

Analysis of X-ray and neutron scattering from biomacromolecular solutions

Maxim V. Petoukhov and Dmitri I. Svergun

EMBL, Hamburg Outstation, Notkestraße 85, D-22603 Hamburg, Germany, and Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia

Corresponding author: Dmitri Svergun, EMBL c/o DESY
Notkestrasse 85, D-22603 Hamburg, GERMANY
Tel: +49 40 89902 125, Fax: +49 40 89902 149
E-mail: Svergun@EMBL-Hamburg.DE

Running title: Biological small-angle solution scattering

Keywords: small-angle scattering, solution scattering, macromolecular structure, functional complexes, *ab initio* methods, rigid body modelling, flexible macromolecules

Summary

New developments in small-angle X-ray and neutron scattering studies of biological macromolecules in solution are presented. Small-angle scattering is rapidly becoming a streamline tool in structural molecular biology providing unique information about overall structure and conformational changes of native individual proteins, functional complexes, flexible macromolecules and assembly processes.

Introduction

The tremendous recent progress in high resolution structure determination of individual proteins due to structural genomics initiatives [1] underlined the ultimate importance of shedding light on the structure of macromolecular complexes. The latter are accomplishing most important cellular functions and the focus of modern structural biology is increasingly towards their study [2]. Structural analysis of the functional complexes, which are often of transient and flexible nature, requires a synergy of complementary approaches covering a broad range of macromolecular sizes, experimental conditions and temporal and spatial resolution.

Small-angle scattering (SAS) probes the structure of native biological macromolecules in solutions at a low (1-2 nm) resolution [3]. The dissolved macromolecules are exposed to a collimated X-ray (SAXS) or neutron (SANS) beam and the scattered intensity I is recorded by the detector (Figure 1). For dilute solutions (concentrations in mM range) the particles are chaotically distributed leading to isotropic intensity depending only on the scattering angle 2θ between the incident and scattered beam (here, only elastic scattering is considered, and the radiation wavelength λ remains unchanged). For monodisperse solutions, the intensity $I(s)$ after subtraction of the separately measured solvent scattering is proportional to the scattering from a single particle averaged over all orientations (here, $s = 4\pi\sin\theta/\lambda$ is the momentum transfer). The magnitude of the useful signal is

proportional to the number of particles in the illuminated volume (i.e. to their concentration) and to the squared contrast, which is the difference between the scattering length density of the particle and of the solvent. For X-rays, electron density contrast counts; for neutrons, nuclear and /or spin density contrast is relevant (importantly, neutrons are sensitive to isotopic H/D exchange, which is experimentally used for contrast variation).

The SAS patterns directly provide parameters such as molecular mass (MM), radius of gyration (R_g), hydrated volume (V) and maximum diameter (D_{max}). Since nineteen sixties, SAS was used to acquire structural information about the overall shapes of proteins in the absence of crystals. During the last decade, novel approaches were developed to interpret SAS data from solutions of biological macromolecules in terms of three-dimensional (3D) models. These approaches together with the progress in instrumentation (most notably, high brilliance synchrotron beamlines) paved the way for the present renaissance of SAS in structural biology (see e.g. [4] for a review).

SAS is hardly limited by the particle size, being equally applicable to smallest proteins and to huge macromolecular machines like ribosomes and largest viruses. SAS experiments are usually fast (down to sub-second range on a third generation synchrotron) and require moderate amount of purified material (about a few mg for a complete study). Structural responses to changes in external conditions (pH or salt concentration, temperature, ligand binding etc) can be directly correlated with functional results (e.g. kinetics, spectroscopy, or interaction studies in solution). SAS is extremely useful in the quantitative characterization of mixtures and flexible systems, including time-resolved experiments to monitor assembly or folding processes.

The variety of structural questions addressed by SAS is schematically illustrated in Figure 1. In the absence of complementary information, low resolution macromolecular shapes can be reconstructed *ab initio* and overall characteristics of flexible systems can be extracted. If a high resolution structure or a predicted model is available, it can be validated against the experimental SAS data, and oligomeric state and/or oligomeric composition in solution can be determined. If an incomplete high resolution model is known, approximate configurations of missing fragments can be obtained. Probably the most important approach for macromolecular complexes is rigid body modelling, when the structures of individual subunits are available and SAS is employed to build the entire complex. SAS can be effectively combined with other structural, computational and biochemical methods as indicated in Figure 1, the major recent advances in the field will be presented below with a special emphasis on the studies of functional complexes.

***Ab initio* methods**

A decade ago it was hardly believed that 3D models can be directly reconstructed, even at low resolution, from the one-dimensional SAS patterns; nowadays, *ab initio* shape determination is an established procedure. The first publicly available *ab initio* method to determine the angular envelope function [5] had limited range of applications to shapes without internal holes. More versatile methods represent the particle by finite volume

elements, and, starting from a random configuration, employ different flavours of Monte-Carlo search to fit the experimental data by a physically sound (compact and interconnected) model. In the first such method [6], genetic algorithm was used to assign beads in a spherical search volume with diameter D_{max} either to the particle or to the solvent to yield a configuration, whose computed scattering fits the experimental profile. Non-spherical search volumes including those from electron microscopy (EM) may also be used. Other methods are now available using models represented by beads or ellipsoids and different minimization procedures [7-9]. Several papers comparing these methods [10,11] and numerous practical applications proved that the shape determination allows one to reliably reconstruct the low resolution macromolecular structure. Given that different shapes are obtained for different random seeds, multiple *ab initio* runs are often performed and analysed to reveal the most persistent features of the model [12,13**].

The resolution of the shape determination methods is limited by the assumption of particle homogeneity. For proteins, the level of detail is improved by the use of dummy residues (DR) [14] more adequately representing the internal structure. In a general multiphase approach for macromolecular complexes [7,15*] the beads correspond to different components of the particle having different contrasts. Instead of "black-and-white" (particle-solvent) models from shape determination, "color" models are built to reveal not only the shape but also the distribution of components inside the particle. These more detailed models can be constructed by simultaneous fitting of contrast variation SANS data (*e.g.* recorded from a nucleoprotein complex in different H₂O/D₂O mixtures utilizing natural contrast between nucleic acids and proteins [3]). Using specific perdeuteration of individual proteins, the approach is effective also for multiprotein complexes [16**]. Multiple curves fitting also applicable in X-rays from multidomain or multisubunit proteins, if the scattering from partial constructs is available, and the contrast the domain/subunit is defined for each pattern by its absence or presence in the given construct [17*].

Ab initio methods are currently used for various macromolecules and complexes, including *e.g.* the studies of the complex between a hamster prion and DNA [18*], of autoinhibited and mutated c-Abl tyrosine kinase [19**], of human fibrillin-1 [20**], of choline binding proteins [21*], of the complex between the DNA ligase and proliferating cell nuclear antigen [22**], of human pyruvate dehydrogenase complex [23*]. Information about particle symmetry, if available can be explicitly used in the *ab initio* programs [7,14,15*], further improving the resolution of the models. Recent examples of symmetric reconstructions are dimeric bacteriophytochrome [24*], hexameric transcriptional activator NtrC [25*], octameric potassium channel/calcium sensor protein complex [26*]. Shape analysis is also applicable to nucleic acids [27*] and to proteins, which can only be solubilized in the presence of detergent (*e.g.* membrane proteins). The detergent contribution is masked out by SANS in solutions containing appropriate amount of D₂O (see a recent application to a human apolipoprotein B-100 [28*]).

High resolution crystal structures of components, if available, are often docked into the obtained low resolution shapes. *Ab initio* analysis is usefully complemented also by other methods, in particular, EM [20**,25*,29*] or NMR [18*]. Attempts at SAS-assisted *ab*

initio folding of proteins did not yet result in publicly available algorithms, although predicted secondary structure elements may be used in the addition of missing fragments to high resolution models (see below).

Rigid body modelling

Although the docking of the high resolution structures into the *ab initio* shapes is a possible way of studying the macromolecular complexes, given the limited resolution of the SAS-derived shapes a direct modelling against the experimental scattering data is preferable. A necessary prerequisite is a rapid computation of the scattering from the complex given the structures of the subunits. SAXS/SANS from atomic models is evaluated e.g. by the programs CRY SOL [30] and CRYSON [31], respectively, and their output can be used in the fast algorithms computing scattering from complexes by spherical harmonics expansion [32]. Interactive modelling programs [33] have already long been used; recently [15*,34] a set of tools for automated modelling was made publicly available. The most comprehensive program SASREF uses simulated annealing (SA) for a simultaneous fitting of multiple SAXS/SANS patterns. The modelling can be constrained by symmetry, inter-residue contacts from mutagenesis, distances from Fourier transform infrared spectroscopy or subunit orientations from residual dipolar coupling in NMR (the latter has been shown to be very useful [35,36]). Other rigid body modelling procedures are also published, employing exhaustive semi-automated search [37*,38] or Monte-Carlo based methods [39].

The rigid body analysis requires complete structures of all subunits. In practice, loops or entire domains are often missing in high resolution crystallographic and NMR models. The methods initially developed to find probable configurations of the missing fragments [40] were coupled with rigid body algorithms in the program BUNCH [13**,34]. SA is employed to move/rotate the domains as rigid bodies and to change the local conformation of the DR chains representing the unknown fragments. The method is applicable to multisubunit complexes but also to multidomain proteins connected by linkers; multiple X-ray data sets (from deletion mutants or partial complexes) can simultaneously be fitted.

Rigid body modelling became a powerful analysis tool in the SAS studies of macromolecular complexes. Often, this analysis complements *ab initio* reconstruction to further refine or validate the docking results [20**,22**,23*,29*]. A simultaneous fitting of X-ray and contrast variation neutron data was successfully employed in the study of the chicken skeletal muscle troponin complex [41] and of the complex between the bacterial histidine kinase and its cognate response regulator [16**]. Addition of missing portions was performed e.g. for the cytosolic factor of NADPH oxidase [42*], for dystrophin myotonia kinase [43] and for polypyrimidine tract binding protein [17*].

Although the rigid body modelling operates with high resolution structures, the SAS-based models are still low resolution ones and, similar to *ab initio* analysis, one should be aware of possible multiple solutions [15*]. Constraints from other methods are important but the models should also be independently validated (e.g. by analysing the feasibility of

the intersubunit interfaces). An example of an *a posteriori* validation is given by the study of sensor histidine kinase PrrB from *Mycobacterium tuberculosis* [44*], where a rigid body model was constructed for the dimeric enzyme consisting of two domains with known structure (dimerisation domain and ATP-binding domain) and a HAMP linker with unknown structure. The model is overlapped in Figure 2 with the crystal structure of the cytoplasmic portion of sensor histidine kinase from *Thermotoga maritima* containing the dimerisation and ATP-binding domains [45]. These two independently determined structures (both papers were submitted at about the same time) show a remarkably similar position of the ATP-binding domain.

Structure validation and mixtures

Validation of the experimental and predicted high resolution structures and selection between alternative models are perhaps the most straightforward applications of SAS. The solution conformations of eucaryotic release factor eRF1 [46] and elongation factor eEF3 [47*] differed significantly from their crystal structures. Based on SAXS, NMR and EM, quaternary structure of tumour suppressor p53 was established [48**] distinctly different from its cryo-EM model [49]. SAXS has recently even helped in distinguishing between alternative NMR models [50*]. Another typical task is determination of the oligomeric state in solution (which was recently used *e.g.* for screening of detergent-solubilized integral membrane proteins from *Thermotoga maritima* [51]). Given the crystal of the monomer, biologically relevant oligomer in solution may be identified by considering possible crystallographic oligomers or by rigid body modelling [26*,52,53].

For a very practically important case of oligomeric mixtures, the measured intensity is a sum of the contributions of different oligomeric components weighted by their volume fractions. The latter are directly computed from the SAS data provided the structures of the components are known. This approach is widely used to study oligomeric equilibria but also complex formation, structural transitions and assembly [54*,55-58]. Processes like amyloid fibrillation can be quantitatively described and the precursors detected may provide important information for drug design [59**].

SAS is one of the few structural methods applicable to flexible macromolecules, and attempts were made to use it for unfolded systems [60] and multidomain proteins with flexible linkers [61]. A joint use of SAXS and NMR provides also information about structural dynamics [62]. The major problem with flexible systems is that the SAS data include not only orientational but also conformational average. Recently a general approach 'ensemble optimization method' (EOM) was proposed allowing for coexistence of multiple conformations [63*]. The EOM may find broad applications for flexible proteins and complexes, intrinsically disordered proteins, and in the time resolved SAXS studies of protein folding, where the high brilliance synchrotrons and rapid mixing devices allowed one to push the time resolution into μsec range [64*,65*].

Conclusions

Biological solution SAS is a rapidly growing field with many novel approaches and exciting applications, especially in the studies of macromolecular complexes. One should however always keep in mind that the SAS models are low resolution ones. Typically, *ab initio* bead models utilize the scattering data up to about 1.5-2 nm resolution (momentum transfer to about $s=0.3-0.4 \text{ nm}^{-1}$), but also for rigid body modelling this range is most sensitive to changes in the quaternary structure. Incorporation of higher resolution data (to 0.5-1 resolution) is useful for domain structure analysis and the methods employing dummy residues, which also provide somewhat more detailed - but still low resolution - models. Wide-angle X-ray scattering patterns up to $s=2-3 \text{ nm}^{-1}$ (0.2-0.3 nm resolution) are sensitive to the internal structure and can be utilized to probe e.g. ligand-induced conformational changes in proteins with known high resolution structure [66].

In general, if a model agrees with the SAS data, this does not yet prove its uniqueness at high resolution (albeit at low resolution, with accurate use of the methods, unique reconstructions are achieved). However, any model failing to fit the data definitely does not represent the macromolecular conformation in solution. It is thus always useful to confront SAS-based models with other pieces of information and *vice versa*, SAS is an ideal validation method for high resolution models. The joint use of SAS with other high and low resolution experimental techniques and bioinformatic tools is a powerful instrument for hierarchical systems at different level of structural organization. Further, SAS stays the method of choice for the study of flexible systems, mixtures and structural characterization of kinetic processes.

Many SAS analysis programs described above are publicly available from the EMBL Web page (<http://www.embl-hamburg.de/ExternalInfo/Research/Sax/>), and some of them can now be run on-line. Other useful public tools include PISA server for the analysis of intersubunit interfaces (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html), SITUS package for the docking of high resolution structures into low resolution maps (<http://situs.biomachina.org/>), and ElNemo server to generate protein models using normal mode analysis (<http://igs-server.cnrs-mrs.fr/elnemo/index.html>).

Acknowledgment

The authors thank P.Tucker for helpful comments and acknowledge support from the EU Framework 6 Programme (Design Study SAXIER, RIDS 011934).

Figure captions

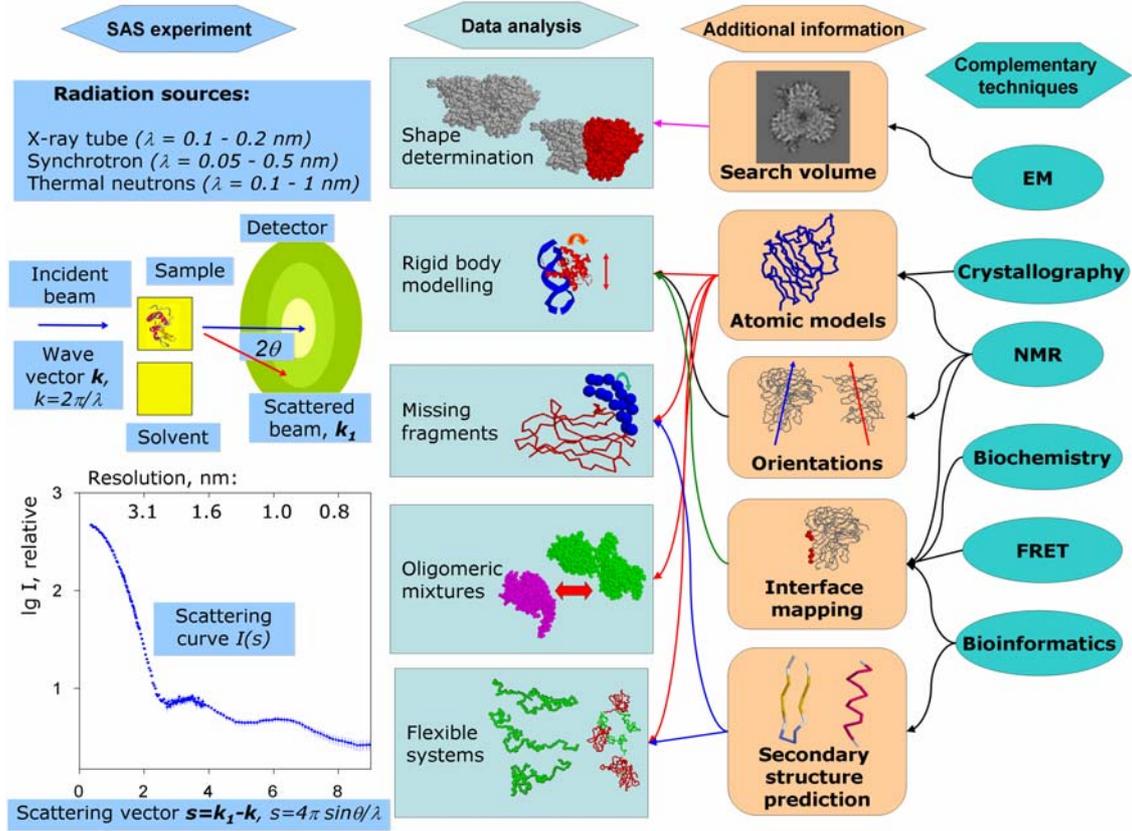


Fig. 1. A scheme of a SAS experiment, structural tasks addressed and the joint use with other methods. The nominal resolution of the scattering data is indicated as $d=2\pi/s$.

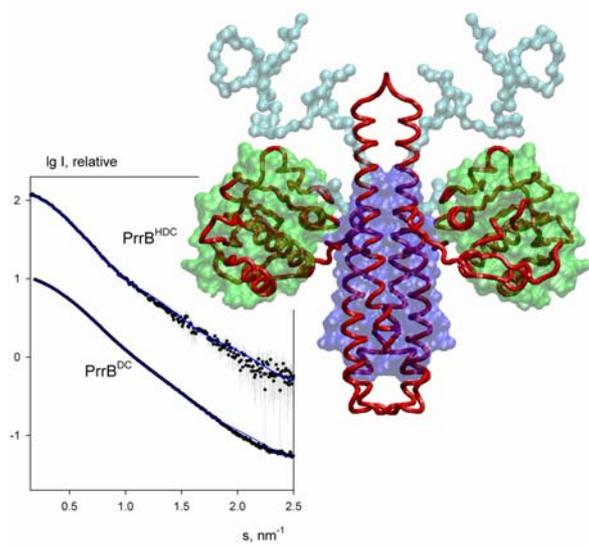


Fig. 2. Rigid body model of sensor histidine kinase PrrB from *Mycobacterium tuberculosis* [44*] displayed as semitransparent beads (blue: dimerisation domain, green: ATP-binding domain, cyan: tentative configuration of the HAMP linker). The three-domain construct fits the upper data set (PrrB^{HDC}) in the plot while the dimerization+ATP-binding domain portion fits the lower curve labeled PrrB^{DC} (dots with error bars: experiment, solid lines: calculated curve). The independent crystallographic model of *Thermotoga maritima* [45] is displayed as red C_{α} -trace.

References and recommended reading

1. Levitt M: **Growth of novel protein structural data.** *Proc Natl Acad Sci U S A* 2007, **104**:3183-3188.
2. Aloy P, Russell RB: **Structural systems biology: modelling protein interactions.** *Nat Rev Mol Cell Biol* 2006, **7**:188-197.
3. Feigin LA, Svergun DI: *Structure analysis by small-angle x-ray and neutron scattering.* New York: Plenum Press; 1987.
4. Koch MH, Vachette P, Svergun DI: **Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution.** *Q Rev Biophys* 2003, **36**:147-227.
5. Svergun DI, Volkov VV, Kozin MB, Stuhrmann HB: **New developments in direct shape determination from small-angle scattering 2. Uniqueness.** *Acta Crystallogr.* 1996, **A52**:419-426.
6. Chacon P, Moran F, Diaz JF, Pantos E, Andreu JM: **Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm.** *Biophys J* 1998, **74**:2760-2775.
7. Svergun DI: **Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.** *Biophys J* 1999, **76**:2879-2886.
8. Walther D, Cohen FE, Doniach S: **Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules.** *J. Appl. Crystallogr.* 2000, **33**:350-363.
9. Heller WT, Abusamhadneh E, Finley N, Rosevear PR, Trehwella J: **The solution structure of a cardiac troponin C-troponin I-troponin T complex shows a somewhat compact troponin C interacting with an extended troponin I-troponin T component.** *Biochemistry* 2002, **41**:15654-15663.
10. Takahashi Y, Nishikawa Y, Fujisawa T: **Evaluation of three algorithms for ab initio determination of three-dimensional shape from one-dimensional solution scattering profiles.** *J. Appl. Crystallogr.* 2003, **36**:549-552.
11. Zipper P, Durchschlag H: **Modeling of protein solution structures.** *J. Appl. Crystallogr.* 2003, **36**:509-514.
12. Volkov VV, Svergun DI: **Uniqueness of ab initio shape determination in small angle scattering.** *J. Appl. Crystallogr.* 2003, **36**:860-864.
- **13. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI: **ATSAS 2.1, a program package for small-angle scattering data analysis.** *J. Appl. Crystallogr.* 2006, **39**:277-286.

A comprehensive set of tools is described for SAS data analysis from biological macromolecules implemented in the program package ATSAS 2.1. These programs cover major processing and interpretation steps from primary data reduction to three-dimensional model building. The presented methods allow the user to perform data manipulations (averaging, buffer subtraction), to compute the overall particle parameters (R_g , D_{max} , MM , V), to analyze the composition of mixtures, to restore the overall shape *ab initio*. Recent improvements are described in the methods to compute the SAXS/SANS patterns from high resolution structures. Based on the known structures of subunits interactive and automated rigid body modelling methods are presented to analyze the quaternary structure of homo- and heterodimers, symmetric oligomers,

multidomain proteins and multisubunit complexes. In latter case a global simulated annealing based search is applied taking into account possible complementary information from other methods.

14. Svergun DI, Petoukhov MV, Koch MHJ: **Determination of domain structure of proteins from X-ray solution scattering.** *Biophys J* 2001, **80**:2946-2953.
- *15. Petoukhov MV, Svergun DI: **Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution.** *Eur Biophys J* 2006, **35**:567-576.

New *ab initio* and rigid body modelling methods are described for simultaneous analysis of multiple X-ray and neutron scattering patterns from complexes in solution utilizing simulated annealing. They are applicable for studying complexes of proteins, nucleic acids and other biological macromolecules. The data which can be taken into account include contrast variation series from selectively deuterated complexes and the scattering patterns from partial constructs. The search volume for *ab initio* analysis (usually, a sphere of diameter D_{max}) is filled with densely packed beads, which may belong to one of the components in the complex or to the solvent. The use of multiphase approach allows one not only to restore the overall shape but also to gain the information on internal structure of multicomponent particles. A possibility to account for the symmetry of the complex reduces the ambiguity of *ab initio* reconstruction. Yet more detailed modelling of macromolecular complexes is done using rigid body modelling. The positions and orientations of rigid subunits yielding interconnected models without steric clashes and displaying a pre-defined symmetry are optimized to simultaneously fit multiple solution scattering curves. The reliability of the models is enhanced by using additional information (e.g. distance restrains between specific residues or nucleotides and orientational restrains from NMR residual dipolar couplings). Still, examples are presented where multiple solutions are possible and ways of reducing the ambiguity are discussed.

- **16. Whitten AE, Jacques DA, Hammouda B, Hanley T, King GF, Guss JM, Trewhella J, Langley DB: **The Structure of the KinA-Sda Complex Suggests an Allosteric Mechanism of Histidine Kinase Inhibition.** *J Mol Biol* 2007, **368**:407-420.

A complex of 54 kDa histidine kinase KinA and 13 kDa DNA-damage checkpoint inhibitor Sda is studied by small-angle X-ray scattering and neutron contrast variation. The experimental MM and R_g values from SAXS indicate homodimeric architectures for individual species and a 2:2 stoichiometry for their complex. A significant decrease in D_{max} of the complex compared to that of free KinA dimer points to KinA₂ compaction upon Sda binding. Further insight into complex organization is gained by a combination of SAXS with SANS. Using neutron scattering on the complex with perdeuterated Sda, the contrast of this protein was highlighted to compensate for its small size. A SAXS pattern together with seven SANS data sets in different H₂O/D₂O mixtures were used in rigid body modelling of the dimeric complex with P2 symmetry constraint. Fourteen independent runs yielded similar models of the complex with Sda molecules and CA domains at the opposite ends of the DHp stalk of KinA. As an additional check, *ab initio* multiphase bead modelling was also performed providing similar results. The model built

from solution scattering data suggests an allosteric mode of inhibition of autokinase activity by Sda binding.

- *17. Petoukhov MV, Monie TP, Allain FH, Matthews S, Curry S, Svergun DI: **Conformation of polypyrimidine tract binding protein in solution.** *Structure* 2006, **14**:1021-1027.

An example of *ab initio* and rigid body modelling of multidomain protein against multiple scattering data sets from its deletion mutants. Simulated annealing-based optimization is applied to analyse the polypyrimidine tract binding protein (PTB) comprised by four domains (each about 10 kDa, structures solved by NMR) interconnected by linkers. A total of seven SAXS curves from the full-length protein and from all possible sequential combinations of domains were fitted simultaneously to analyze the PTB structure. The multiphase *ab initio* approach yields an elongated shape of PTB with essentially linear distribution of domains. This result is further supported and refined by molecular modelling which provided not only the mutual arrangement of domains but also possible conformations of the linkers. The proposed model is compatible with the possible role of PTB in bridging RNA sequence motifs.

- *18. Lima LM, Cordeiro Y, Tinoco LW, Marques AF, Oliveira CL, Sampath S, Kodali R, Choi G, Foguel D, Torriani I, et al.: **Structural insights into the interaction between prion protein and nucleic acid.** *Biochemistry* 2006, **45**:9180-9187.

A combined SAXS/NMR study of the high-affinity complex between a hamster prion protein (MM=16 kDa) and an 18 bp DNA sequence. *Ab initio* SAXS modelling of the free protein and its truncated form indicated an elongated particle with two distinct domains; a globular domain resembles the known murine PrP high-resolution structure and a disordered domain. Analysis of the mixtures of the truncated prion protein with DNA confirms the formation of a tight complex with 1:1 stoichiometry and suggests that DNA binding involves the globular domain. This finding is further corroborated by the measurements of DNA binding affinity for the PrP lacking part of the disordered domain. The NMR chemical shifts between the free and DNA-bound prion suggest the same overall fold and secondary structure arrangement of the protein in the both states and localize the residues involved in DNA binding: a cluster in the globular domain of the PrP and another cluster in the disordered portion. The nucleic acid could therefore be a likely chaperone candidate converting the normal, cellular form into the disease-causing isoform.

- **19. Nagar B, Hantschel O, Seeliger M, Davies JM, Weis WI, Superti-Furga G, Kuriyan J: **Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase.** *Mol Cell* 2006, **21**:787-798.

This paper reports a crystal structure of the autoinhibited state of 54 kDa c-Abl tyrosine kinase fragment containing an N-terminal cap segment, not visualized in previous structures, SH3-SH2 substructure and the kinase domain. The c-Abl fragment and its mutated form, in which the N-terminal cap and two other key contacts in the autoinhibited state are deleted, were analysed by SAXS. The experimental R_g and D_{max} values of non-mutated c-Abl are compatible with those calculated from the crystal structure. The SAXS measurements on the mutated form reveal an increase of 20% in R_g

and 40% in D_{max} pointing to a significantly more extended conformation. Further analysis of structural dissimilarities between the autoinhibited and mutated forms was performed by DR modelling. The averaged *ab initio* model of the inactive form has a globular shape matching well the crystal structure. The reconstruction of the mutated Abl produced a more elongated shape, showing an extended configuration of the SH3, SH2 and kinase domains. This alternative conformation of mutated c-Abl is likely to prolong the active state of the kinase and suggests a critical role of N-terminal cap segment in autoinhibition.

20. Baldock C, Siegler V, Bax DV, Cain SA, Mellody KT, Marson A, Haston JL, Berry R, Wang MC, Grossmann JG, et al.: **Nanostructure of fibrillin-1 reveals compact conformation of EGF arrays and mechanism for extensibility. *Proc Natl Acad Sci U S A* 2006, **103**:11922-11927.

Solution structure of human fibrillin-1, a multidomain extracellular matrix protein with MM=330 kDa, is analyzed using SAXS, light scattering and EM. Fibrillin-1 was believed to have a uniform rod shape, however *ab initio* structure determination of nine overlapping fragments, ranging in size from 5 to 13 contiguous domains revealed a nonlinear hook-shaped conformation of calcium-binding EGF arrays in solution. Particularly, a very compact, globular region of the fibrillin-1 containing the integrin and heparan sulfate-binding sites was found by SAXS and further confirmed using EM and single-particle image analysis. Based on homology models of the domains a combined rigid body and *ab initio* modelling approach was also employed to get further structural insights into domain arrangement. One-, two- and three-domain subunits were positioned and the missing linkers between the fragments were built by fitting the scattering data. A putative structure of fibrillin-1 was generated by aligning the SAXS structures from TB1 to the C terminus. The methodology used can be applied to other EGF-containing extracellular matrix and membrane proteins.

*21. Buey RM, Monterroso B, Menendez M, Diakun G, Chacon P, Hermoso JA, Diaz JF: **Insights into molecular plasticity of choline binding proteins (pneumococcal surface proteins) by SAXS.** *J Mol Biol* 2007, **365**:411-424.

The experimental SAXS data from monomeric phage encoded Cpl-1 lysozyme (40 KDa) cannot be fitted by its model in the crystal. Neither of the two dimeric assemblies according to the crystal packing fits the data of Cpl-1 with bound choline, which is known to induce dimerization. *Ab initio* low resolution models of the monomeric and dimeric states were built from SAXS data. The three domains of Cpl-1 identified in the original crystal structure were docked manually into the monomer shape accounting for the predicted hydrophobic contacts. This monomer was docked into the dimeric envelope yielding the choline-binding module at the interface between the monomers. Both monomeric and dimeric models of Cpl-I in solution provide a much better fit to the SAXS data. The scattering from a shorter C-Cpl-1 construct containing the choline-binding module and lacking the catalytic module also indicates dimerization upon choline binding, and the *ab initio* shapes for monomeric and dimeric states of C-Cpl-1 agree well with the appropriate portions of the SAXS-based models of the full-length protein.

- **22. Pascal JM, Tsodikov OV, Hura GL, Song W, Cotner EA, Classen S, Tomkinson AE, Tainer JA, Ellenberger T: **A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA.** *Mol Cell* 2006, **24**:279-291.

An example of combined use of crystallography and SAXS for macromolecular complexes. The crystallographically determined structures of *Sulfolobus solfataricus* DNA ligase and heterotrimeric proliferating cell nuclear antigen (PCNA) were used to model the structure of their complex against SAXS data. Using SAXS it was confirmed that free ligase in solution has an elongated, anisotropic shape, remarkably similar to the crystallographic model of extended, open conformation whereas a more compact DNA-bound conformation of ligase is not compatible with SAXS. A planar ring heterotrimeric assembly of PCNA in the crystal in the absence of DNA is in a good agreement with the SAXS data confirming the stability of the trimer. The interaction of ligase and PCNA analyzed by SAXS showed formation of a stable 1:1 complex. The *ab initio* shape determination produced stable solutions clearly suggesting the architecture of the complex: the discoidal shape (corresponding to PCNA) is capped on one edge by a projection extending from the circumference in the plane of the ring (corresponding to ligase). The docking of the crystallographic models into the averaged *ab initio* shape and independent rigid body modelling against SAXS data produced similar results showing a single molecule of ligase bound to the PCNA trimer. The SAXS models of the complex suggest that PCNA can serve as a docking station that keeps multiple enzymatic activities in close proximity to DNA, engaging the substrate only at the appropriate step in a reaction pathway.

- *23. Smolle M, Prior AE, Brown AE, Cooper A, Byron O, Lindsay JG: **A new level of architectural complexity in the human pyruvate dehydrogenase complex.** *J Biol Chem* 2006, **281**:19772-19780.

Human pyruvate dehydrogenase (PDC) is a complex consisting of multiple copies of pyruvate decarboxylase (E1), dihydrolipoamide acetyltransferase (E2), and dihydrolipoamide dehydrogenase (E3). In many eukaryotic complexes an accessory E3-binding protein (E3BP) is also present, which mediates stable E3 integration to the E2 icosahedral faces. A 144 kDa subcomplex of E3 with E3BP construct was investigated by SAXS, native PAGE, analytical ultracentrifugation and isothermal titration calorimetry. The overall parameters from SAXS show that that two E3BP constructs interact with a single E3 homodimer to form a 2:1 complex in agreement with the results of other methods. The *ab initio* shape restoration revealed an elongated, asymmetric structure of the E3 - E3BP subcomplex. The structure of the subcomplex was independently analyzed by rigid body modelling to determined the locations of four E3BP domains and of E3 yielding the best agreement to the SAXS data. The resulting rigid body model agrees well with the *ab initio* SAXS envelope. In contrast with crystal structures of human E3 complexed with its cognate subunit binding domain, the observed 2:1 stoichiometry suggests a novel subunit organization implying the existence of a network of E3 “cross-bridges” linking pairs of E3BPs across the surface of the E2 core assembly.

- *24. Evans K, Grossmann JG, Fordham-Skelton AP, Papiz MZ: **Small-angle X-ray scattering reveals the solution structure of a bacteriophytochrome in the catalytically active Pr state.** *J Mol Biol* 2006, **364**:655-666.

SAXS data suggest that the catalytic Pr state of light-sensing bacteriophytochrome controlling gene expression is dimeric in solution (167 kDa), and a Y-shape reconstructed *ab initio* correlates well with earlier EM images of pea phytochrome. Homologues with known atomic structure were found for the chromophore binding (CBD) and dimerization+catalytic domains but not for the phytochrome N terminal domain (PHY). The base of the Y-shaped SAXS density resembles closely the shape of dimeric CHK, the peripheral part of the Y arms agrees well with the shape of CBD and the rest density is able to accommodate the missing PHY domain. The proposed model helps in understanding the autophosphorylation mechanism of a histidine in the dimerization domain of bacteriophytochrome.

- *25. De Carlo S, Chen B, Hoover TR, Kondrashkina E, Nogales E, Nixon BT: **The structural basis for regulated assembly and function of the transcriptional activator NtrC.** *Genes Dev* 2006, **20**:1485-1495.

A structural study of the activated, full-length nitrogen-regulatory protein C (NtrC; MM of a monomer 51 kDa). Analysis of the SAXS data was facilitated by the known hexameric assembly of the object revealed by EM, so that the *ab initio* low resolution shape analysis was performed with P6 symmetry constraint. The averaged displays a flat star-shaped planar structure with a pore in the center. Molecular docking of available high resolution structures of the domains to SAXS model started from the location of a hexameric ring of the AAA+ ATPase domains. Then the six receiver domains were docked to the peripheral knobs to satisfy NMR chemical shift and biochemical Fe-BABE cleavage data pointing to proximity of helix 4 of a receiver domain to helix 1 of an ATPase domain. Finally, a full model of the activated, full-length NtrC was produced by adding the three copies of a dimeric DNA-binding domain using three-fold symmetry to fill the unoccupied cap-like portion surrounding the central pore. The DNA-binding helix was kept at the outer surface of the model. An independent 3D EM reconstruction yielded a very similar model. Based on the revealed architecture of NtrC a novel mechanism for regulation of AAA+ ATPase assembly via the juxtaposition of the receiver domains and ATPase ring was proposed.

- *26. Pioletti M, Findeisen F, Hura GL, Minor DL, Jr.: **Three-dimensional structure of the KChIP1-Kv4.3 T1 complex reveals a cross-shaped octamer.** *Nat Struct Mol Biol* 2006, **13**:987-995.

This paper presents and validates the crystal structure of KChIP1-Kv4.3 T1 complex influencing potassium channel trafficking and gating. Crystallographic model solved at 0.34 nm resolution shows a cross-shaped octamer (MM=142 Kda) having a T1 tetramer at the center and the individual KChIPs extending radially. The SAXS data were used to probe the complex structure in solution and to test whether the flat, cruciform octamer is the preferred arrangement or a consequence of crystallization. The overall structural parameters of the complex assessed by SAXS agree well with those computed from the crystal structure and a low resolution DR model reconstructed *ab initio* using P4 symmetry strongly resembles the overall shape of the crystallographic model. In contrast,

an alternative four-fold symmetric arrangement of the KChIP-Kv4 T1 complex with a square shape proposed from low-resolution negative-stain EM images shows poor overlap with the DR model and yields substantially smaller overall sizes and worse fit to the SAXS experimental data. Thus, solution scattering allowed one to discriminate between the concurrent models of the complex and helped in understanding how does KChIPs modulate Kv4 function.

- *27. Lipfert J, Das R, Chu VB, Kudaravalli M, Boyd N, Herschlag D, Doniach S: **Structural transitions and thermodynamics of a glycine-dependent riboswitch from *Vibrio cholerae***. *J Mol Biol* 2007, **365**:1393-1406.

Conformational changes of 74 kDa tandem aptamer riboswitch (VCI-II) as a function of Mg^{2+} and glycine concentration are investigated by SAXS complemented by hydroxyl radical footprinting. SAXS profiles at taken different conditions demonstrate significant structural rearrangement upon glycine binding or Mg^{2+} addition. Singular value decomposition (SVD) analysis suggest at least three components identified as unfolded structure in the absence of Mg^{2+} , the glycine-bound compact conformation of tandem aptamer and a partially folded intermediate without glycine in the presence of Mg^{2+} . The energetic coupling between magnesium-induced folding and glycine binding are described by a simple three-state thermodynamic model which allowed to design VCI-II conformational landscape. Further structural insights were obtained by *ab initio* modelling of the three conformational states which showed some marginal similarity between the glycine-bound conformation and the intermediate whereas the unfolded state model is dissimilar from both other states.

- *28. Johs A, Hammel M, Waldner I, May RP, Laggner P, Prassl R: **Modular structure of solubilized human apolipoprotein B-100. Low resolution model revealed by small angle neutron scattering**. *J Biol Chem* 2006, **281**:19732-19739.

A SANS study of solubilized human apolipoprotein B-100 where the delipidated protein (550 kDa) was surrounded by a non-ionic detergent. The experiment with apoB-100 was performed in a solvent at the matching point of the detergent (18% D_2O) so that the net scattering curve contained mostly by the scattering from the pure protein. The SANS data yielded D_{max} of 60 nm and the distance distribution function computed as a Fourier transformation of the intensity displayed several distinct maxima suggesting that apoB-100 consists of independent modules. A typical *ab initio* model reconstructed from SANS data has an elongated bowed shape with a central cavity and also displays several distinct modules. To further characterize the lipoprotein, the sequence-based secondary structure prediction of the apoB-100 was mapped onto the low resolution model. The hypothetical organization of an LDL was suggested by accommodating the apoB-100 shape onto a sphere resulting in a ring-like conformation with the termini close to each other.

- *29. Gherardi E, Sandin S, Petoukhov MV, Finch J, Youles ME, Ofverstedt LG, Miguel RN, Blundell TL, Vande Woude GF, Skoglund U, et al.: **Structural basis of hepatocyte growth factor/scatter factor and MET signalling**. *Proc Natl Acad Sci U S A* 2006, **103**:4046-4051.

An extensive study of 90 KDa hepatocyte growth factor/scatter factor (HGF/SF) and its interaction with the receptor tyrosine kinase MET (118 KDa) using SAXS and cryo-EM.

Both techniques reveal major structural transition by conversion of single-chain HGF/SF into the active two-chain form. The *ab initio* shape of 1:1 complex between two-chain HGF/SF and MET reconstructed from SAXS data is compatible with EM maps. Rigid body models of 1:1 complex and 2:2 complex formed with the truncated form of MET were built by simulated annealing based on the atomic models of individual domains of HGF/SF and MET and taking into account the information on binding sites predicted by mutagenesis experiments. Both models show the binding of HGF/SF to the propeller domain of MET. The dimerization interface of 2:2 complex resembles that seen in the crystal structure of the NK1 fragment of HGF/SF.

30. Svergun DI, Barberato C, Koch MHJ: **CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates.** *J. Appl. Crystallogr.* 1995, **28**:768-773.
31. Svergun DI, Richard S, Koch MHJ, Sayers Z, Kuprin S, Zaccai G: **Protein hydration in solution: experimental observation by x-ray and neutron scattering.** *Proc Natl Acad Sci U S A* 1998, **95**:2267-2272.
32. Svergun DI: **Restoring three-dimensional structure of biopolymers from solution scattering.** *J. Appl. Crystallogr.* 1997, **30**:792-797.
33. Konarev PV, Petoukhov MV, Svergun DI: **MASSHA - a graphic system for rigid body modelling of macromolecular complexes against solution scattering data.** *J. Appl. Crystallogr.* 2001, **34**:527-532.
34. Petoukhov MV, Svergun DI: **Global rigid body modelling of macromolecular complexes against small-angle scattering data.** *Biophys J* 2005, **89**:1237-1250.
35. Mattinen ML, Paakkonen K, Ikonen T, Craven J, Drakenberg T, Serimaa R, Waltho J, Annala A: **Quaternary structure built from subunits combining NMR and small-angle x-ray scattering data.** *Biophys J* 2002, **83**:1177-1183.
36. Grishaev A, Wu J, Trewhella J, Bax A: **Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data.** *J Am Chem Soc* 2005, **127**:16621-16628.
- *37. Fernando AN, Furtado PB, Clark SJ, Gilbert HE, Day AJ, Sim RB, Perkins SJ: **Associative and structural properties of the region of complement factor H encompassing the Tyr402His disease-related polymorphism and its interactions with heparin.** *J Mol Biol* 2007, **368**:564-581.

A combination of analytical ultracentrifugation and SAXS for the study of SCR-6/8 construct of factor H (MM=21 kDa) and its complex with a heparin decasaccharide. Both methods indicate a liner arrangement of SCR-6,-7 and -8 domains in the presence of heparin. The sedimentation coefficient and the values of R_g and D_{max} in the unbound state suggest still elongated but a bent conformation of free SCR-6/8. To further prove this finding, the scattering profile of SCR-6/8 at the lowest concentration corresponding predominantly to the monomers was used for a constrained rigid body modelling. A total of 2000 models were randomly generated based on the crystallographic and NMR homology models of the three domains and a library of linkers and terminal peptides produced by molecular dynamics. By screening of these models against scattering data a single best-fit family of structures was identified demonstrating a bent domain arrangement. The heparin-binding residues in the family were exposed on the outside

curvature, which may explain the formation of a more linear structure in the complex caused by heparin binding.

38. Augustus AM, Reardon PN, Heller WT, Spicer LD: **Structural basis for the differential regulation of DNA by the methionine repressor MetJ.** *J Biol Chem* 2006, **281**:34269-34276.
39. Nollmann M, Stark WM, Byron O: **A global multi-technique approach to study low-resolution solution structures.** *J. Appl. Cryst.* 2005, **38**:874-887.
40. Petoukhov MV, Eady NA, Brown KA, Svergun DI: **Addition of missing loops and domains to protein models by x-ray solution scattering.** *Biophys J* 2002, **83**:3113-3125.
41. King WA, Stone DB, Timmins PA, Narayanan T, von Brasch AA, Mendelson RA, Curmi PM: **Solution structure of the chicken skeletal muscle troponin complex via small-angle neutron and X-ray scattering.** *J Mol Biol* 2005, **345**(4):797-815.
- *42. Durand D, Cannella D, Dubosclard V, Pebay-Peyroula E, Vachette P, Fieschi F: **Small-angle X-ray scattering reveals an extended organization for the autoinhibitory resting state of the p47(phox) modular protein.** *Biochemistry* 2006, **45**:7185-7193.

A SAXS study of the cytosolic factor p47^{phox} in autoinhibitory resting state, a modular protein of 46 KDa promoting the assembly of the NADPH oxidase complex. The experimental MM value monomeric state of the protein in solution whereas D_{max} and R_g values were significantly larger than those expected for a globular protein with this MM. *Ab initio* modelling approaches generated elongated models with one thinner end presumably corresponding to a flexible terminal portion. Further structural analysis was performed by rigid body docking of crystallographic and NMR models of the PX domain and SH3 tandem into the typical *ab initio* shape followed by the addition of missing loops composed of DRs. The final solution scattering based model of p47^{phox} unambiguously demonstrates extended conformation, which is unusual for autoinhibitory resting state.

43. Garcia P, Ucurum Z, Bucher R, Svergun DI, Huber T, Lustig A, Konarev PV, Marino M, Mayans O: **Molecular insights into the self-assembly mechanism of dystrophin myotonia kinase.** *Faseb J* 2006, **20**:1142-1151.
- *44. Nowak E, Panjikar S, Morth JP, Jordanova R, Svergun DI, Tucker PA: **Structural and functional aspects of the sensor histidine kinase PrrB from Mycobacterium tuberculosis.** *Structure* 2006, **14**:275-285.

The paper reports the high-resolution crystal structure of the ATP-binding domain (108 residues) of the sensor histidine kinase PrrB from *Mycobacterium tuberculosis* and the solution structures of two- and three domain constructs containing additionally a 67 residues phosphorylation domain and also a HAMP linker (69 residues). Both constructs were found to be dimeric in solution, with the dimer formation mediated *via* the phosphorylation domain. The scattering patterns from the two and three domain constructs were simultaneously fitted by a model assuming a 2-fold symmetry axis. The dimeric phosphorylation domain was fixed as in the available crystallographic structure of a related CheA protein, the ATP-binding domain was moving as a rigid body attached to the appropriate terminus of the phosphorylation domain and the HAMP linker was

represented by a dummy residues loop. The resulting model showed a significantly closer contact between the ATP-binding and dimerization domains than that observed for CheA, and the (potentially flexible) HAMP domain occupying the part of the molecule above the catalytic domain.

45. Marina A, Waldburger CD, Hendrickson WA: **Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein.** *Embo J* 2005, **24**:4247-4259.

46. Vestergaard B, Sanyal S, Roessle M, Mora L, Buckingham RH, Kastrop JS, Gajhede M, Svergun DI, Ehrenberg M: **The SAXS solution structure of RF1 differs from its crystal structure and is similar to its ribosome bound cryo-EM structure.** *Mol Cell* 2005, **20**:929-938.

*47. Andersen CB, Becker T, Blau M, Anand M, Halic M, Balar B, Mielke T, Boesen T, Pedersen JS, Spahn CM, et al.: **Structure of eEF3 and the mechanism of transfer RNA release from the E-site.** *Nature* 2006, **443**:663-668.

The paper reports a high-resolution crystal structure of the 110 KDa elongation factor eEF3 from *Saccharomyces cerevisiae* which serves an essential function in the translation cycle of fungi. The pattern computed from the rather compact crystal structure does not fit the experimental SAXS data collected at pH 7.2. A modelling of the ATP state of eEF3 in solution was performed based on the arrangement of ATP-binding cassette domains ABC1 and ABC2 in the ATP-induced dimer of MJ0796, that resulted in elongated molecule with the chromodomain insert in ABC2 located at the tip. Cryo-EM structure of the ATP-bound form of eEF3 in complex with the post-translocational-state 80S ribosome from yeast also obtained in this study suggests that the chromodomain plays a role in stabilizing the ribosomal L1 stalk in an open conformation. SAXS data collected at the crystallization pH of 5.2 indicate a more compact conformation of eEF3 pointing to pH-induced conformational change within eEF3 in solution.

48. Tidow H, Melero R, Mylonas E, Freund SMV, Grossmann JG, Carazo JM, Svergun DI, Valle M, Fersht AR: **Quaternary structures of tumour suppressor p53 and a specific p53-DNA complex. *Proc Natl Acad Sci U S A* 2007, **in press**.

The human tumour suppressor p53 is a homotetrameric transcription factor (four times 393 residues) playing a central role in the cell cycle. It contains a folded core and tetramerization domains, linked and flanked by intrinsically disordered segments. *Ab initio* and rigid body SAXS modelling accounting for NMR-derived interfaces revealed an extended cross-shaped structure with tetrameric contacts and a pair of loosely coupled core domain dimers at the ends. In contrast, the calculated scattering from the recent rather compact cryo-EM structure of murine p53 [49] showing dissociated tetramerization domains did not fit the experimental SAXS data. The structure of the complex of p53 with 24 bp DNA independently determined by SAXS and negative stain EM displays a compact complex with the core domains closing around DNA. Interestingly, negatively stained EM analysis of the conformationally mobile, unbound p53 selected a minor compact conformation (less than 20% of the adsorbed particles). The study underlines the significance of the synergistic use of different techniques for the structural characterization of a rapidly growing number of proteins with inherent disorder.

49. Okorokov AL, Sherman MB, Plisson C, Grinkevich V, Sigmundsson K, Selivanova G, Milner J, Orlova EV: **The structure of p53 tumour suppressor protein reveals the basis for its functional plasticity.** *Embo J* 2006, **25**:5191-5200.

*50. Nicastro G, Habeck M, Masino L, Svergun DI, Pastore A: **Structure validation of the Josephin domain of ataxin-3: Conclusive evidence for an open conformation.** *J Biomol NMR* 2006, **36**:267-277.

Two high resolution models were proposed for the Josephin domain of ataxin-3, a 182 residues protein involved in the ubiquitin/proteasome pathway, using NMR (PDB codes 1yzb and 2aga). They display distinctly different overall shapes with an open cleft in the former model and a closed cleft in the latter one. SAXS data was used together with include Bayesian methods to validate the models and to demonstrate that only one of them, with the open cleft, is compatible with the experimental information.

51. Columbus L, Lipfert J, Klock H, Millett I, Doniach S, Lesley SA: **Expression, purification, and characterization of *Thermotoga maritima* membrane proteins for structure determination.** *Protein Sci* 2006, **15**:961-975.

52. Mendillo ML, Putnam CD, Kolodner RD: **Escherichia coli muts tetramerization domain structure reveals that stable dimers but not tetramers are essential for DNA mismatch repair in vivo.** *J Biol Chem* 2007.

53. Rosenberg OS, Deindl S, Sung RJ, Nairn AC, Kuriyan J: **Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme.** *Cell* 2005, **123**:849-860.

*54. Graziano V, McGrath WJ, Yang L, Mangel WF: **SARS CoV main proteinase: The monomer-dimer equilibrium dissociation constant.** *Biochemistry* 2006, **45**:14632-14641.

Oligomeric state of the SARS coronavirus main proteinase, a 33 KDa protein processing viral polyproteins was studied by SAXS, chemical cross-linking and enzyme kinetics. The enzyme requires homodimeric state for its activity and characterization of the dimer-monomer equilibrium is important. SAXS patterns of proteinase at different protein concentrations were compared with the theoretical scattering curves computed from the known structures of the monomer and dimer. The dimeric species dominated in solutions above the concentration of 5.5 mg/ml, but equilibrium mixtures of monomers and dimers were observed at lower concentrations. The volume fractions of the monomeric and dimeric components determined from SAXS yielded the dimer-monomer dissociation constant K_D of 5.8 μ M. This value agreed well with The K_D estimates from chemical cross-linking and enzyme kinetics experiments. No tendency to form higher-order oligomers even at very high protein concentrations was observed.

55. Debela M, Magdolen V, Grimminger V, Sommerhoff C, Messerschmidt A, Huber R, Friedrich R, Bode W, Goettig P: **Crystal structures of human tissue kallikrein 4: activity modulation by a specific zinc binding site.** *J Mol Biol* 2006, **362**:1094-1107.

56. Zhang X, Konarev PV, Petoukhov MV, Svergun DI, Xing L, Cheng RH, Haase I, Fischer M, Bacher A, Ladenstein R, et al.: **Multiple assembly States of lumazine synthase: a model relating catalytic function and molecular assembly.** *J Mol Biol* 2006, **362**:753-770.

57. Lee KK, Tsuruta H, Hendrix RW, Duda RL, Johnson JE: **Cooperative reorganization of a 420 subunit virus capsid.** *J Mol Biol* 2005, **352**:723-735.
58. Fetler L, Kantrowitz ER, Vachette P: **Direct observation in solution of a preexisting structural equilibrium for a mutant of the allosteric aspartate transcarbamoylase.** *Proc Natl Acad Sci U S A* 2007, **104**:495-500.
- **59. Vestergaard B, Groenning M, Roessle M, Kastrup JS, van de Weert M, Flink JM, Frokjaer S, Gajhede M, Svergun DI: **A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils.** *PLoS Biol* 2007, **5**:e134.

The amyloid fibril formation is a cause of many critical diseases and also commonly observed for a number of protein drugs, such as insulin. A series of time resolved SAXS data from fibrillating insulin could not be adequately fitted by linear combinations of the scattering from insulin monomers and mature fibrils. SVD decomposition and analysis of the residuals of the two-component fits yielded the scattering pattern from the third major component, an intermediate oligomeric species. Low-resolution models were constructed *ab initio* for the fibril repeating unit and for the oligomer, the latter being a helical unit composed of five to six insulin monomers. The growth rate of the fibrils was proportional to the amount of the intermediate present in solution, suggesting that these oligomers (and not monomers, as widely believed) elongate the fibrils. Based on the shape and size of this fibrillation precursor, a novel elongation pathway of insulin and the results suggest a conceptually new basis of structure-based drug design against amyloid diseases.

60. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M: **A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering.** *Proc Natl Acad Sci U S A* 2005, **102**:17002-17007.
61. von Ossowski I, Eaton JT, Czjzek M, Perkins SJ, Frandsen TP, Schulein M, Panine P, Henrissat B, Receveur-Brechot V: **Protein disorder: conformational distribution of the flexible linker in a chimeric double cellulase.** *Biophys J* 2005.
62. Marino M, Zou P, Svergun D, Garcia P, Edlich C, Simon B, Wilmanns M, Muhle-Goll C, Mayans O: **The Ig doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from titin.** *Structure* 2006, **14**:1437-1447.
- *63. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI: **Structural characterization of flexible proteins using small-angle X-ray scattering.** *J Am Chem Soc* 2007, **129**:5656-5664.

A novel approach is proposed for structural characterisation of inherently dynamic macromolecules including intrinsically disordered and multidomain proteins with flexible linkers based on SAXS data. To take the macromolecular flexibility into account, coexistence of different protein conformations contributing to the scattering is assumed. In the ensemble optimization method, a genetic algorithm is applied to select few conformations from a pool containing a large number of randomly generated models covering the protein configurational space whose linear combination fits the experimental scattering data. If the scattering data from deletion mutants are available, they can be fitted simultaneously, taking appropriate residues ranges in the conformers. The efficiency of the method is demonstrated on simulated and practical examples of unfolded and multidomain proteins.

- *64. Uzawa T, Kimura T, Ishimori K, Morishima I, Matsui T, Ikeda-Saito M, Takahashi S, Akiyama S, Fujisawa T: **Time-resolved small-angle X-ray scattering investigation of the folding dynamics of heme oxygenase: implication of the scaling relationship for the submillisecond intermediates of protein folding.** *J Mol Biol* 2006, **357**:997-1008.

A time-resolved SAXS study of the folding dynamics of highly helical 22 kDa heme oxygenase (HO). Polypeptide collapse caused by the coil-globule transition at the initial folding stage is analysed in terms of the scaling behavior between R_g and chain length for the initial intermediates. Sub-millisecond time frames of synchrotron radiation SAXS patterns from initially unfolded HO demonstrated a rapid decrease of R_g corresponding to the formation of the burst phase intermediate. The collapse was followed by the increase of R_g after several tens of milliseconds reaching its maximum at about 0.5 s, pointing to the formation of oligomeric intermediates. It was demonstrated that the amplitude of the burst phase (monomeric intermediates) is independent of the concentration of HO whereas the amplitude of the oligomeric intermediate phase showed concentration dependence. The study revealed a complex folding mechanism but its initial stage could be described by Flory's theory of polymers.

- *65. Arai M, Kondrashkina E, Kayatekin C, Matthews CR, Iwakura M, Bilsel O: **Microsecond hydrophobic collapse in the folding of Escherichia coli dihydrofolate reductase, an alpha/beta-type protein.** *J Mol Biol* 2007, **368**:219-229.

The early folding events of an 18 kDa α/β -type protein dihydrofolate reductase (DHFR) from *E. coli* is investigated by SAXS using a continuous-flow mixing device. The scattering patterns were measured from 300 μ s to 10 ms after the refolding reaction initiation. The analysis of the R_g values and the behavior of the Kratky plots at higher angles ($s^2I(s)$ vs. s) indicated that significant compaction (about 50% of the total change between the unfolded and folded states) occurs within 300 μ s of refolding. DHFR at 300 μ s represents an intermediate globule and not a linear combination of the native and the unfolded states. The protein conformation stayed unchanged from 300 μ s to 10 ms confirming that the subsequent folding after the major chain collapse occurs on a considerably longer timescale. The DHFR folding trajectory is therefore described by a hydrophobic collapse rather than the framework model of protein folding.

66. Fischetti RF, Rodi DJ, Gore DB, Makowski L: **Wide-angle X-ray solution scattering as a probe of ligand-induced conformational changes in proteins.** *Chem Biol* 2004, **11**:1431-1443.