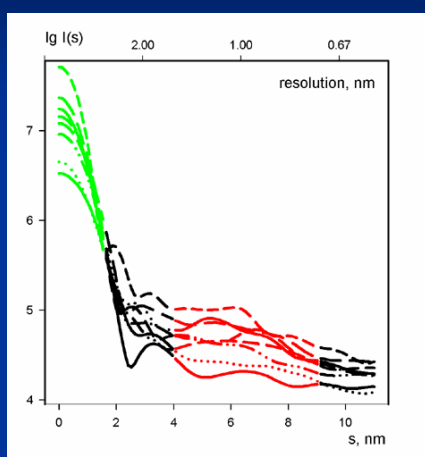
Maffei, M. & Svergun, D.I. (2001).  
to be submitted



## Other approaches/programs II

- J. Bardhan, S. Park and L. Makowski (2009) SoftWAXS: a computational tool for modeling wide-angle X-ray solution scattering from biomolecules *J. Appl. Cryst.* **42**, 932-943  
*A program to compute WAXS*
- Schneidman-Duhovny D, Hammel M, Sali A. (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38** Suppl:W540-4.  
*Debye-like computations, Web server*
- Grishaev A, Guo L, Irving T, Bax A. (2010) Improved Fitting of Solution X-ray Scattering Data to Macromolecular Structures and Structural Ensembles by Explicit Water Modeling. *J Am Chem Soc.* Oct 19. [Epub ahead of print]  
*They really do fit the data to the models...*

## DARA, a database for rapid characterization of proteins



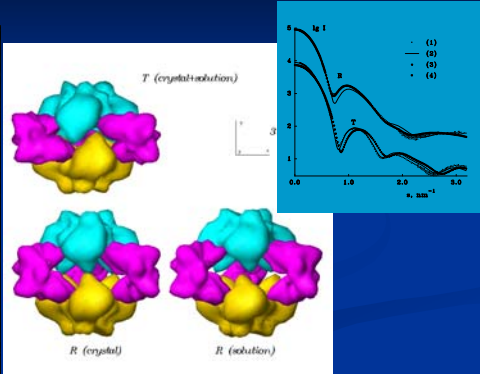
<http://dara.embl-hamburg.de/>

About 15000 atomic models of biologically active molecules are generated from the entries in Protein Data Bank and the scattering patterns computed by CRY SOL

Rapidly identifies proteins with similar shape (from low resolution data) and neighbors in structural organization (from higher resolution data)

Sokolova, A.V., Volkov, V.V. & Svergun, D.I. (2003) *J. Appl. Crystallogr.* **36**, 865-868

# Validation of high resolution models

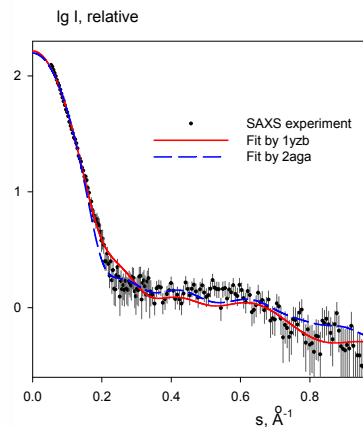
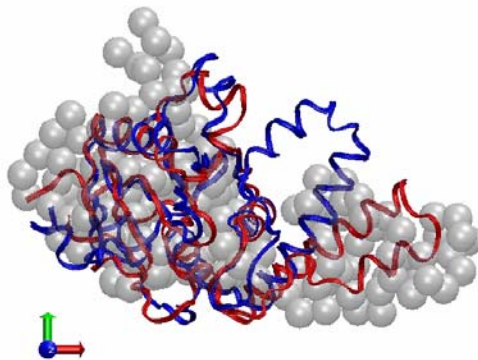


Crystallographic packing forces are comparable with the intersubunit interactions. The solution structures of multisubunit macromolecules could be significantly different from those in the crystal

Packing forces in the crystal restrict the allosteric transition in aspartate transcarbamylase

Svergun, D.I., Barberato, C., Koch, M.H.J., Fetler, L. & Vachette, P. (1997). *Proteins*, **27**, 110-117

# Validation of high resolution models

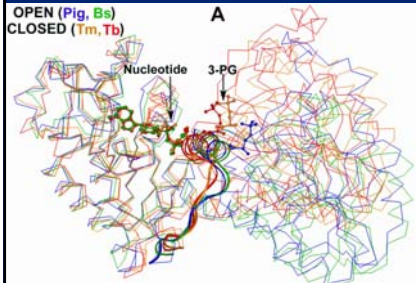


NMR models of the Josephin domain of ataxin-3: red curve and chain: **1yzb**, Nicastro et al. (2005) *PNAS USA* **102**, 10493; blue curve and chain: **2aga**, Mao et al. (2005) *PNAS USA* **102**, 12700.

Nicastro, G., Habeck, M., Masino, L., Svergun, D.I. & Pastore, A. (2006) *J. Biomol. NMR*, **36**, 267.

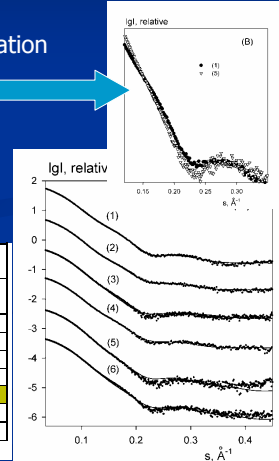
## Domain Closure in 3-Phosphoglycerate Kinase

Closure of the two domains of PGK upon substrate binding is essential for the enzyme function. Numerous crystal structures do not yield conclusive answer, which conditions are required for the closure



A SAXS fingerprint of open/closed conformation (human PGK)

SAXS proves that binding of both substrates induces the closure



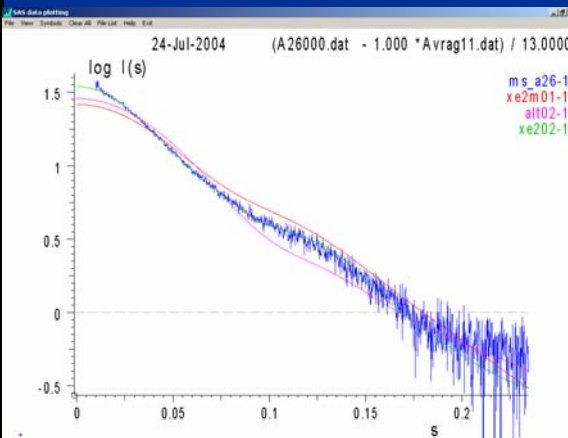
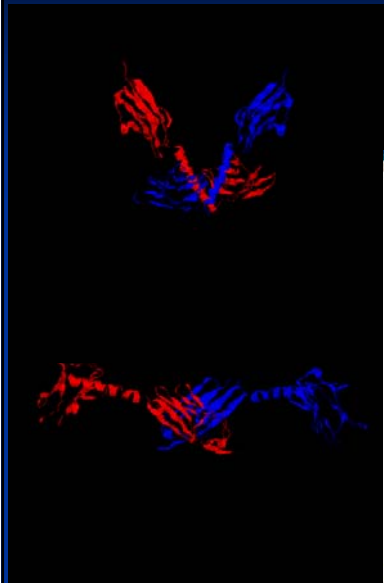
Ligands/ Parameters	Pig PGK	Bs PGK	Pig PGK				Tm PGK	Tb PGK
	Substr. free	MgADP binary	MgATP binary	3-PG binary	<sup>a</sup> tern1	<sup>a</sup> tern2	<sup>a</sup> tern1	<sup>a</sup> tern2
No	2.746	4.332	3.524	3.158	3.664	4.767	9.135	9.560
3-PG	2.678	5.329	3.297	1.958	3.655	4.234	6.052	6.125
MgATP	3.855	2.848	2.409	3.389	7.827	7.766	3.179	3.910
MgADP	1.486	3.235	1.627	1.140	1.780	2.463	5.151	6.193
MgATP*3-PG	6.140	6.044	4.656	5.307	5.146	4.805	2.247	1.611
MgADP*3-PG	2.303	3.522	2.795	2.049	2.712	2.810	2.018	2.922
R <sub>s</sub> (theor), A	24.25	24.34	24.02	23.97	24.24	24.16	23.26	22.64

Varga, A., Flachner, B., Konarev, P., Grácz, E., Szabó, J., Svergun, D., Závodszky, P. & Vas, M. (2006) *FEBS Lett.* **580**, 2698-2706.

## Identification of biologically active oligomers

### Biologically active dimer of myomesin-1

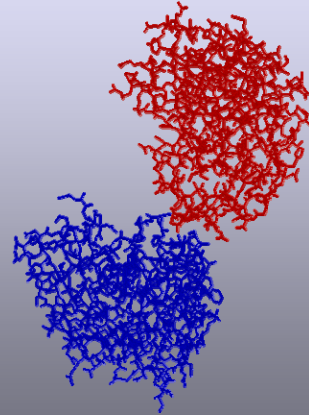
Experiment started: 24-07-2004 at 21:09  
Final result obtained: 24-07-2004 at 21:48



Pinotsis, N., Lange, S., Perriard, J.-C., Svergun, D.I. & Wilmanns, M. (2008) *EMBO J.* **27**, 253-264

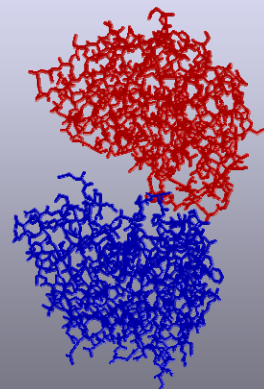
## The idea of rigid body modeling

- The structures of two subunits in reference positions are known.
- Arbitrary complex can be constructed by moving and rotating the second subunit.
- This operation depends on three Euler rotation angles and three Cartesian shifts.

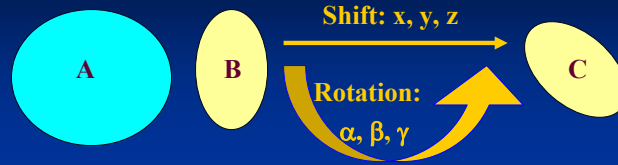


## The idea of rigid body modeling

- The structures of two subunits in reference positions are known.
- Arbitrary complex can be constructed by moving and rotating the second subunit.
- This operation depends on three Euler rotation angles and three Cartesian shifts.



## Equation for rigid body modeling



Using spherical harmonics, the amplitude(s) of arbitrarily rotated and displaced subunit(s) are analytically expressed *via* the initial amplitude and the six positional parameters:  $C_{lm}(s) = C_{lm}(B_{lm}, \alpha, \beta, \gamma, x, y, z)$ .

The scattering from the complex is then rapidly calculated as

$$I(s) = I_A(s) + I_B(s) + 4\pi^2 \sum_0^{\infty} \sum_{-l}^l \text{Re} [A_{lm}(s) C_{lm}^*(s)]$$

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

## Constraints for rigid body modelling



- *Interconnectivity*
- *Absence of steric clashes*
- Symmetry
- Intersubunit contacts (from chemical shifts by NMR or mutagenesis)
- Distances between residues (FRET or mutagenesis)
- Relative orientation of subunits (RDC by NMR)
- Scattering data from subcomplexes

Petoukhov & Svergun  
(2005) *Biophys J.* **89**, 1237;  
(2006) *Eur. Biophys. J.* **35**,  
567.

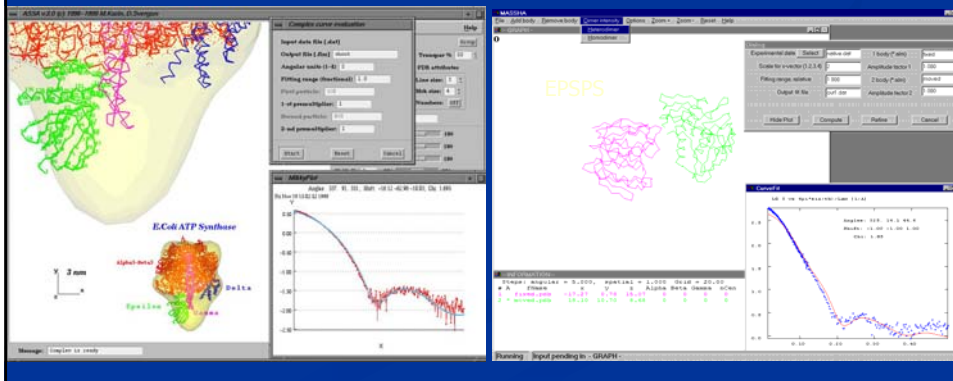
# Interactive and local refinement

◆ **ASSA** (SUN/SGI/DEC)

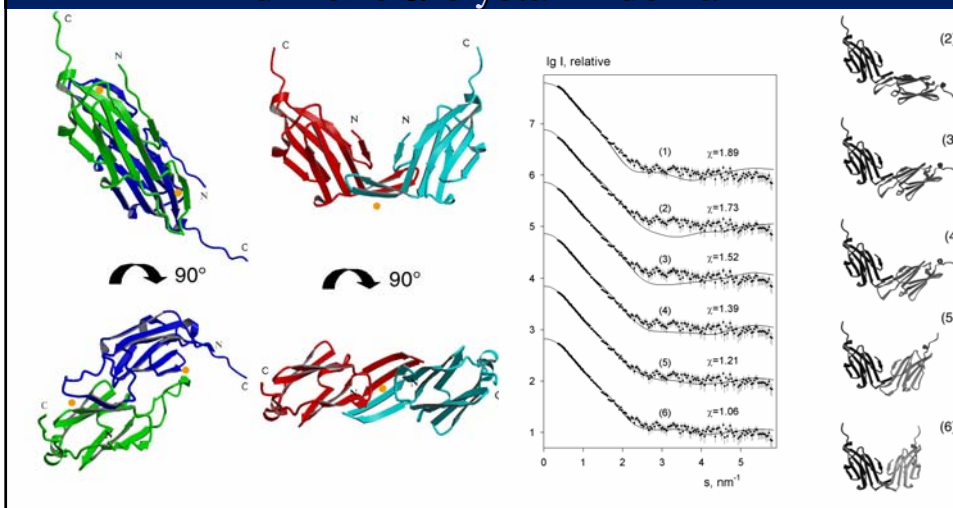
Kozin & Svergun (2000). *J. Appl. Cryst.* **33**, 775-777

◆ **MASSHA** (Win9x/NT/2000)

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



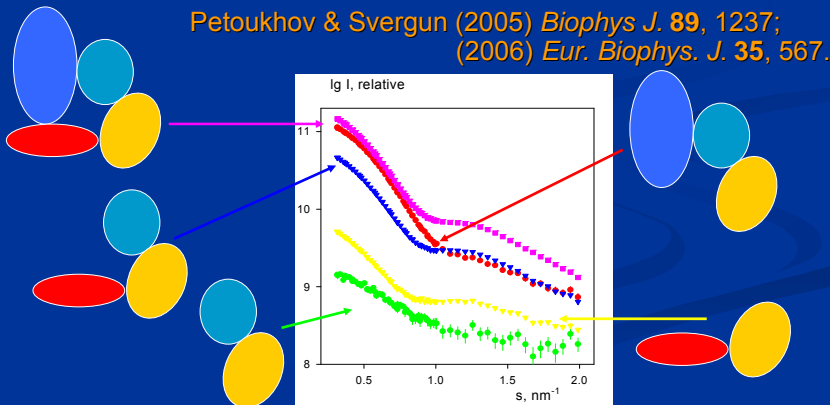
# Manual refinement: quaternary structure of the dimeric $\alpha$ -crystallin domain



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

## Global rigid body modelling (SASREF)

- Fits (multiple X-ray and neutron) scattering curve(s) from partial constructs or contrast variation using simulated annealing
- Requires models of subunits, builds interconnected models without steric clashes
- Uses constraints: symmetry, distance (FRET or mutagenesis) relative orientation (RDC from NMR), if applicable



## Local and global search



- Local search always goes to a better point and can thus be trapped in a local minimum
- To get out of local minima, global search must be able to (sometimes) go to a worse point
- Pure Monte-Carlo search always goes to the closest local minimum (nature: rapid quenching and vitreous ice formation)
- Slower annealing allows to search for a global minimum (nature: normal, e.g. slow freezing of water and ice formation)

# 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:  $M \approx 10^3$  variables (e.g. 0 or 1 bead assignments in DAMMIN)**

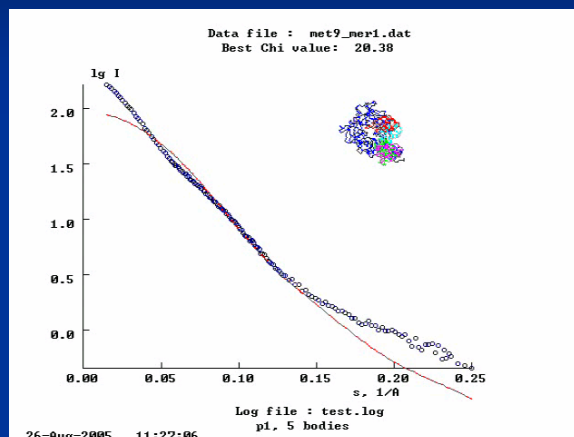
**Rigid body methods:  $M \approx 10^1$  variables (positional and rotational parameters of the subunits)**

**$f(x)$  is always (Discrepancy + Penalty)**

## A global refinement run with distance constraints

A tyrosine kinase MET (118 kDa) consisting of five domains

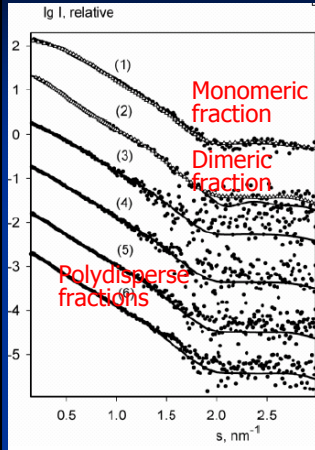
Program  
SASREF



Single curve  
fitting with  
distance  
constraints:  
C to N  
termini  
contacts

Gherardi, E., Sandin, S., Petoukhov, M.V., Finch, J., Youles, M.E., Ofverstedt, L.G., Miguel, R.N., Blundell, T.L., Vande Woude, G.F., Skoglund, U. & Svergun, D.I. (2006) *PNAS USA*, **103**, 4046.

# Quaternary structure of tetanus toxin



Receptor binding H(C) domain reveals concentration-dependent oligomerization

100 : 0

0 : 100

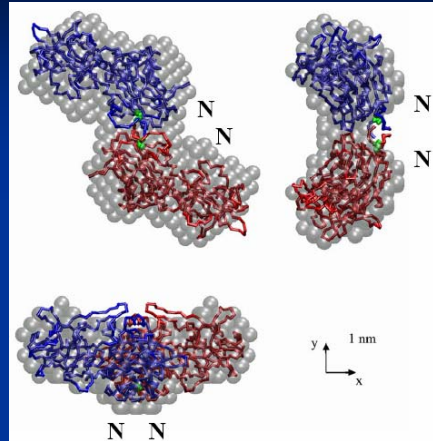
64 : 36

43 : 57

21 : 79

14 : 86

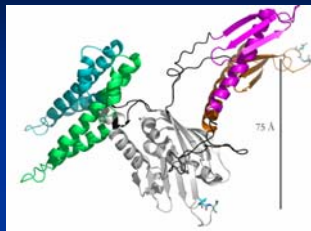
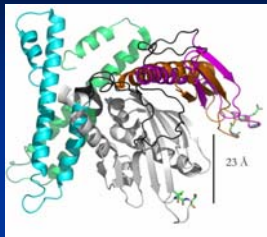
Mon:Dim



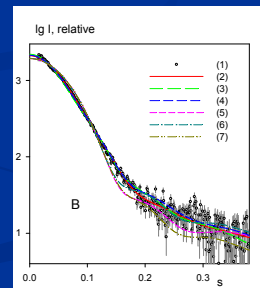
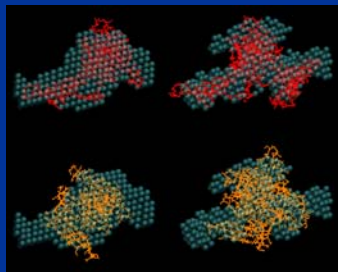
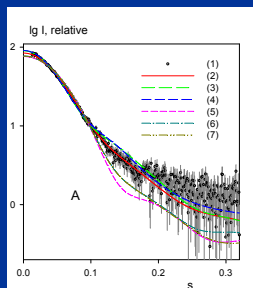
*Ab initio* and rigid body analysis of the dimeric H(C) domain using the structure of the monomer in the crystal (1FV2) and accounting that the mutant Cys869Ala remains always monomeric yield a unique model of the dimer

Qazi, O., Bolgiano, B., Crane, D., Svergun, D.I., Konarev, P.V., Yao, Z.P., Robinson, C.V., Brown, K.A. & Fairweather N. (2007) *J Mol Biol.* **365**, 123–134.

# Solution structure of eucaryotic release factor RF1

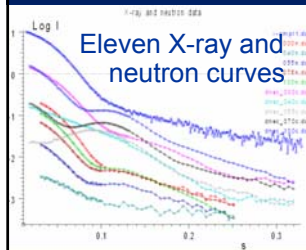


- Cryo-EM: extended; spans the distance between the ribosomal decoding and peptidyl transferase centers
- Crystal: compact, does not span this distance



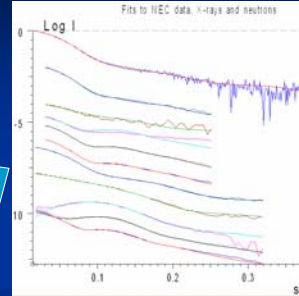
Vestergaard, B., Sanyal, S., Roessle, M., Mora, L., Buckingham, R. H., Kastrop, J. S., Gajhede, M., Svergun, D. I. & Ehrenberg, M. (2005) *Mol. Cell*, **20**, 929–938.

## Rigid body modelling of the Xpot ternary complex

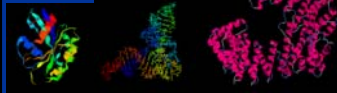


complex

SASREF

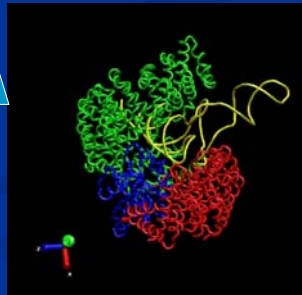


Atomic and homology models



Distance restraints from tRNA footprinting (Arts et al. (1998) *EMBO J.* 17, 7430)

Fukuhara et al. (in preparation)



## Addition of missing fragments



- Flexible loops or domains are often not resolved in high resolution models or genetically removed to facilitate crystallization
- Tentative configuration of such fragments are reconstructed by fixing the known portion and adding the missing parts to fit the scattering from the full-length macromolecule.

## Building native-like folds of missing fragments

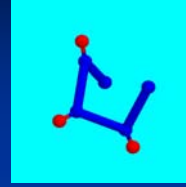
- Using DR-type models and protein-specific penalty functions



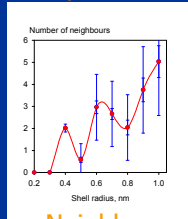
Primary  
sequence



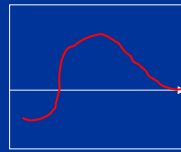
Secondary  
structure



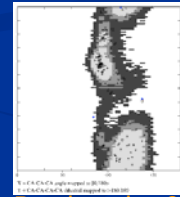
Excluded  
volume



Neighbors  
distribution



Knowledge-based  
potentials

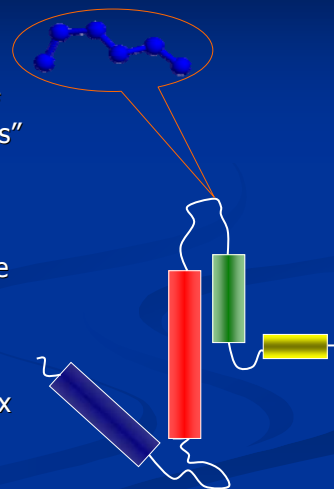


Bond angles &  
dihedrals distribution

Petoukhov, M.V., Eady, N.A.J., Brown, K.A. & Svergun, D.I. (2002) *Biophys. J.* **83**, 3113

## Addition of missing fragments: BUNCH

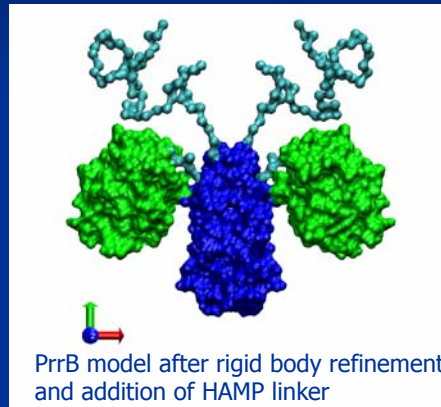
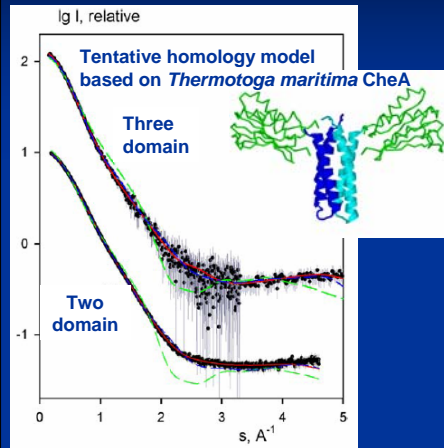
- BUNCH combines rigid body and *ab initio* modelling to find the positions and orientations of rigid domains and probable conformations of flexible linkers represented as "dummy residues" chains
- Multiple experimental scattering data sets from partial constructs (e.g. deletion mutants) can be fitted simultaneously with the data of the full-length protein.
- BUNCH accounts for symmetry, allows one to fix some domains and to restrain the model by contacts between specific residues



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

## Structure of sensor histidine-kinase PrrB

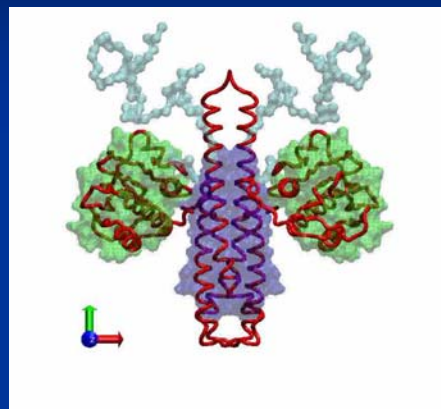
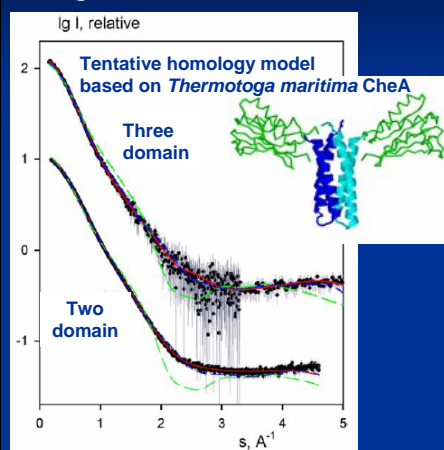
The dimeric sensor histidine-kinase PrrB from *Mycobacterium tuberculosis* contains ATP binding and dimerization domains and a 59 aas long (flexible) HAMP linker



Nowak, E., Panjikar, S., Morth, J. P., Jordanova R., Svergun, D. I. & Tucker, P. A. (2006) *Structure*, 14, 275

## Structure of sensor histidine-kinase PrrB

The dimeric sensor histidine-kinase PrrB from *Mycobacterium tuberculosis* contains ATP binding and dimerization domains and a 59 aas long (flexible) HAMP linker

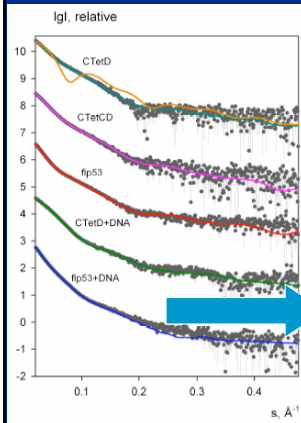


Superposition with the independently determined sensor histidine-kinase from *Thermotoga maritima* (Marina A. et al. (2005) *Embo J.* 24, 4247)

Nowak, E., Panjikar, S., Morth, J. P., Jordanova R., Svergun, D. I. & Tucker, P. A. (2006) *Structure*, 14, 275

## Tumour suppressor p53 and its complex with DNA

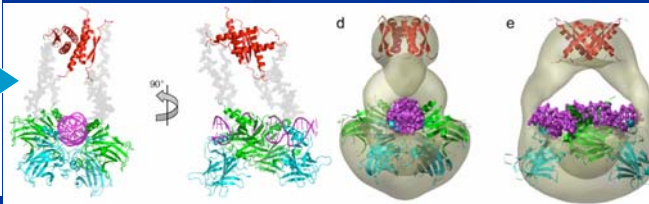
The homotetrameric p53 plays a central role in the cell cycle and maintaining genomic integrity. It consists of folded core and tetramerization domains, linked and flanked by intrinsically disordered segments.



Cross-shaped extended p53 from SAXS and NMR

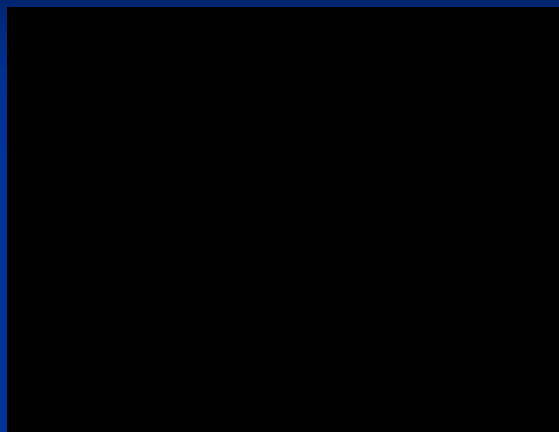


Compact p53/DNA from SAXS and an independent EM reconstruction



Tidow, H., Melero, R., Mylonas, E., Freund, S.M., Grossmann, J.G., Carazo, J.M., Svergun, D.I., Valle, M. & Fersht, A.R. (2007) *Proc Natl Acad Sci USA*, **104**, 12324

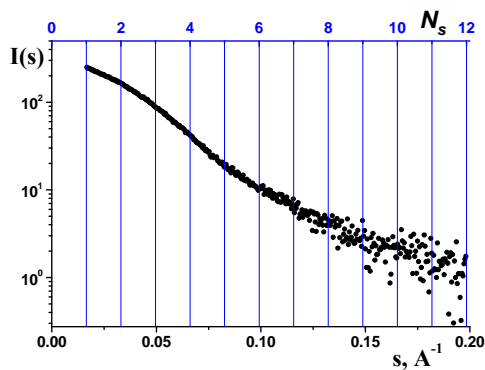
## Some words of caution



Or Always remember about ambiguity!

## Information content in SAS: simple explanation

$$I(s) = \sum_{k=1}^{\infty} s_k I(s_k) \left[ \frac{\sin D(s - s_k)}{D(s - s_k)} - \frac{\sin D(s + s_k)}{D(s + s_k)} \right]$$



A solution scattering curve from a particle with maximum size  $D$  can be represented by its values taken at discrete points (Shannon channels)

$$s_k = k\pi/D$$

In a typical SAS experiment,  
 $N_s \approx 5-15$

C. E. Shannon & W. Weaver (1949).  
The mathematical theory of  
communication. University of Illinois  
Press, Urbana.

## Simple explanations do not work in SAS

Shape determination:  $M \approx 10^3$  variables (e.g. 0 or 1 bead assignments in DAMMIN)

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

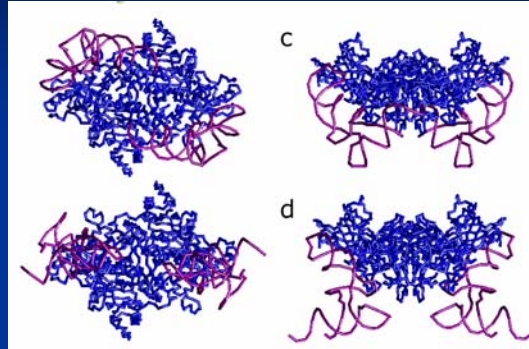
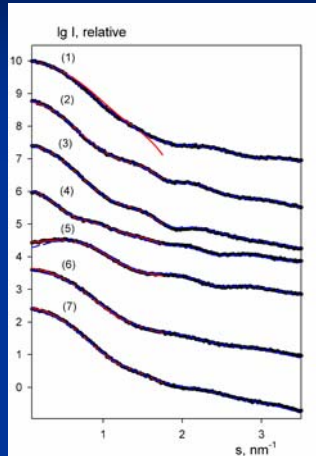
From the informational point of view, rigid body modeling should provide unique or at least much less ambiguous models than shape determination

**NO WAY**

As all the problems are non-linear, the number of Shannon channels does not give you exact number of parameters, which is possible to extract from the scattering data (depending on accuracy, this number varies between zero and infinity).

Further, uniqueness of reconstruction depends largely on the complexity of the function  $f(x)$  to be minimized

## Ambiguity of rigid body analysis



■ A synthetic example: two different orientations of tRNA in a dimeric complex with aspartyl-tRNA synthetase obtained by rigid body modelling and compatible with X-ray and contrast variation neutron scattering data

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

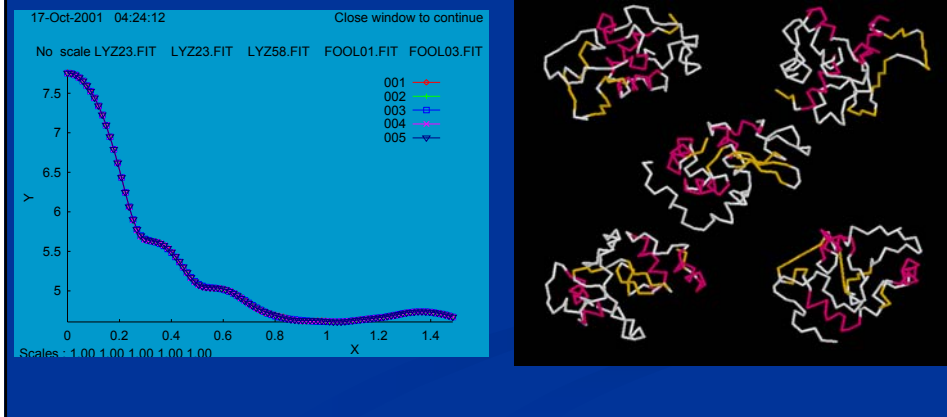
## Constraints and restrains used in global modelling procedures

- Information about contacting residues from other experiments (spin labelling, site-directed mutagenesis, FRET, chemical shifts etc)
- Information about symmetry
- Avoiding steric clashes
- For missing loops and linkers: contiguous chain, excluded volume, Ramachandran plot for Ca, knowledge-based potentials etc

AND STILL, one must always cross-validate SAS models against all available biochemical/biophysical information

# By the way, can X-ray scattering yield the fold?

- Lysozyme and its near-native scattering mates

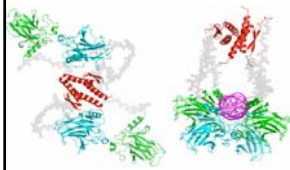


## Recent 'hybrid' projects at EMBL-HH

### Complexes and assemblies

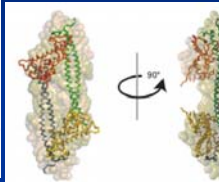
### Domain and quaternary structure

Tumor suppressor p53 and its complex with DNA



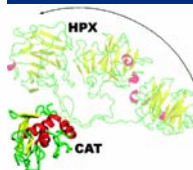
Tidow *et al*  
PNAS USA (2007)

Cdt1-Geminin complex



De Marco *et al*  
PNAS (2009)

Reciprocal domain reorientation in MMP-12



Bertini *et al*, JACS (2008)

Myomesin-1

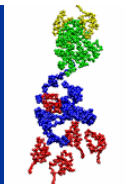


Pinotsis *et al*  
EMBO J (2008)

### Flexible macromolecules

### Structural transitions

Pex5.Pex14.PTS1 complex



Shiozawa *et al*  
JBC (2009)

Ig super-motifs in titin



von Castelmur *et al*  
PNAS USA (2008)

eRF3/eRF1 interactions



Cheng *et al*  
Genes Dev (2009)

Nucleoplasmin/histone



Taneva *et al*  
JMB (2009)



And now, let us awake for the tutorials:



- P.Konarev:  
**MASSHA**  
**(interactive)**
- M.Petoukhov:  
**SASREF, BUNCH**  
**CORAL**  
**(automated)**