Structure analysis of macromolecular solutions with small-angle X-ray scattering

Dmitri Svergun

EMBL

Biological SAXS @ EMBL-HH

Group leader: D. Svergun

Staff : C.Jeffries, C.Blanchet, D.Franke, A.Kikhney, H.Mertens, C.Kerr, C.Borges

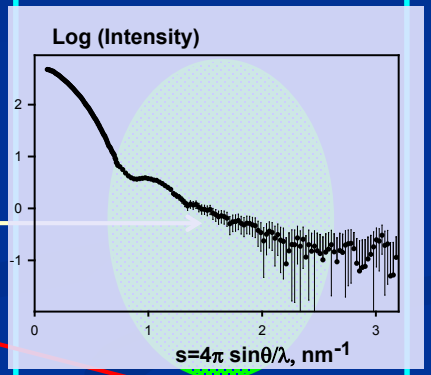
Postdocs: A.Tuukkanen, M.Graewert , A.Spilotros, A.Panjkovich, M.Schroer, A.Gruzinov

Predocs: N.Hajizadeh, K.Malanastas

Major tasks:

- Development of data analysis methods
- Running and developing SAXS beamlines
- User support and collaborative projects
- Interactions, education and training

Small-angle scattering: experiment



Detector

Sample

Monochromatic beam

Wave vector ***k***, $k=2\pi/\lambda$

2θ

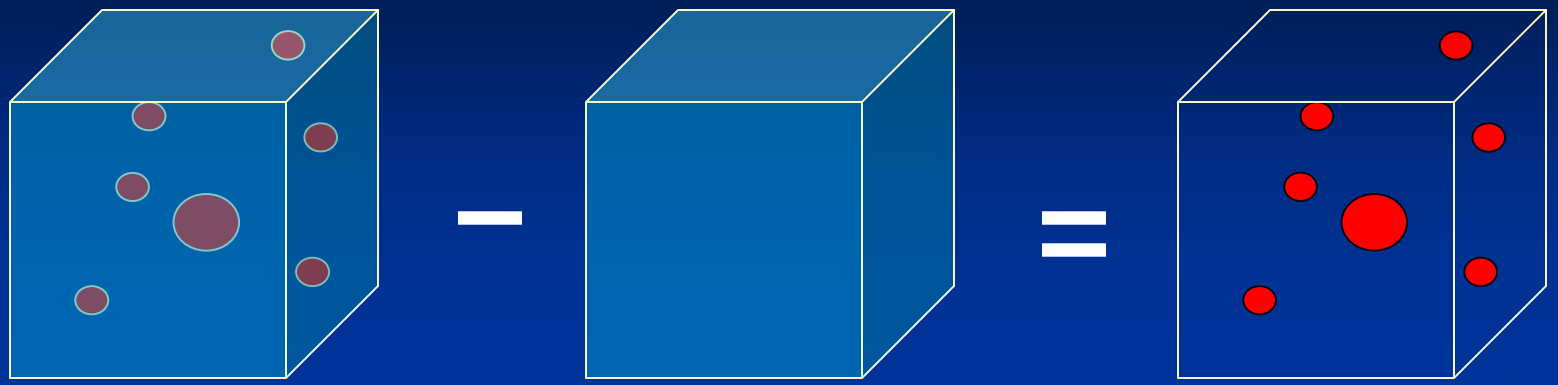
k_1

Radiation sources:

X-ray generator *(λ = 0.1 - 0.2 nm)*
Synchrotron *(λ = 0.03 - 0.35 nm)*
Thermal neutrons *(λ = 0.2 - 1 nm)*

Scattering vector $s=k_1-k$,
$s=4\pi \sin\theta/\lambda$

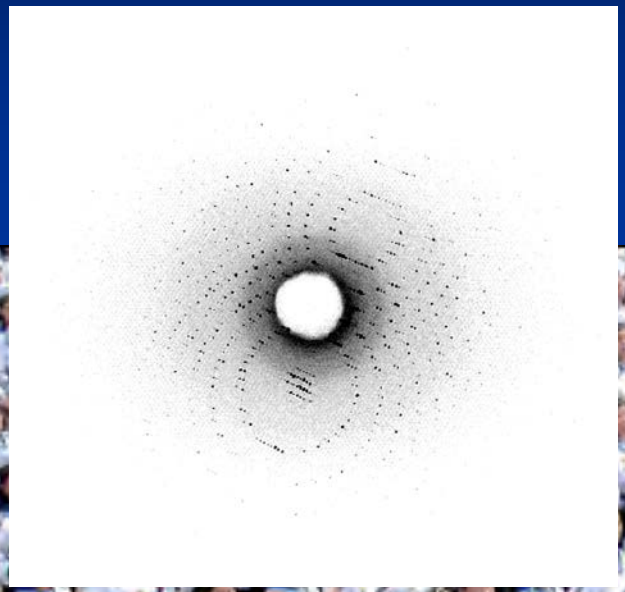
Small-angle scattering: solvent

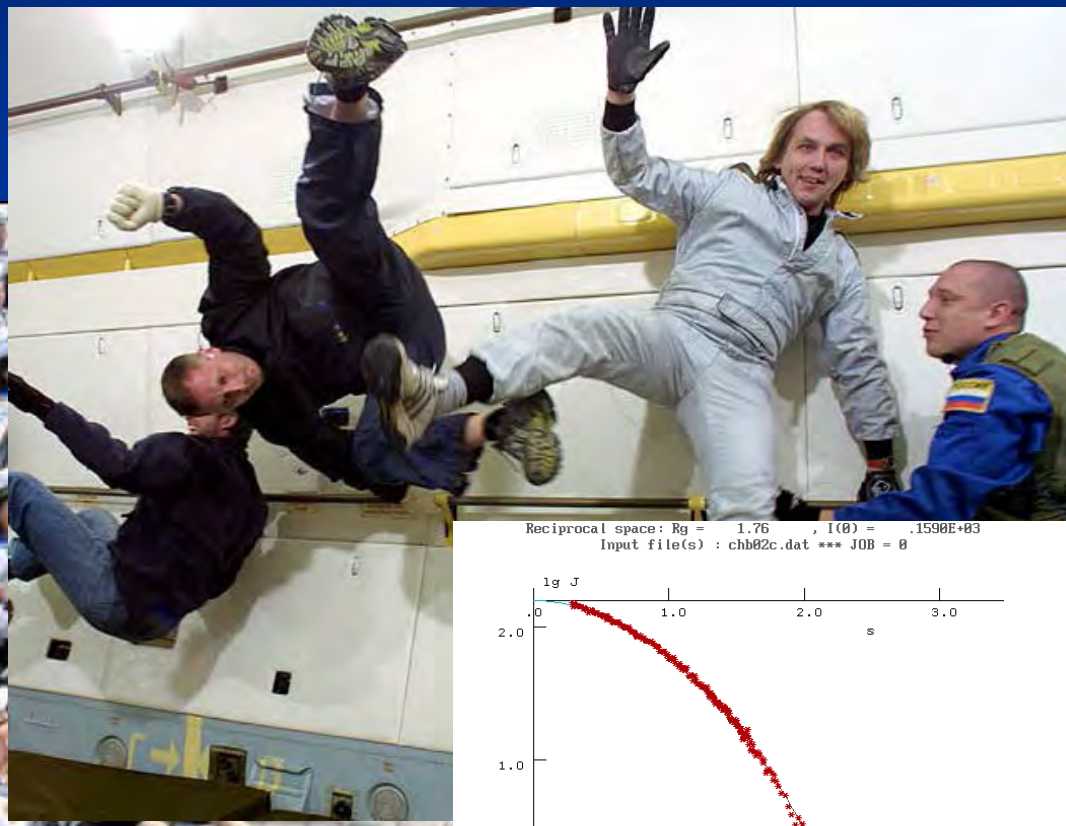


$I_{sample}(s)$ $I_{matrix}(s)$ $I_{particle}(s)$

- To obtain scattering from the particles, matrix scattering must be subtracted, which also permits to significantly reduce contribution from parasitic background (slits, sample holder etc)
- **Contrast $\Delta\rho = \langle\rho(r) - \rho_s\rangle$**, where ρ_s is the scattering density of the matrix, may be very small for biological samples

Crystal versus solution



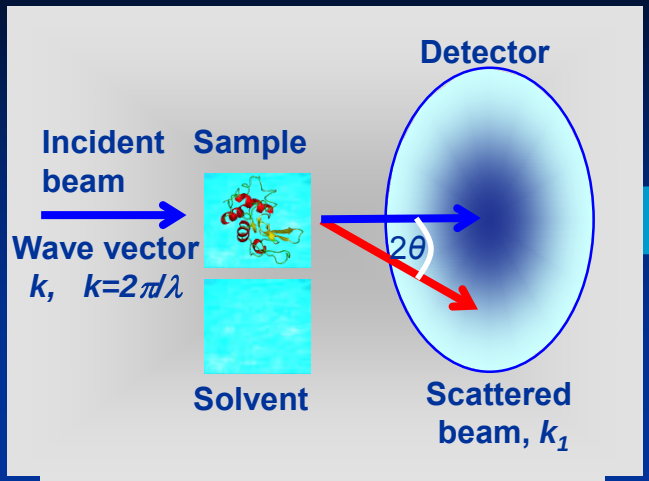


Crystal versus solution

- In solution, no crystallographic packing forces are present

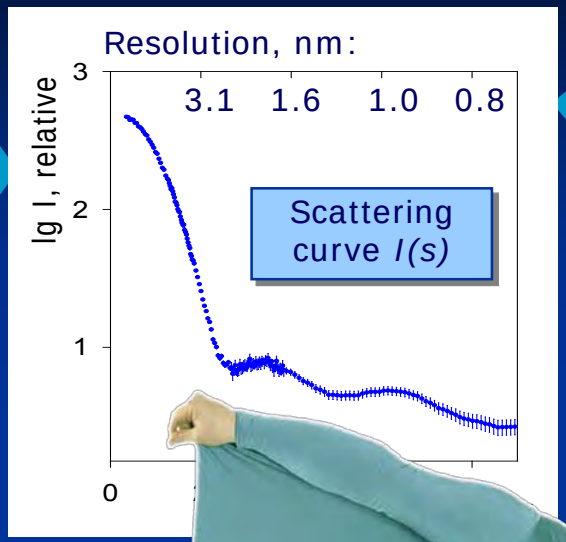
- For SAXS solution studies, one does not need to grow crystals
- SAXS is not limited by molecular mass and is applicable under nearly physiological conditions
- Using solution SAXS, one can more easily observe responses to changes in conditions
- SAXS permits for quantitative analysis of complex systems and processes

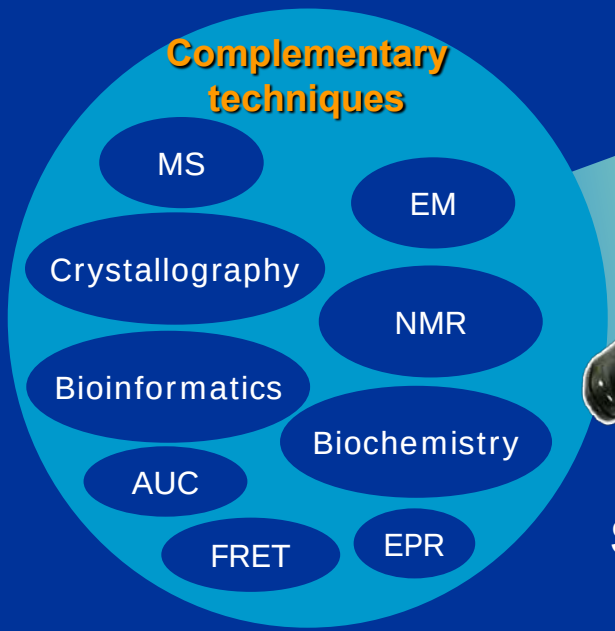
Small-angle scattering in structural biology



Radiation sources:

X-ray tube	(λ = 0.1 - 0.2 nm)
Synchrotron	(λ = 0.05 - 0.5 nm)
Thermal neutrons	(λ = 0.1 - 1 nm)





SAXSMAN © A.Kikhney

Data analysis

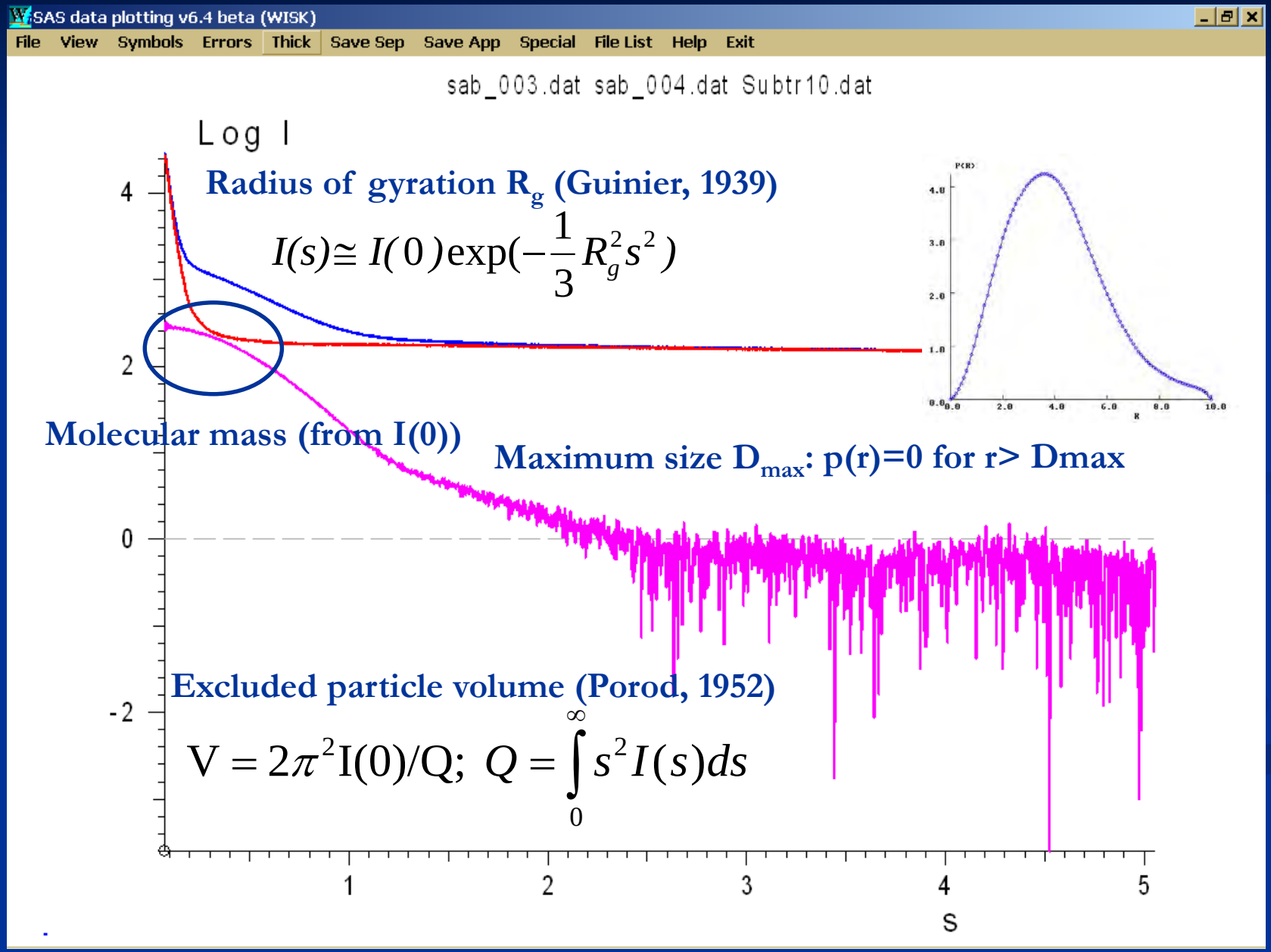
- Shape determination
- Rigid body modelling
- Missing fragments
- Oligomeric mixtures
- Hierarchical systems
- Flexible systems

Scattering from dilute macromolecular solutions (monodisperse systems)

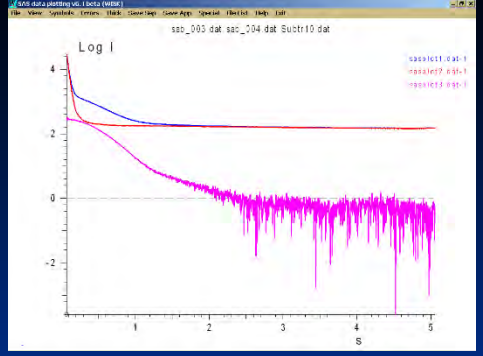
$$I(s) = 4\pi \int_0^D p(r) \frac{\sin sr}{sr} dr$$

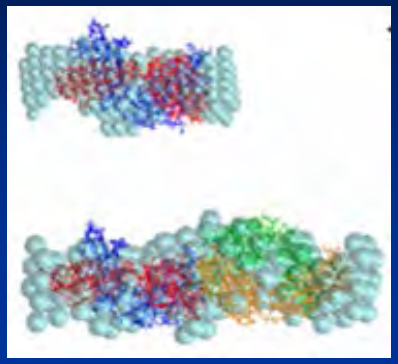
The scattering is proportional to that of a single particle averaged over all orientations, which allows one to determine size, shape and internal structure of the particle at low (*1-10 nm*) resolution.

Overall parameters



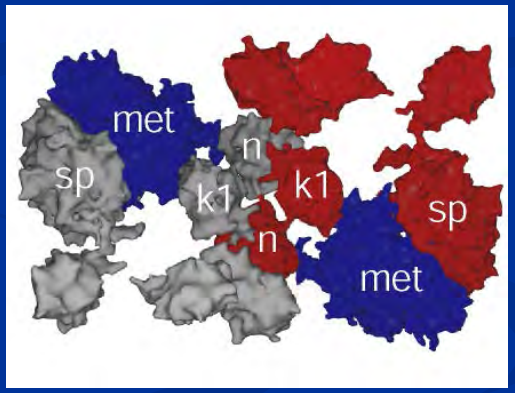
Low resolution structures of macromolecules in solution



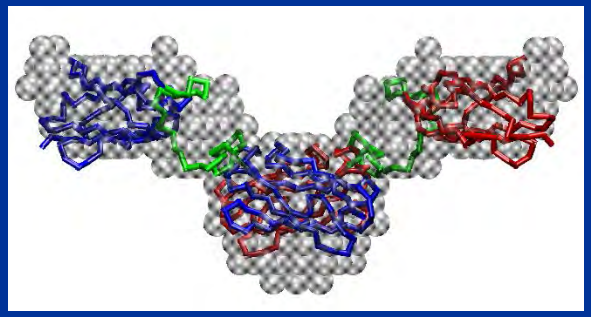


Shape and conformational changes of macromolecules and complexes

Validation of high resolution models and oligomeric organization



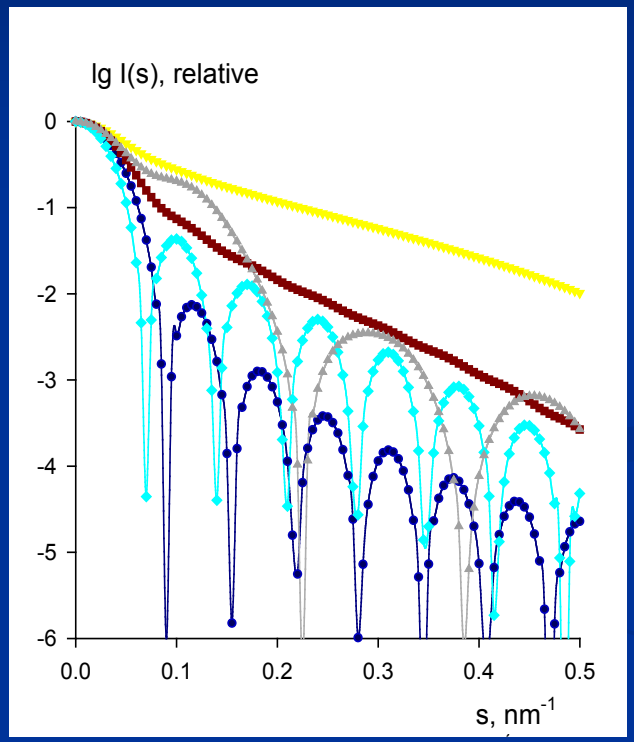
Rigid body models of complexes using high resolution structures

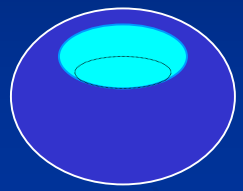


Addition of missing fragments to high resolution models

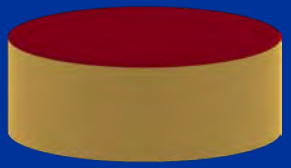
The scattering is related to the shape (or low resolution structure)

Solid sphere

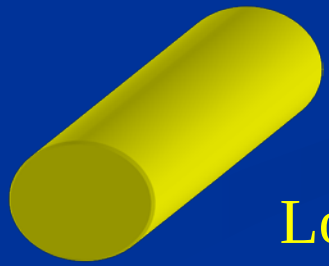




Hollow sphere



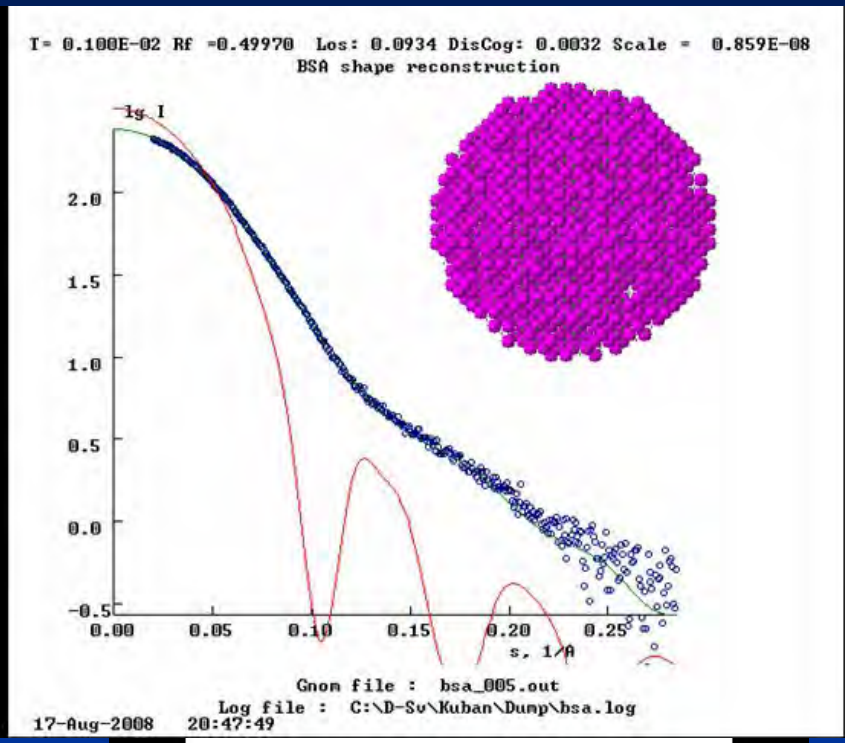
Flat disc



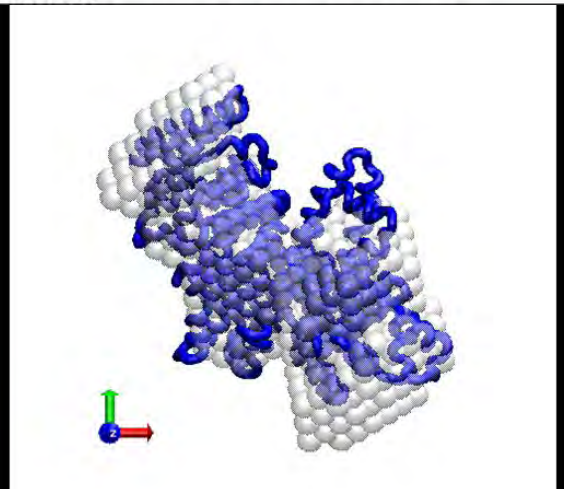
Long rod

Dumbbell

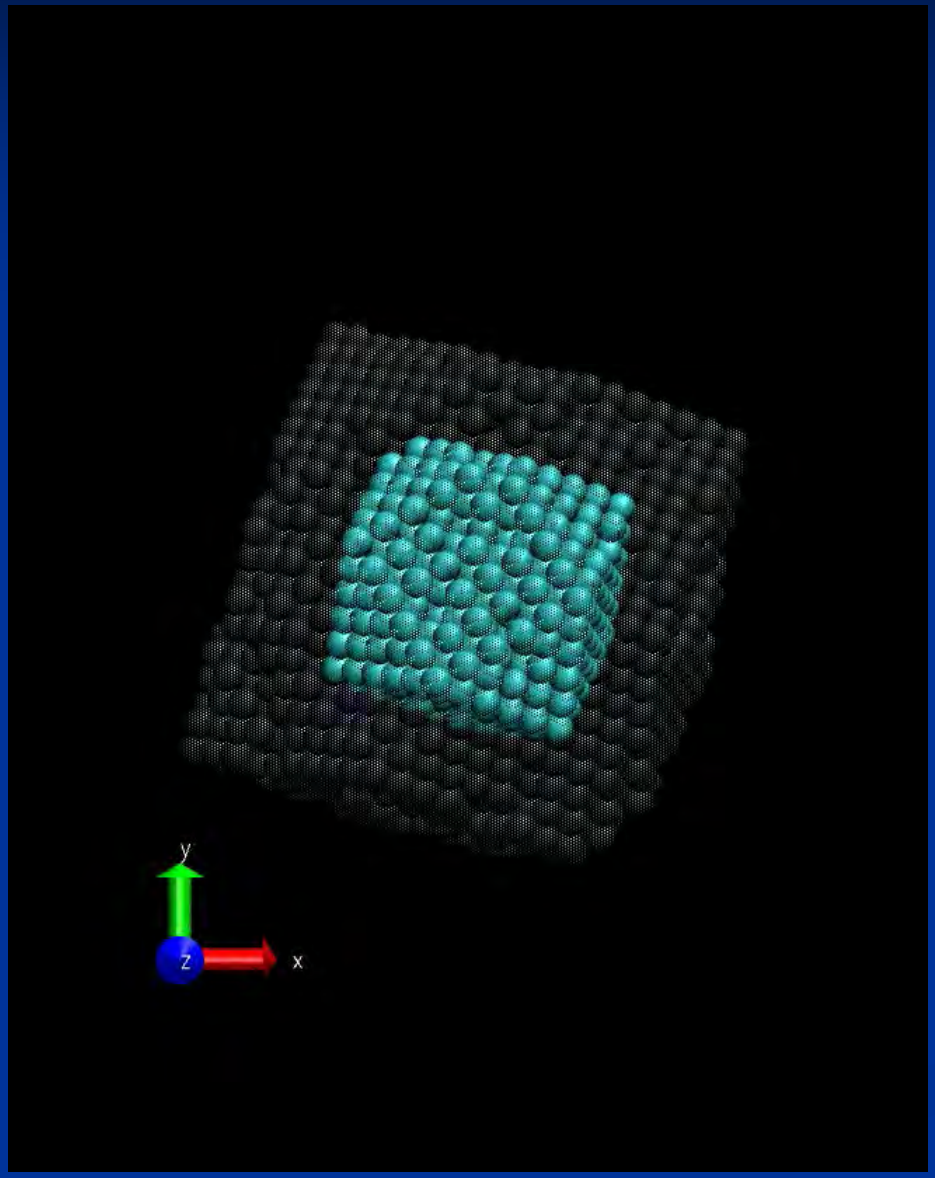
How to reconstruct 3D from 1D



DAMMIN: uses beads packed on a regular grid and simulated annealing to generate a (most possible) compact model fitting the experimental data



DAMMIF, a fast DAMMIN



DAMMIF is a completely reimplemented DAMMIN written in object-oriented code

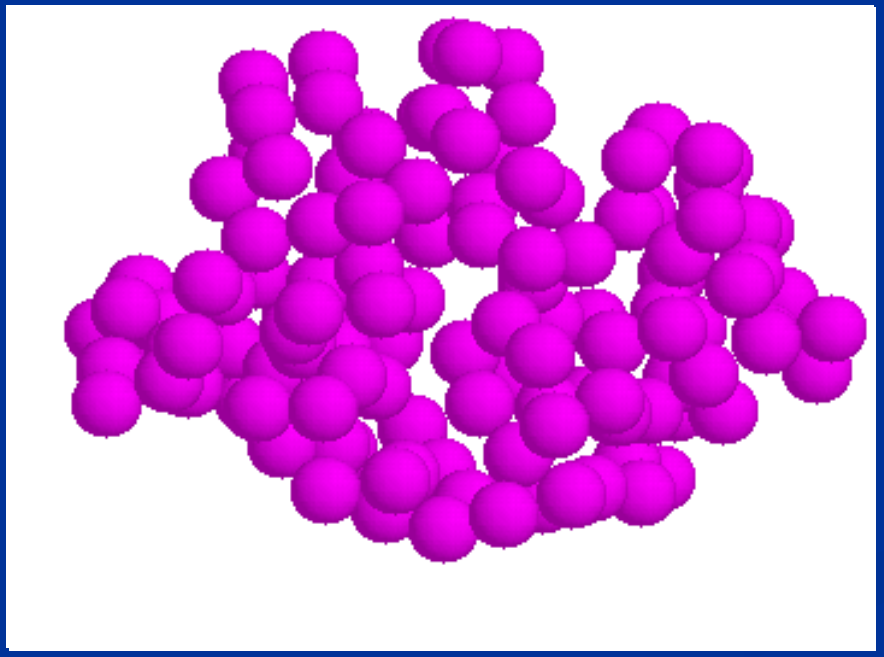
- **About 25-40 times faster than DAMMIN (in fast mode, takes about 1-2 min on a PC)**
- **Employs adaptive search volume**
- **Makes use of multiple CPUs**

Franke, D. & Svergun, D. I. (2009)
J. Appl. Cryst. **42**, 342–346

Ab initio dummy residues model

- Proteins typically consist of folded polypeptide chains composed of amino acid residues

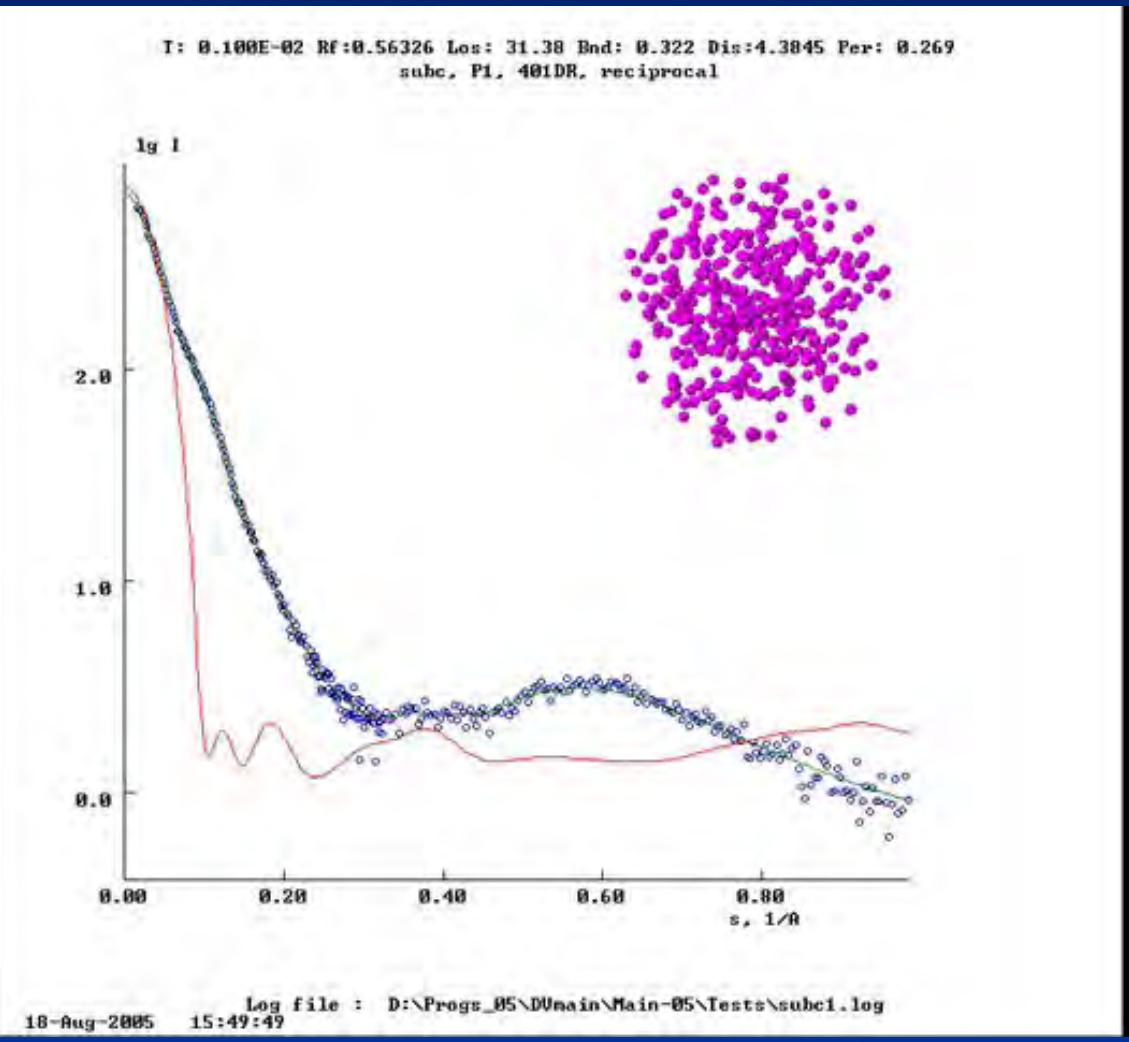
At a resolution of 0.5 nm a protein can be represented by an ensemble of K dummy residues centered at the Cα positions with coordinates $\{\boldsymbol{r}_i\}$



Scattering from such a model is computed using the Debye (1915) formula.

Starting from a random model, simulated annealing is employed similar to DAMMIN

GASBOR run on C subunit of V-ATPase



Starting from a random “gas” of 401 dummy residues, fits the data by a locally chain-compatible model

GASBOR run on C subunit of V-ATPase

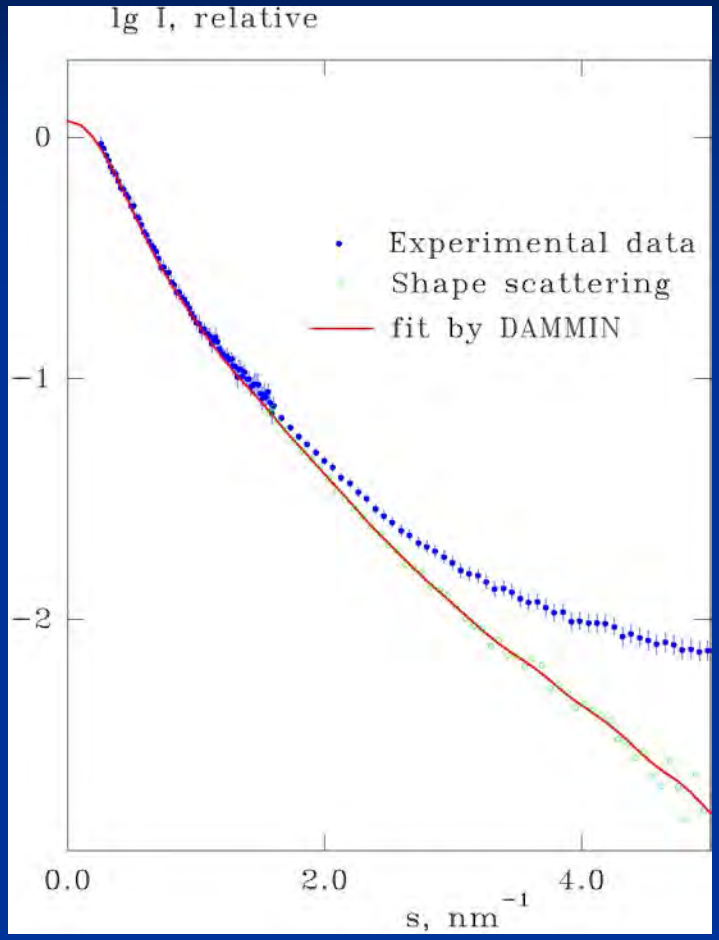
Beads: Ambruster *et al.* (2004, June) *FEBS Lett.* **570,** 119

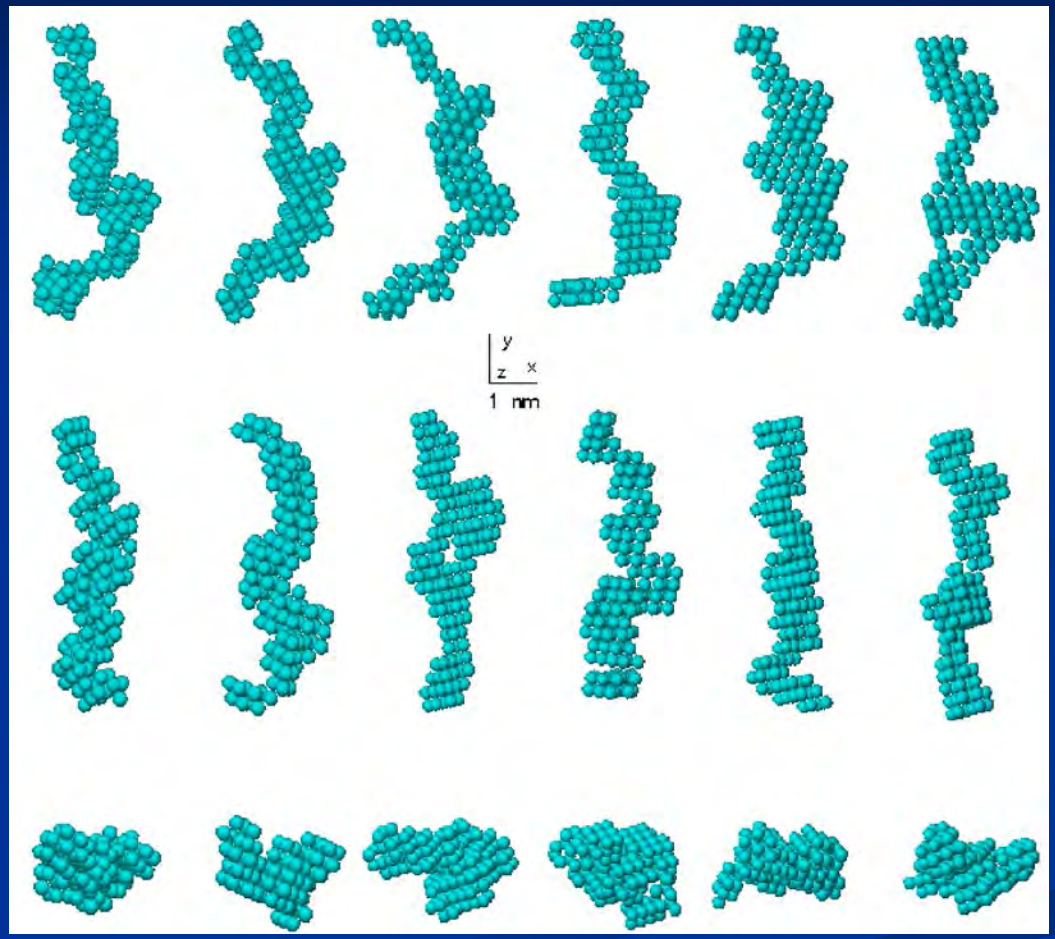
C_α trace: Drory *et al.* (2004, November), *EMBO reports,* **5**, 1148

Some words of caution

Or Always remember about ambiguity!

Shape determination of 5S RNA: a variety of DAMMIN models yielding identical fits

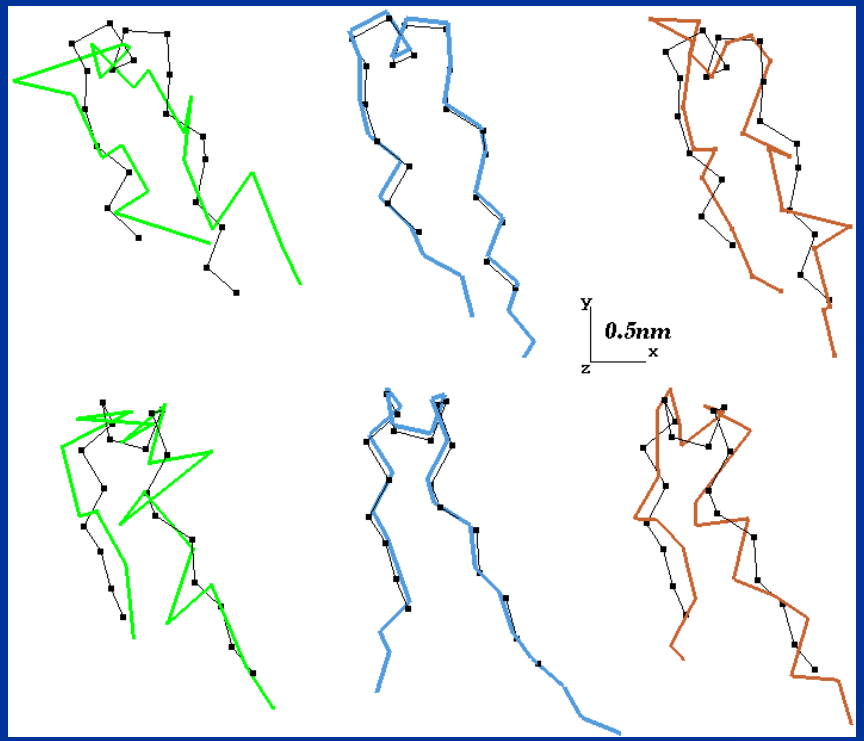


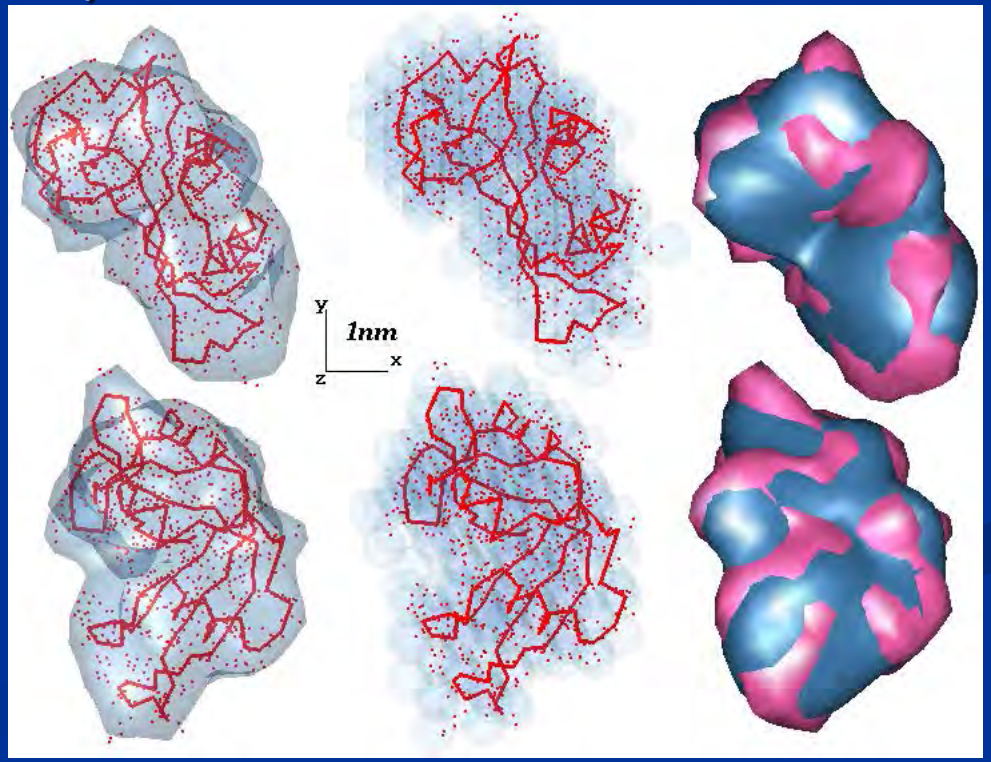


Funari, S., Rapp, G., Perbandt, M., Dierks, K., Vallazza, M., Betzel, Ch., Erdmann, V. A. & Svergun, D. I. (2000) *J. Biol. Chem.* **275**, 31283-31288.

Program SUPCOMB – a tool to align and conquer

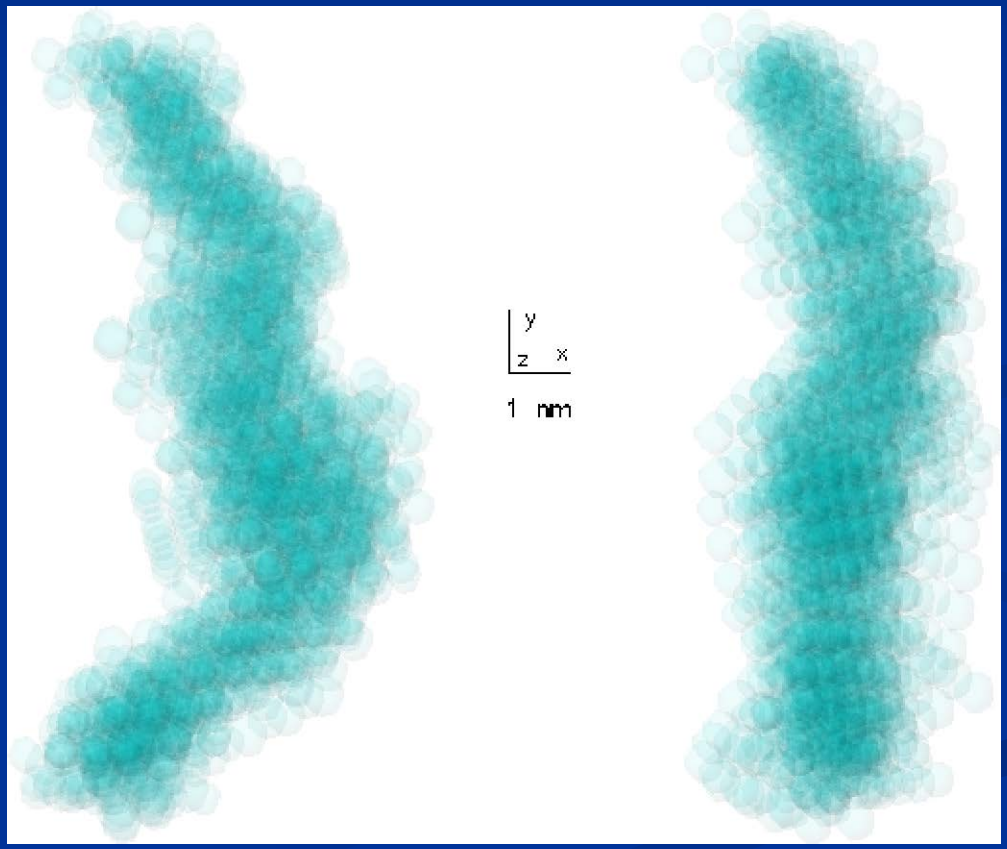
- Aligns heterogeneous high- and low-resolution models and provides a dissimilarity measure (NSD). For every point in the first model, the minimum value among the distances between this point and ALL points in the second model is found; the same is done for the second model. These distances are added and normalized against the average distances between the neighbouring points for the two models (computation time ~ N1*N2).





Kozin, M.B. & Svergun, D.I. (2001) *J. Appl. Crystallogr.* **34,** 33-41

5S RNA: ten shapes superimposed



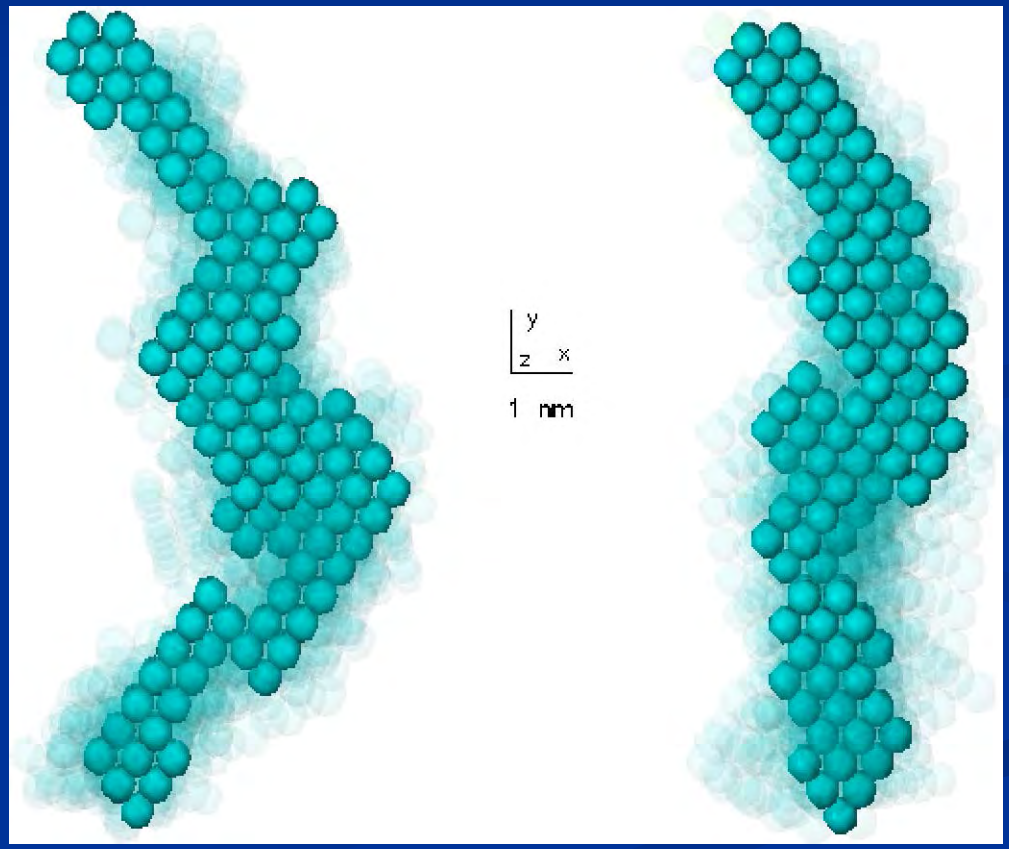
Most populated volume

Automated analysis of multiple models

1. Find a set of solutions starting from random initial seeds and superimpose all pairs of models with SUPCOMB.
2. Find the most probable model (which is on average least different from all the others) and align all the other models with this reference one.
3. Remap all models onto a common grid to obtain the solution spread region and compute the spatial occupancy density of the grid points.
4. Reduce the spread region by rejecting knots with lowest occupancy to find the most populated volume
5. These steps are automatically done by a package called DAMAVER if you just put all multiple solutions in one directory

Program DAMAVER, Volkov & Svergun (2003) *J. Appl. Crystallogr*. 36, 860

5S RNA: final solution



The final model obtained within the solution spread region

When biologists go for SAXS

SAXSMAN © A.Kikhney

This is just trivial case: SAS yields much more

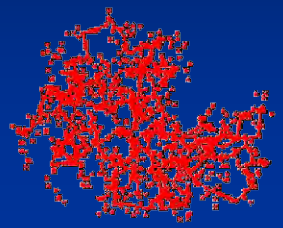
Modern life sciences widely employ hybrid methods

The most known and popular tool is, of course, Photoshop

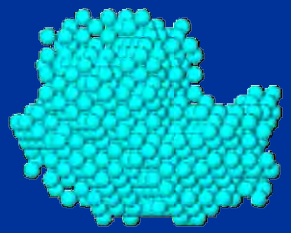
SAXS also allows for a very effective hybrid model building where high resolution portions are positioned to fit the low resolution scattering data

Scattering from a macromolecule in solution

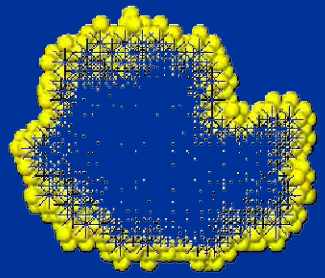
$$I(s) = \left\langle \left| A(\mathbf{s}) \right|^2 \right\rangle_\Omega = \left\langle \left| A_a(\mathbf{s}) - \rho_s A_s(\mathbf{s}) + \delta\rho_b A_b(\mathbf{s}) \right|^2 \right\rangle_\Omega$$



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



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

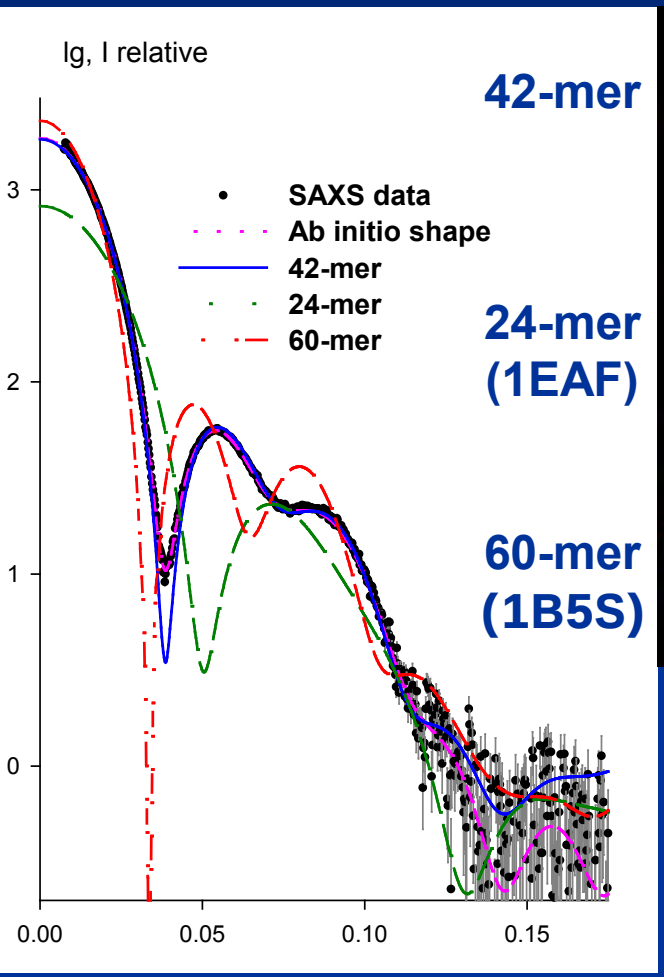


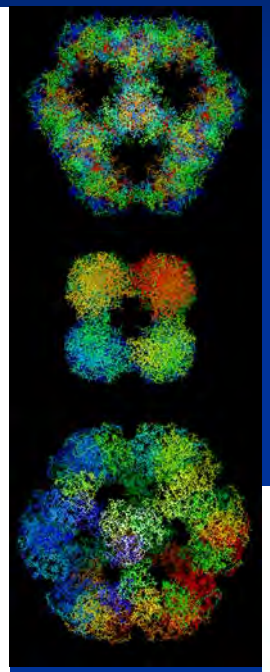
- $A_b(\mathbf{s})$: scattering from the hydration shell

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

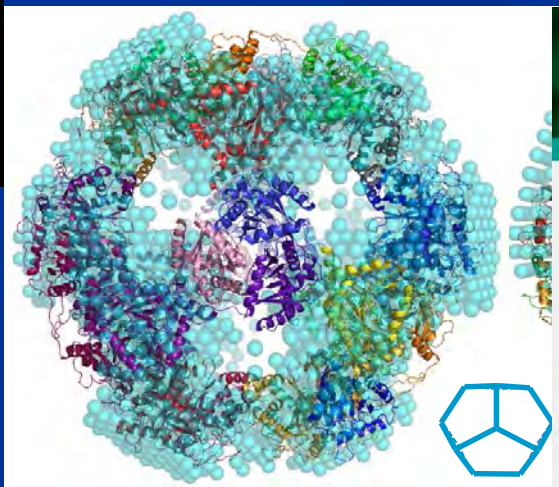
***CRYSON* (neutrons):** Svergun et al. (1998) *P.N.A.S. USA*, **95**, 2267

Catalytic core of E2 multienzyme complex is an irregular 42-mer assembly



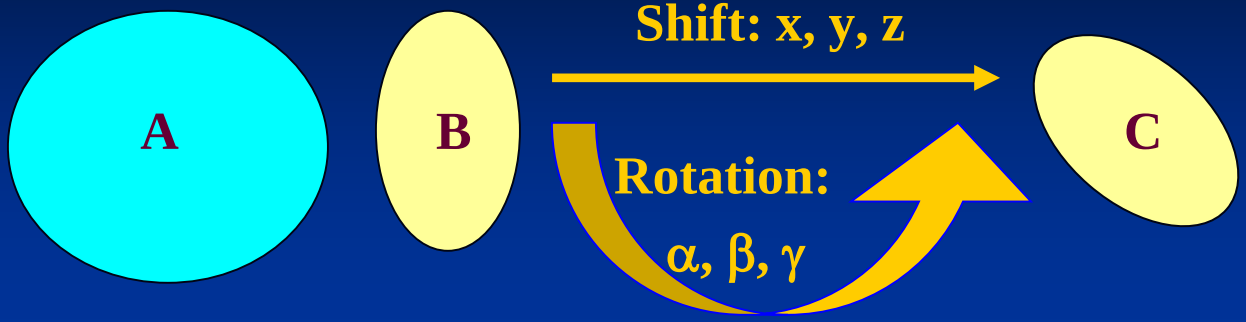


The E2 cores of the dihydrolipoyl acyl-transferase (E2) enzyme family form either octahedral (24-mer) or icosahedral (60-mer) assemblies. The E2 core from Thermoplasma acidophilum assembles into a unique 42-meric oblate spheroid. SAXS proves that this catalytically active 1.08 MDa unusually irregular protein shell does exists in this form in solution.



Marrott NL, Marshall JJ, Svergun DI, Crennell SJ, Hough DW, Danson MJ & van den Elsen JM. (2012) *FEBS J.* **279**, 713-23

Principle of rigid body modelling



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

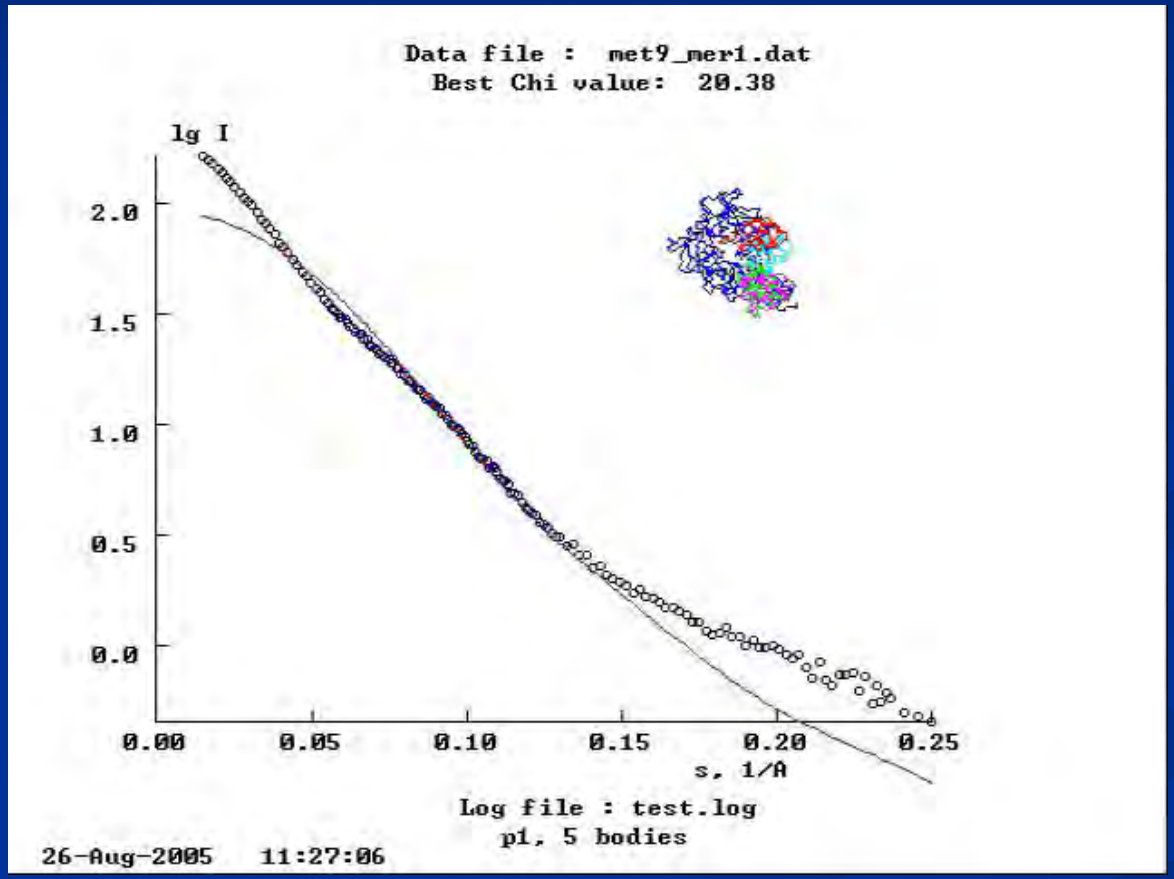
$$I(s) = I_A(s) + I_B(s) + 4\pi^2 \sum_{0}^{\infty} \sum_{-l}^{l} \mathrm{Re}\left[A_{lm}(s) C_{lm}^{*}(s)\right]$$

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

A global refinement run with distance constraints

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

Program SASREF



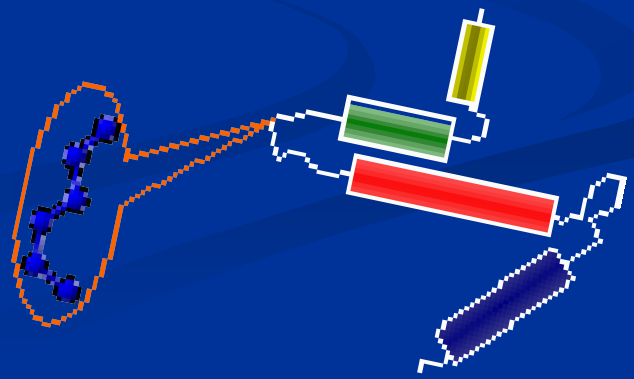
Single curve fitting with distance constraints: C to N termini contacts

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.

Addition of missing fragments

- Flexible loops or domains are often not resolved in high resolution models
- Their tentative configuration can be reconstructed by fixing the known portion and adding the missing parts to fit the scattering from the full-length macromolecule.

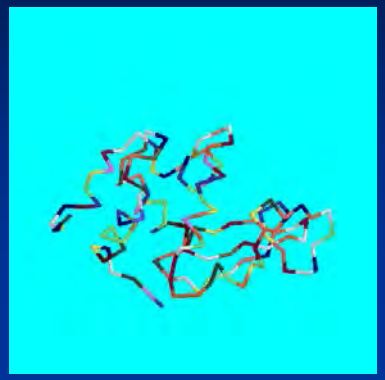
Moreover, addition of missing fragments can be combined with rigid body refinement (programs BUNCH and CORAL)



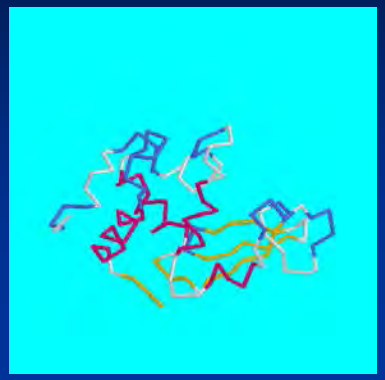
Petoukhov, M. V. & Svergun, D. I. (2005). *Biophys. J.* **89**, 1237-1250

Building native-like folds of missing fragments

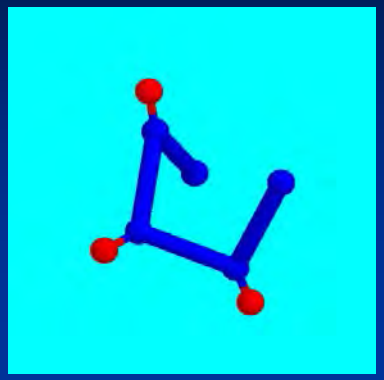
- Using DR-type models and protein-specific penalty functions



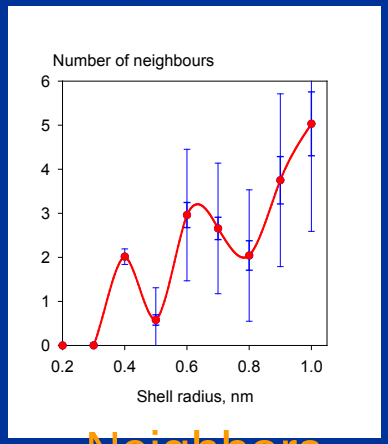
Primary sequence



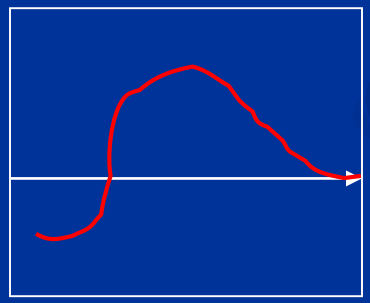
Secondary structure



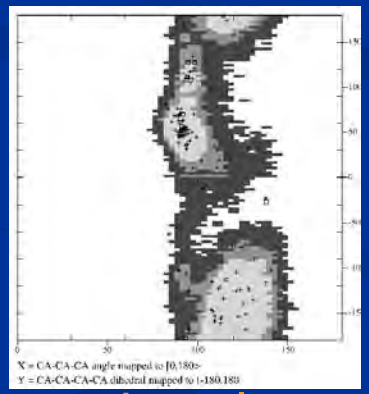
Excluded volume



Neighbors distribution



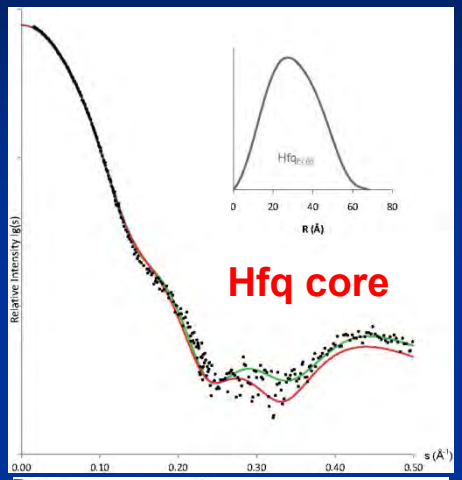
Knowledge-based potentials

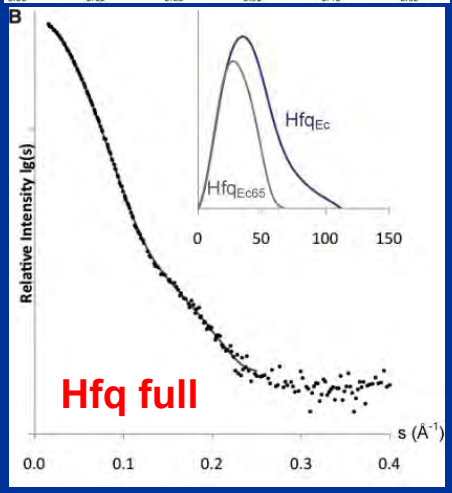


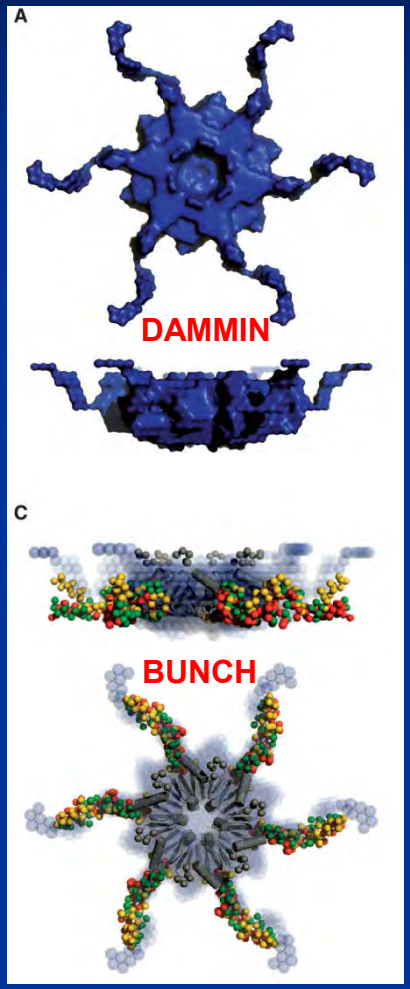
Bond angles & dihedrals distribution

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

Dynamics and function of the C-terminus of the *E. coli* RNA chaperone Hfq





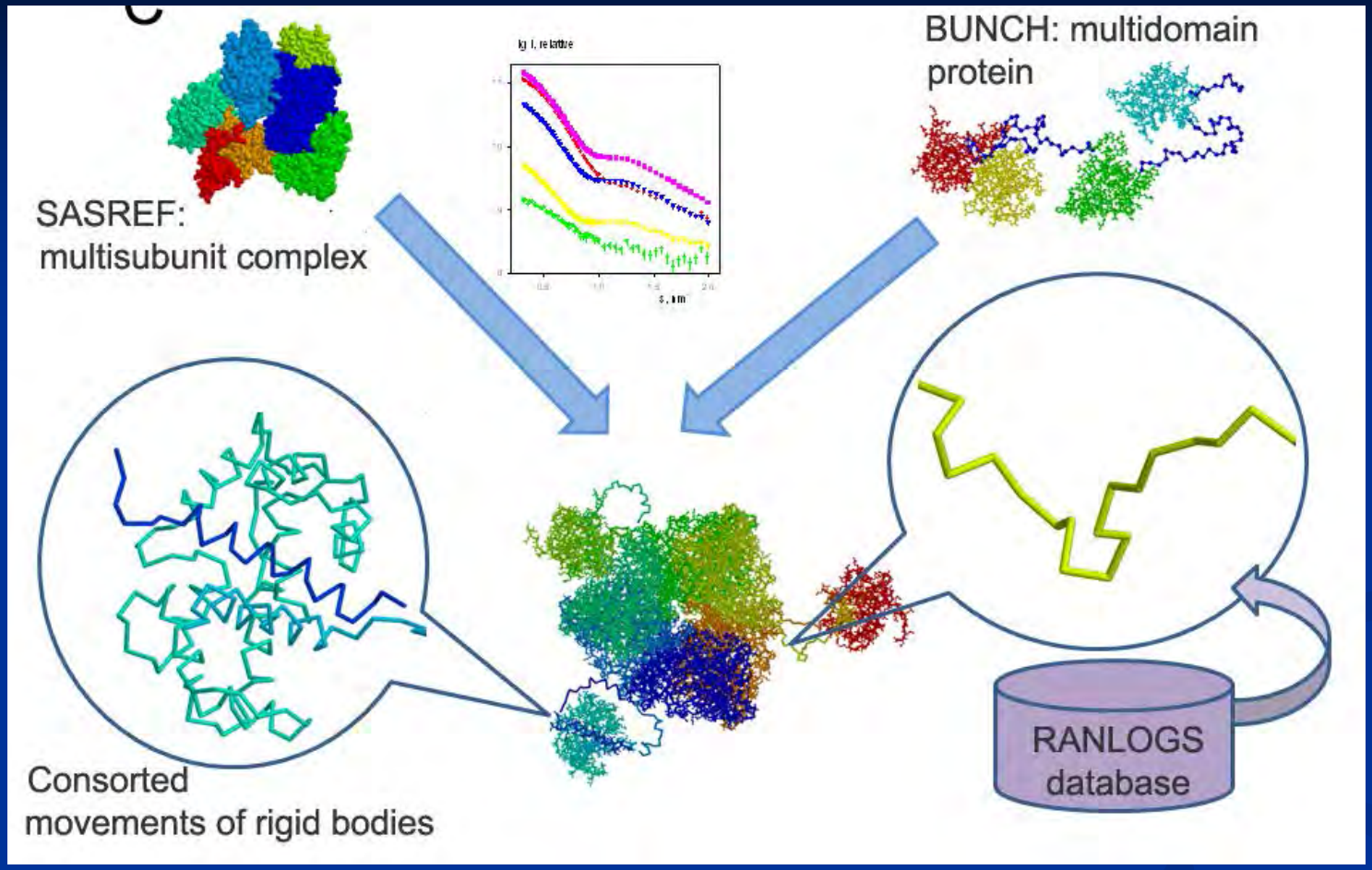


The hexameric Hfq (HfqEc) is involved in riboregulation of target mRNAs by small trans-encoded RNAs. Hfq proteins of different bacteria comprise an evolutionarily conserved core, whereas the C-terminus is variable in length.

By bioinfomatics, NMR, synchrotron CD and SAXS the C-termini are demonstrated to be flexible and to extend laterally away from the hexameric core. The flexible C-terminal moiety is capable of tethering long and structurally diverse RNA molecules.

Beich-Frandsen M, Vecerek B, Konarev PV, Sjöblom B, Kloiber K, Hämmerle H, Rajkowitsch L, Miles AJ, Kontaxis G, Wallace BA, Svergun DI, Konrat R, Bläsi U and Djinovic-Carugo K. (2011) Nucleic Acids Res. **39,** 4900-15

Addition of missing fragments: CORAL

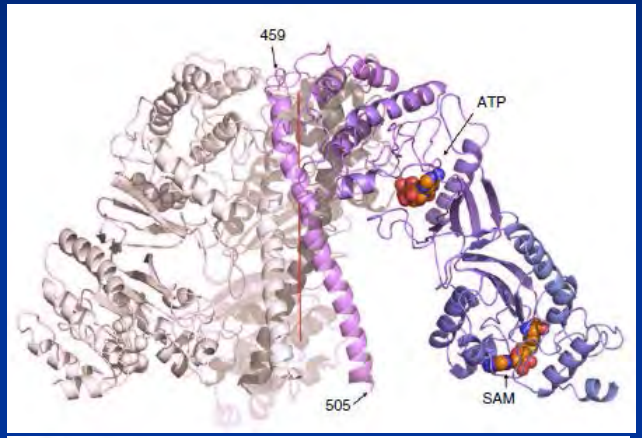


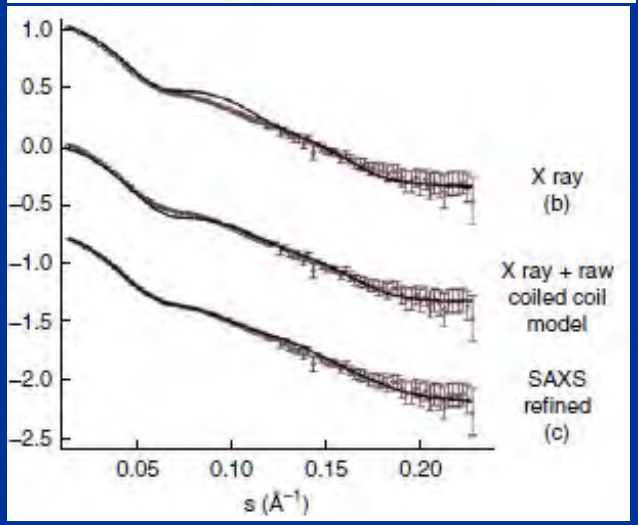
- A merger of SASREF and BUNCH: advanced methods to account for missing loops in multi-subunit protein structures (RANLOGS, CORAL)

M.V. Petoukhov, D. Franke, A. Shkumatov, G. Tria, A.G. Kikhney, M. Gajda, C. Gorba, H.D.T. Mertens, P.V. Konarev, D.I. Svergun (2012). *J. Appl. Cryst.* **45**, 342-350.

C-terminal domain of WbdD as a molecular ruler

In Escherichia coli O9a, a large extracellular carbohydrate with a narrow size distribution is polymerized from monosaccharides by a complex of two proteins, WbdA (polymerase) and WbdD (terminating protein).

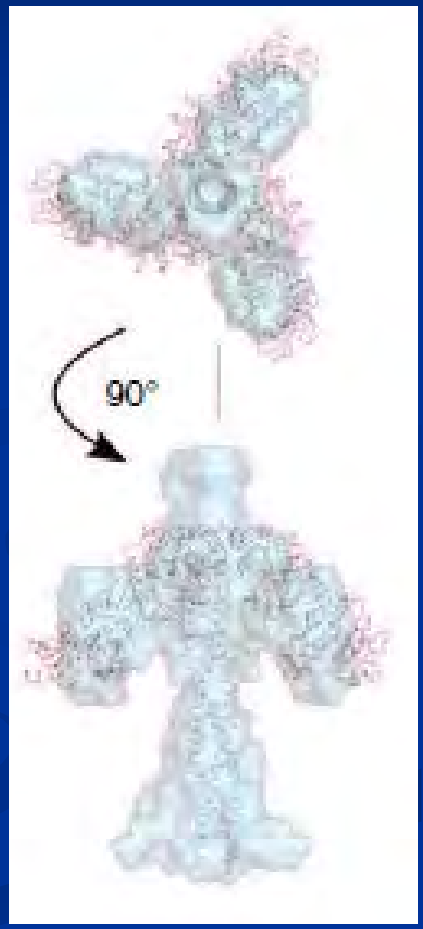




A truncated construct WbdD$^{1\text{-}459}$ is monomeric. For the construct WbdD$^{1\text{-}556}$ MX yields an active trimer but AAs 505-556 are not seen in the crystal.

SAXS *ab initio* shape reveals that the C-terminal is further extended. A rigid body model was constructed using coiled-coil C-terminal and refining the position of the catalytic domains.

In vivo analysis of insertions and deletions in the coiled-coil region revealed that polymer size is controlled by varying the length of the coiled-coil domain.



Hagelueken G., Clarke B. R., Huang H., Tuukkanen A. T., Danciu I., Svergun D. I., Hussain R., Liu H., Whitfield C. & Naismith, J. H. (2015) *Nat. Struct. Mol. Biol.*, **22**, 50-56.

Scattering from mixtures

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

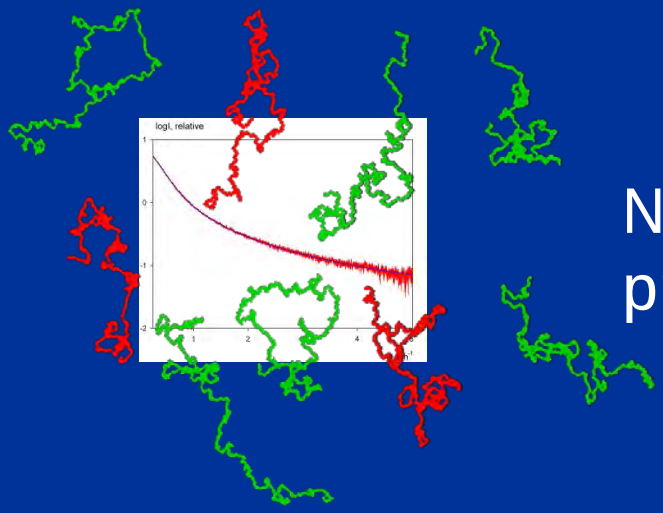
The scattering is proportional to that of a single particle averaged over all orientations, which allows one to determine size, shape and internal structure of the particle at low (*1-10 nm*) resolution.

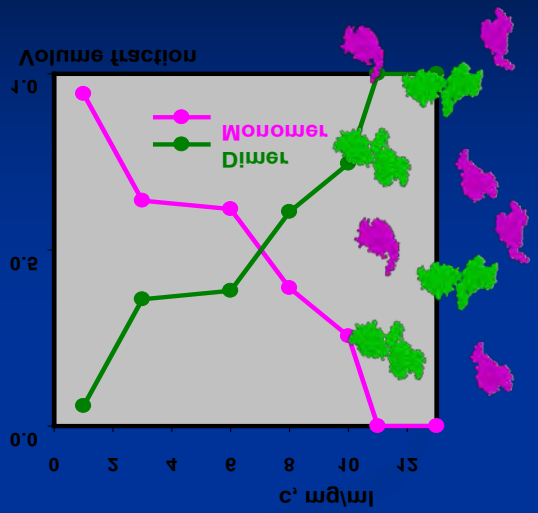
For equilibrium and non-equilibrium mixtures, solution scattering permits to determine the number of components and, given their scattering intensities $I_k(s)$, also the volume fractions

Flexible systems, interactions, mixtures and processes

Equilibrium oligomeric mixtures

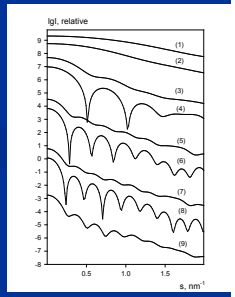
Stoichiometry and complex formation



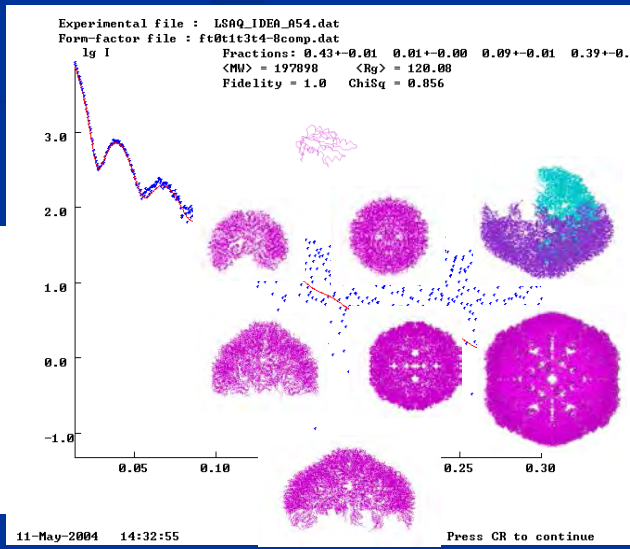


Natively unfolded proteins and multidomain proteins with flexible linkers

Protein folding/unfolding kinetics



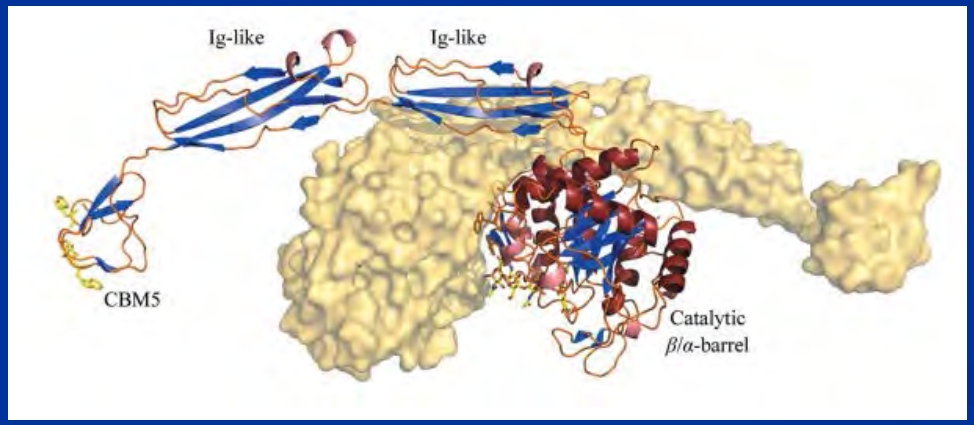
Assembly/disassembly processes

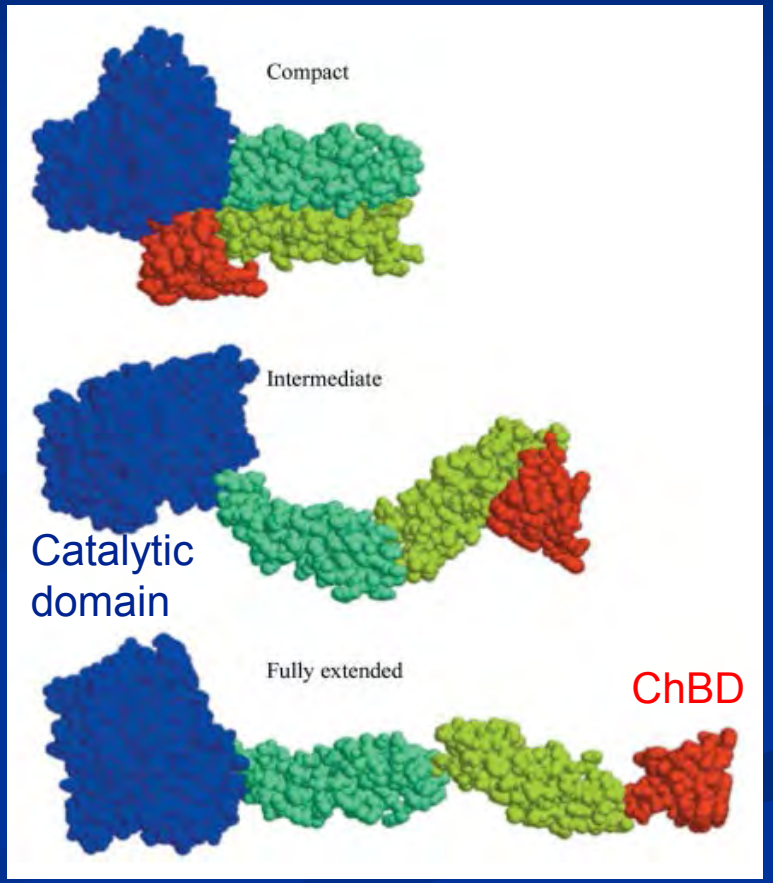


Crystal structures of substrate-bound chitinase from *Moritella marina* and its structure in solution

Chitinases break down glycosidic bonds in chitin and only few crystal structures are reported because of the flexibility of these enzymes.

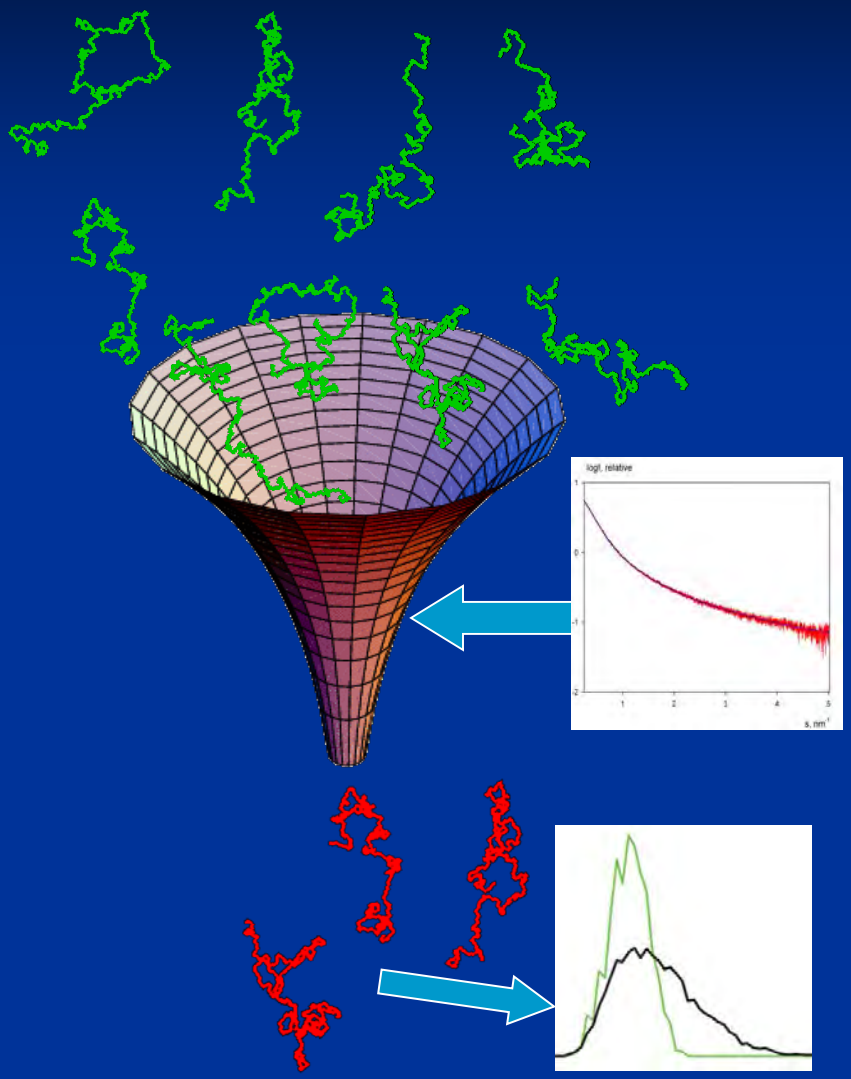
The dimeric crystal structure (at BESSY) of chitinase 60 from M. marina (MmChi60) contains four domains: catalytic, two Ig-like, and chitin-binding (ChBD). SAXS (at EMBL) demonstrates that MmChi60 is monomeric and flexible in solution. The flexibly hinged Ig-like domains may thus allow the catalytic domain to probe the surface of chitin.





P. H. Malecki, C. E. Vorgias, M. V. Petoukhov, D. I. Svergun and W. Rypniewski. *Acta Cryst.* (2014) **D70**, 676-684

Quantitative assessment of flexibility



EOM, Bernadó et al. (2007)
J. Am. Chem. Soc. **129**, 5656.

- Automated classification (folded, partially or completely unfolded) is available

D.Franke

DATCLASS

- In ensemble methods, one generates a large pool covering the conformational space and selects sub-ensemble(s) fitting the available experimental data
- EOM 2.0 (**G.Tria, 2015**): advanced pool generation, e.g. use of (partial) point symmetry
- Quantification of flexibility using enthropy and variation

ATSAS (All That SAS) roadmap

- World most comprehensive program suite for small-angle scattering data analysis from biomacromolecular solutions
- Consists of more then 25 programs developed at EMBL-HH since 1991
- Available for download since 1999, accessible on-line since 2006
- Presently, ATSAS has over 12,000 users from over 50 countries
- On-line usage: 750 users/30,000 jobs per year

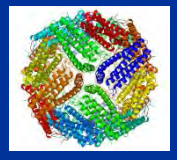
Raw data

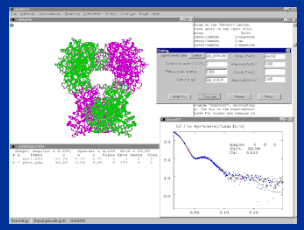
High resolution models

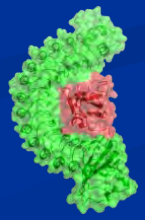
Computati… solution scattering from atomic models (X-rays and neutrons)

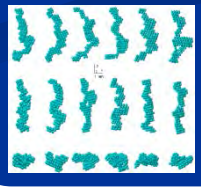
Multisubunit complexes modelling
Multidomain proteins modelling against multiple data sets

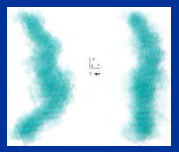
Models superposition, averaging and clustering











Recent ATSAS methods developments

Correlation Map: quantification of data fitting and an alternative to χ^2 when the experimenrtal error estimates are not available
D.Franke, C.Jeffries & D. Svergun (2015) *Nat. Methods*, **12**, 419-422

Automated determination of the useful data range by finding the number of reliable Shannon channels
P.Konarev & D.Svergun, (2015). *IUCrJ,* **2**, 352-360

Advanced ensemble analysis of flexibility: EOM version 2.0 including symmetry and quantitative characterization of the results
G.Tria, H.Mertens, M.Kachala & D.Svergun (2015) *IUCrJ*, **2**, 207-217

Intrinsic ambiguity of SAXS data: calculation of a propensity that a given scattering pattern yields ambiguous shape reconstruction
M.Petoukhov & D.Svergun, (2015) *Acta Cryst.* **D71**, 1051–1058

Correlation map (CM)

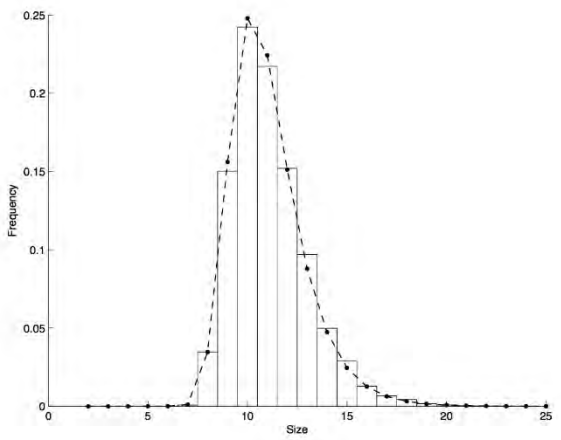
Lysozyme shape determination shown as signs of residuals

Parametric distribution for the probability of having the longest stretch of constant sign of the deviation (expressed analytically)

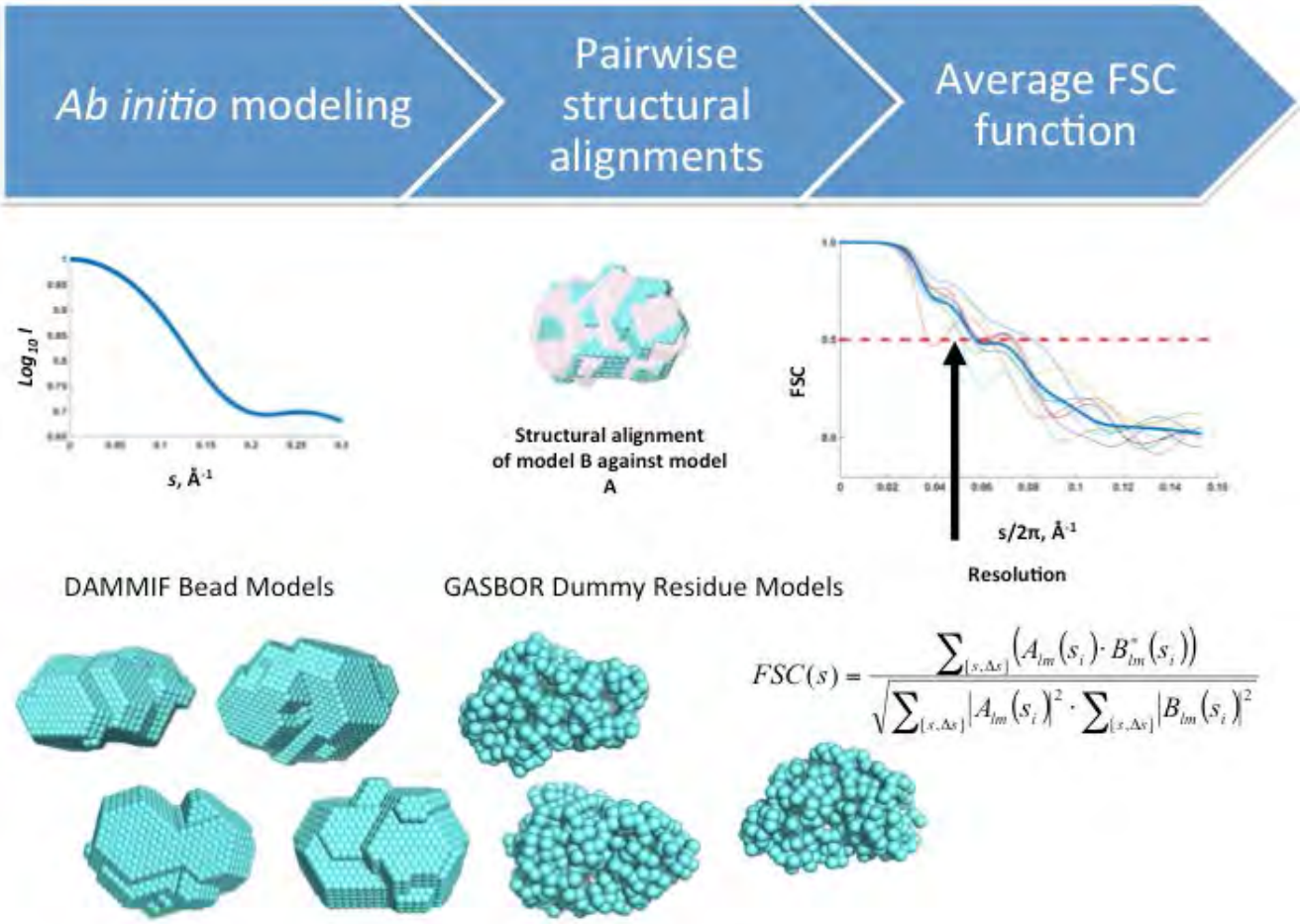
With this distribution, CM provides a p-value saying whether the fit is statistically acceptable. Numerous simulations demonstrated that CM has essentially the same statistical power as χ^2, but without the need of knowing the associated errors in the data

D.Franke, C.Jeffries

DATCMP



Variability and resolution of *ab initio* models



Analysis of the Fourier shell correlation functions between the multiple aligned models allows one (like in cryo-EM) to assess the resolution of the ensemble by identifying the point where the averaged correlation drops below 0.5. The approach is tested and validated in hundreds of model and real examples.

A.Tuukkanen, G.Kleywegt & D.Svergun, (2016), submitted.

Program SASRES, available online and for download in ATSAS 2.8

Standardization, databases, web servers

Report of the wwPDB Small-Angle Scattering Task Force http://www.wwpdb.org/task/sas Trewhella J. et al., (2013) Structure, **21**, 875

sasCIF format http://www.wwpdb.org/task/sas Kachala M., Westbrook J., Svergun D. (2016) J Appl Crystallogr. **49**, 302-310.

pE-DB: a database of intrinsically disordered and unfolded proteins http://pedb.vib.be Varad M. et al., (2014) Nucleic Acids Res., **42**, 326)

SASBDB, a Web repository of biological SAS data and models, www.sasbdb.org Valentini E, Kikhney AG, Previtali G, Jeffries CM, Svergun DI (2015) Nucleic Acids Res. **43**, 357-363

DARA, a Web server for rapid structural search using SAXS (http://dara.embl-hamburg.de) A.G. Kikhney, A. Panjkovich, A.V. Sokolova, D.I. Svergun (2016) Bioinformatics, **32**, 616-8.

DANESSA, an expert system for automated interpretation of a SAXS experiment given the data and available a priori information http://www.embl-hamburg.de/biosaxs/atsas-online/danessa.php (M. Petoukhov, in preparation)

"Table 1" and other publication rules for SAS

Recommend-dations of the SAS Task Force of the IUCr commission

Table 2. SAXS Data collection and derived parameters for CD27L.

	CD27L (wild-type)	CD27L (C238R)
Data collection parameters		
Instrument	EMBL X33 beam line (DORIS-III, DESY, Hamburg)	EMBL P12 beam line (PETRA-III, DESY, Hamburg)
Beam geometry	2.0×0.6 mm^2	0.2×0.12 mm^2
Wavelength (Å)	1.54	1.24
s range ($Å^{-1}$)[a]	0.01–0.6	0.01–0.46
Exposure time (s)	8×15	1 (20×0.05 s)
Concentration range (mg/mL)	0.9–4.0	1.0–8.5
Temperature (K)	283	283
Structural parameters[b]		
$I(0)$ (relative) [from $p(r)$]	44±2	3653±14
R_g (Å) [from $p(r)$]	33±1	43±2
$I(0)$ (cm^{-1}) (from Guinier)	45.6±0.5	3664±14
R_g (Å) (from Guinier)	33±1	42±1
D_{max} (Å)	106	147
Porod volume estimate ($Å^3$)	72151±10000	91690±10000
Excluded volume estimate ($Å^3$)	94000±10000	123000±10000
Dry volume calculated from sequence ($Å^3$)	39121/78219 (mon/dim)	
Molecular-mass determination		
$I(0)$ (cm^{-1}) BSA (66,000 Da)	71.4±0.4	3791±10
Molecular mass M_r (Da) [from $I(0)$]	42150±5000	63780±5000
Molecular mass M_r (Da) [from Porod volume (V_p/1.6)]	45094±5000	57306±5000
Molecular mass M_r (Da) [from excluded volume (V_{ex}/2)]	47000±5000	61500±5000
Calculated monomeric M_r from sequence	~32335	
Software employed		
Primary data reduction	RADAVER	
Data processing	PRIMUS/Qt	
Ab initio analysis	DAMMIF	
Validation and averaging	DAMAVER	
Rigid-body modeling	CORAL	
Equilibrium analysis	OLIGOMER	
Computation of model intensities	CRYSOL	
3D graphics representations	PyMOL, UCSF Chimera	

Abbreviations: M_r: molecular mass; R_g: radius of gyration; D_{max}: maximal particle dimension; V_p: Porod volume; V_{ex}: Particle excluded volume.

[a]Momentum transfer $|s| = 4\pi\sin(\theta)/\lambda$.

[b]Values reported for merged data sets (wild-type: 0.9 & 4.0 $mg.mL^{-1}$, C238R: 1 & 8.4 $mg.mL^{-1}$).

doi:10.1371/journal.ppat.1004228.t002

Sign in | Register

SASBDB

Small Angle Scattering Biological Data Bank

Search

Advanced search E.g. SASDBF4, Lyz, Nucleic Acids Res

Home | Browse | Submit data | About SASBDB | Help

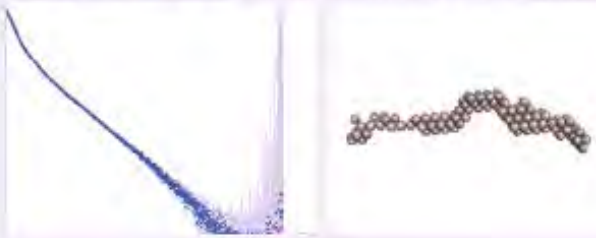
Curated repository for small angle scattering data and models

Small angle scattering (SAS) of X-ray and neutrons provides structural information on biological macromolecules in solution at a resolution of 1-2 nm.
SASBDB is a fully searchable curated repository of freely accessible and downloadable experimental data, which are deposited together with the relevant experimental conditions, sample details, derived models and their fits to the data.

SASBDB currently contains:
249 experimental data sets
394 models
103 experimental data sets on hold
153 models on hold

Recent depositions:

SASDBC4 – Plakin domain of Human plectin (spectrin repeats: SR3-SR9)



Sample: Plakin domain fragment of Human plectin encompassing spectrin repeats SR3-SR9 monomer, *Homo sapiens* protein

Buffer: sodium phosphate, pH: 7.5

Experiment: SAXS data collected at P12, Petra-III 2013-Aug-13

$R_g^{Guinier}$ 8.5 nm

D_{max} 35.0 nm

$Volume^{Porod}$ 135 nm^3

The Structure of the Plakin Domain of Plectin Reveals an Extended Rod-like Shape.
J Biol Chem 2016 Jul 13;
Ortega E, Manso JA, Buey RM, Carballido AM, Carabias A, Sonnenberg A, de Pereda JM

SASDB74 – CyaA Block I-V

Calcium-Driven Folding of RTX I

SASDBM4 – The 1:1:9:1 crRNA

Modulating the Cascade architec

SASDBS4 – Glutamate decarb

X-Ray Solution Scattering Study

SASDBU4 – Vaccinia primase

Domain organization of vaccinia

SASDAD7 – Structure of a con

Crystal Structure of the CTP1L E

Database development and submission curation is done in Hamburg.
Presently offers 249 data sets and 394 macromolecular models (world largest)

DARA, a rapidly searchable database of over 150,000 SAXS patterns from the entire PDB

dara.embl-hamburg.de

EMBL Biological Small Angle Scattering

Home > Web services > ATSAS online > DARA

DARA

Rapid search of structural neighbours using solu

GNOM file (*.out) or experimental data (*.dat) or simulated data (*.int) or model (*.pdb) Browse... No file selected

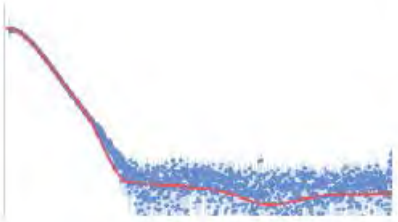
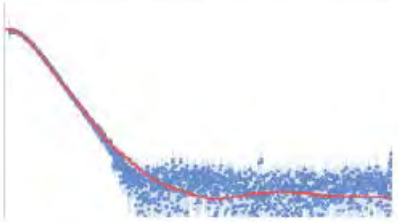
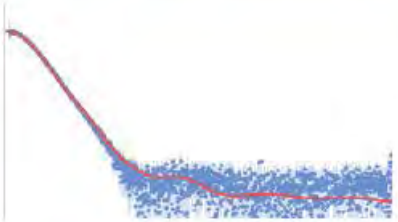
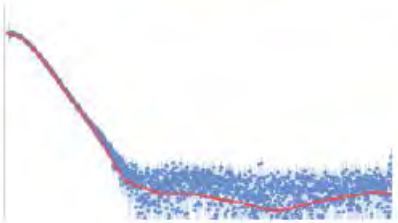
Angular units 1/nm s =

Macromolecule type protein

Show 10 neighbours

Sample input: SAXS data from protein solution

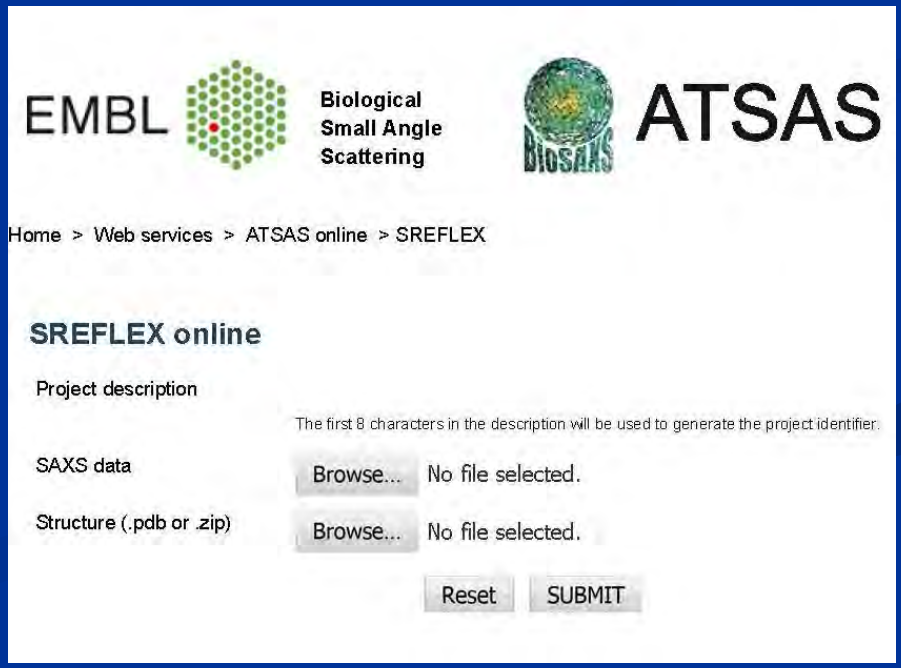
DARA neighbours

	Fit	χ^2	PDB ID	Structure	MW	Volume	R_g	D_{max}
1		1.67	1YUC	 69% α 3% β	59.8 kDa	103 420 Å^3	31 Å	110 Å
2		1.69	1ZW0	 85% α 0% β	53.3 kDa	92 800 Å^3	31 Å	106 Å
3		1.71	3HFD	 47% α 15% β	50.2 kDa	89 960 Å^3	30 Å	98 Å
4		1.81	2AWH	 68% α 5% β	58.6 kDa	102 810 Å^3	29 Å	91 Å

A Brand New Method and Server

Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis

A. Panjkovich, D.I. Svergun (2016) *Phys Chem Chem Phys.* **18**, 5707-19



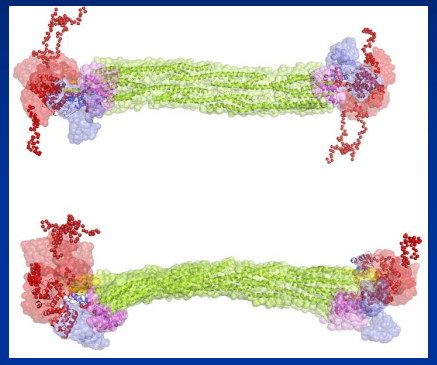
High brilliance EMBL SAXS beamline P12

- Robotic EMBL/ESRF sample changer
- Automated FPLC/HPLC in parallel with biophysical sample characterisation

- About 10^{13} ph/sec in 200*120 mm^2
- Energy between 4 and 20 keV (3.0 to 0.6 Å)
- Divergence below 0.05*0.05 $mrad^2$
- Multilayer monochromator mode: over 5×10^{14} ph/sec
- SASFLOW pipeline for on-line data processing and analysis
- Full automation, remote and mail-in access

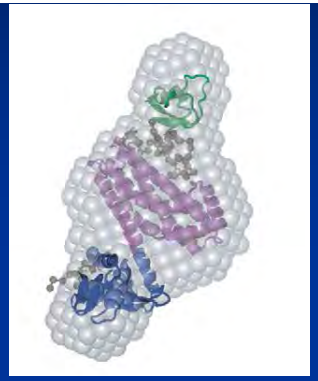
Highlights of EMBL SAXS user publications

Human Muscle α-Actinin



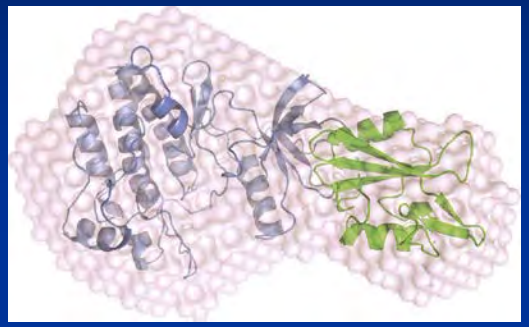
Ribeiro *et al*
Cell (2014)

Conformational switch in collybistin



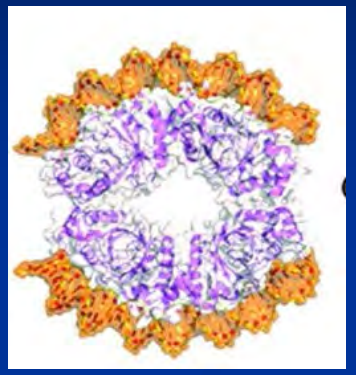
Soykan *et al*
EMBO J (2014)

KD/SH2 domains of Abl kinase



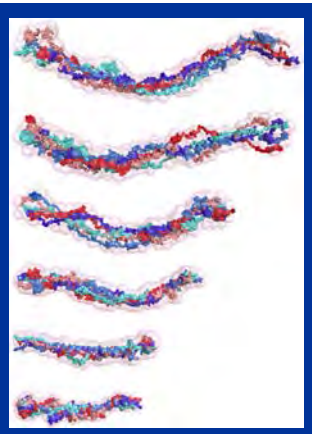
Lamontrara *et al*
Nat. Comm. (2014)

kLANA/DNA complexes



Ponnusamy *et al*
NAR (2015)

Surface protein SASG

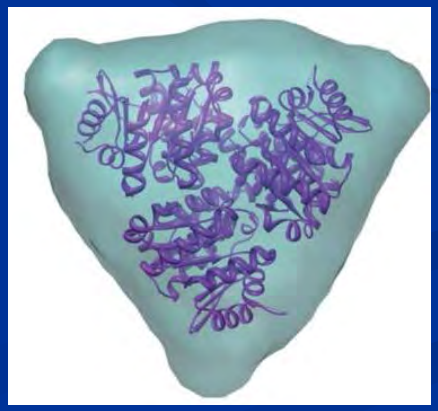


Gruszka *et al.*
Nat. Comm. (2015)

WbdD as a molecular ruler

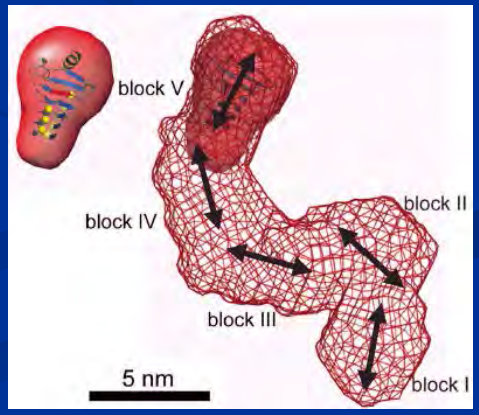
Hagelueken *et al*
NSMB (2015)

SaThiM from vitamin B1 synthetic pathway



Drebes *et al*
Sci. Rep. (2016)

Folded RTX Domain of CyaA



Bumba *et al*
Mol. Cell (2016)

What does SAS tell about biological macromolecules

- Nothing known: *ab initio* low resolution structure
- Incomplete high resolution structure known: probable configuration of missing portions
- Complete high resolution structure known: validation in solution and biologically active oligomers
- High resolution structure of domains/subunits known: quaternary structure by rigid body refinement
- Mixtures/assemblies: volume fractions of components
- Flexible systems: quantitative analysis of configurational ensembles

What does SAS tell about biological macromolecules

- Nothing known: *ab initio* low resolution structure
- Incomplete high reso… structure kn… configu…
- Com… in s…
- High… quat…
- Mixtu…

…ghput SAXS

…ents

Not just high throughput: important present and future applications of SAXS are functional complexes and processes (flexible, dynamic, transient, evolving), where SAXS is among the few methods providing quantitative structural information

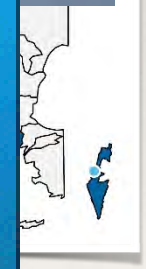
Syn… …aterial, concentration from 0.5 mg/ml, exposure times: (sub)seconds

Some words of caution

- Always check your samples BEFORE doing SAS!
- Use the other methods and NEVER trust them blindly!
- Always check integral parameters BEFORE 3D modelling!

*i*NEXT – Infrastructure for Structural Biology

*i*nfrastructure for **N**MR, **E**M & **X**-rays for **T**ranslational research



2016

NEXT

Acknowledgments

- **All BioSAXS Group** and especially C. Blanchet, C. Jeffries, D. Franke, A.Kikhney, H.Mertens, A.Panjkovich, M. Graewert, C.Kerr, A.Tuukkanen, N.Hajizadeh
- **Former BioSAXS Group members** and especially M. Petoukhov, P.Konarev, E.Valentini, M.Kachala, G.Tria
- **EMBL-HH:** R.Meijers, C.Loew, S.Fiedler, M.Wilmanns
- **Other EMBL units:** E.Lemke, G.Kleywegt, F.Cipriani
- **External collaborators**: J. Trewhella (Sydney), P.Tompa (Brussels), J.Westbrook (RCSB Rutgers), V.Volkov (Moscow), C.Betzel (Hamburg), M.Roessle (Luebeck), numerous users of P12