Analysis of X-ray and neutron scattering from biomacromolecular solutions
Maxim V Petoukhov\(^1,2\) and Dmitri I Svergun\(^1,2\)

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.

Addresses
\(^1\)EMBL, Hamburg Outstation, Notkestraße 85, D-22603 Hamburg, Germany
\(^2\)Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia

Corresponding author: Svergun, Dmitri I (Svergun@EMBL-Hamburg.DE)

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 increasing 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 molecular 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 2\(\theta\) between the incident and scattered beam (here, only elastic scattering is considered, and the radiation wavelength \(\lambda\) 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 (\(D_{\text{max}}\)). Since 1960s, SAS was used to acquire structural information about the overall shapes of proteins in the absence of crystals. During the past 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 subsecond range on a third generation synchrotron) and require moderate amount of purified material (about a few milligrams 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 \textit{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 could 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 $D_{\text{max}}$ 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, ‘colour’ 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 nucleo-protein complex in different H2O/D2O mixtures utilizing natural contrast between nucleic acids and proteins [3]). Using specific perdeuterium of individual proteins, the approach is effective also for multiprotein complexes [16**]. Multiple curves fitting is also applicable in X-rays from multidomain or multisubunit proteins, if the scattering from partial constructs is available. In this case the contrast of the domain/subunit is defined by its absence or presence in the given construct corresponding to the scattering pattern [17*].

*Ab initio* methods are currently used for various macromolecules and complexes, including, for example, the studies of the complex between a hamster prion and DNA [18*], of autoinhibited and mutated c-Abl tyrosine kinase [19**], of human fibrinilin-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*] and 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. SANS/SANS from atomic models is evaluated, for example, by the programs CRYSOL [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 SXS/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] that 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, for example, for the cytosolic factor of NADPH oxidase [42*], for dystrophia myotonica.
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 eukaryotic 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 T. 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 microsecond 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 $s = 0.3–0.4$ 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 nm 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 $s = 2–3$ nm$^{-1}$ (0.2–0.3 nm resolution) are sensitive to the internal structure and can be utilized to probe, for example, 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-resolution and low-resolution experimental techniques and bioinformatic tools is a powerful instrument for hierarchical systems at different levels 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.biocina.org/) and ElNemo server to generate protein models using normal mode analysis (http://igs-server.cnrs-mrs.fr/elnemo/index.html).

Acknowledgements

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

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


In this paper, a comprehensive set of tools is described for SAS data analysis from biological macromolecules implemented in the program package ATLAS 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 ($R\_exp$, $D\_max$, MM, V), to analyse the composition of mixtures, and 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 analyse the quaternary structure of homodimer and heterodimer, 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.


In this paper, new ab initio and rigid body modelling methods are described or 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 the component 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 successfully fit multiple solution scattering curves. The reliability of the models is enhanced by using additional information (e.g. distance restraints between specific residues or nucleotides and orientation restraints from NMR residual dipolar couplings). Still, examples are presented where multiple solutions are possible and ways of reducing the ambiguity are discussed.


A complex of a histidine kinase inhibitor 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 KinA, Sda 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 H$_2$O/D$_2$O mixtures was used in rigid body modelling of the dimeric complex with P2$_2$ symmetry constraint. Fourteen independent SANS models were constructed and the complex structure as the Sda molecule and CA domains at the opposite ends of the DHp stalk of KinA. As an additional check, ab initio multiphasic 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.


This paper is an example of ab initio and rigid body modelling of multidomain proteins on the example of multiple scattering pattern data sets from point mutants. Simulated annealing-based optimization is applied to analyse the polyproline trinucleotide binding protein (PTB) comprising four domains (each about 10 kDa, structures solved by NMR) interconnected by linkers. A total of seven SAXS curves from the full-length protein are fitted from all possible sequential combinations of domains were fitted simultaneously to analyse the PTB structure. The multiphasic 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.


This is a combined SAXS/NMR study of the high-affinity complex between a hamster prion protein ($MM = 16\, 000$ Da) and a DNA sequence. Ab initio 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. The assignment of the C-terminal domain 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 the DNA-bound prion suggest the same overall fold and secondary structure compared with the structure of the protein in both the 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 probably be a chaperone candidate converting the normal, cellular form into the disease-causing isoform.


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 of the c-Abl fragment. The model for the N-terminal and two other key contacts in the autoinhibited state, 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 with 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 promote the active state of the kinase and suggests a critical role of N-terminal cap segment in autoinhibition.


In this paper, solution structure of human fibrillin-1, a multdomain extracellular matrix protein with MM = 330 kDa, is analysed using SAXS, light scattering and EM. Fibrillin-1 was found 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 non-linear 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 used to model the complex. Molecular insights into domain arrangement. One-domain, two-domain 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 151 to the C terminus. The methodology used can be applied to other EGF-containing extracellular matrix and membrane proteins.


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 Cpa1-1 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 Cpa1-1 in solution provide a much better fit to the SAXS data. The scattering from a shorter C-terminus 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 constructs of Cpa1-1 agree well with the appropriate portions of the SAXS-based models of the full-length protein.


This paper is 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 monomeric DNA-bound conformation of ligase is not compatible with SAXS. A planar ring heterotetrameric assembly of PCNA in the crystal in the absence of DNA is in good agreement with the SAXS data confirming the stability of the trimer. The interaction of ligase and PCNA analysed by SAXS indicates formation of a stable DNA complex. The ab initio 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 planar 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 model of the complex suggests 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.


Human pyruvate dehydrogenase (PHD) is a complex consisting of multiple copies of pyruvate dehydrogenase (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 cosasheiral faces. A 144 kDa complex of E3 with E3BP construct was investigated by SAXS, native PAGE, analytical ultracentrifugation and isothermal titration calorimetry. The overall parameters from SAXS show that the two E3BP constructs interact with a single E3 homodimer to form the PHD complex 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 analysed by rigid body modelling against the location of E3BP in the SAXS data and yielding the best agreement to the SAXS data. The resulting rigid body model agrees well with the ab initio SAXS envelope. In contrast with the rigid body 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 E3BP across the surface of the E2 core assembly.


SAXS data suggest that the catalytic Pr state of light-sensitive bacteriophage T4 controlling gene expression is dimeric in solution (167 kDa), and a Y-shape reconstructed ab initio correlates well with earlier EM images of pea phycotozym. Homologues with known atomic structure were found for the chromophore binding (CBD) and dimerization + catalytic domains but not for the phycotozym N terminal domain (PHY). The base of the Y-shaped SAXS density resembles closely the H-shaped dimeric CHK, the peripheral part of the Y arms agrees well with the shape of CB and the rest density is able to accommodate the missing PHY domain. The proposed model helps in understanding the autoregulatory mechanism of a histidine in the dimerization domain of bacteriophytochrome.


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 centre. 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 with the helix 1 of an ATPase domain. Finally, the full model, activated, full-length NtrC was produced by adding the three copies of a dimeric DNA-binding domain using three-fold symmetry to fill the occupied 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.


This paper presents and validates the crystal structure of KChIP1–Kv4.3 T1 complex influencing potassium channel trafficking and gating. Crystallographic refinement of a stable T1–KChIP1 complex yielded rigid body modelling against SAXS data produced similar results showing a single molecule of ligase bound to the PCNA trimer. The SAXS model of the complex suggests 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.


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


A SAXS 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% D2O) so that the net scattering curve could be obtained. Mostly the scattering function of apoB-100 is dominated by a single maximum (D_00) of 60 nm and the distance distribution function computed as a Fourier transformation of the intensity displayed several distinct maxima suggesting independent compact folded domains. The ab initio model reconstructed from SAXS 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, 
Ohrstedt U, Miguel RN, Blomberg M, Vande Woude GF, 
Skoglund U et al.: Structural basis of hepatocyte growth factor/sca

factor and MET signalling. Proc Natl Acad Sci USA 2006, 
103:4046-4051.

This is 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 apoB-100 shape of the 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.


37. Fernandez AN, Furtado PB, Clark SJ, Gilbert HE, Day AJ, Sim RB, Perkins SJ: Association and structural properties of the region of complement factor H encompassing the Tyr402His disease-


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 deacetylatedheparin. Both methods indicate a linear arrangement of SCR-6, SCR-7 and SCR-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 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.


42. Durand D, Cannella D, Duboscvard V, Pebay-Peyroula E, 

A SAXS study of the cytosolic factor p47phox in autoinhibitory resting state, a modular protein of 46 kDa, provides the assembly of the NADPH oxidase complex. The experimental MM value was compatible with a 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. An ab initio modelling based on a single tertiary map provided a closer model that used the experimental map, but the result was not very significant due to a flexible segment.


44. Novak E, Panjikar S, Morth JP, Jordanova R, Svergun DI, 

The paper reports the high-resolution crystal structure of the ATP-binding domain (108 residues) of the sensor histidine kinase PrbB from Mycobacterium tuberculosis and the solution structures of two-domain and three-domain constructs containing additionally a 67-residue phosphorylation domain and also a HAMP (hemopexin) linker. Both constructs were found to be dimeric in solution, with the dimer formation mediated via the phosphorylation domain. The scattering patterns from the two-domain and three-domain constructs were simultaneously fitted by a model assuming a two-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 (potentially flexible) HAMP domain occupying the part of the molecule above the catalytic domain.


46. Vestergaard B, Sanyal S, Roessle M, Mora L, Buckingham RH, 
Kastrup JS, Gajhede M, Svergun DI, Ehrenberg M: The SAXS 
structure solution 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, 

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 E. coli. 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 chromo-
domain insert in ABC2 located at the tip. Cryo-EM structure of the ATP-bound form of eEF3 in complex with the post-translational-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 stabilization pH of 5.2 indicates a more compact nature of eEF3 pointing to pH-induced conformational change within eEF3 in solution.

The human tumour suppressor p53 is a homotetrameric transcription factor (more than 390 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 crosstranscript domain structure [49] showing dissociated tetramerization domains did not fit the experimentally SAXS data. The structure of the complex of p53 with 24 bp DNA independently determined by SAXS and negative-stain EM displays a complex compact with the core domains closing around DNA. Interestingly, negatively stained EM analysis of the conformationally coupled core domain dimers at the ends. In contrast, the calculated cross-shaped structure with tetrameric contacts and a pair of loosely modelling accounting for NMR-derived interfaces revealed an extended functional plasticity of p53 tumour suppressor protein reveals the basis for its modulatory interaction with other proteins. *EMBO J* 2006, 25:5191-5200.


Two high-resolution models were proposed for the Josephin domain of ataxin-3, a 182-residue protein involved in the ubiquitin/proteasome pathway, using NMR (PDB codes 1y2b 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 including Bayesian methods to validate the models and to demonstrate that only one of them, with the open cleft, is compatible with the experimental information.


54. Graziano V, McGrath WJ, Yang L, Mangel WF: SARS CoV main proteinase: the monomer–dimer equilibrium dissociation constant determined from initially unfolded HO demonstrated a rapid decrease of the intermediate present in solution, suggesting that these oligomers (and not monomers, as widely believed) elongate the fibril. The amyloid fibril formation is a cause of many critical diseases and also commonly observed for a number of protein drugs, such as insulin. *J Mol Biol* 2006, 362:1094-1107.


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 fibril. 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.


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 conformational contributions to the scattering is assumed. In the ensemble optimization method, a genetic algorithm is applied to select few conformational models characterizing the protein with conformational 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.


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 behaviour between $R_g$ and chain length for the initial intermediates. Submillisecond 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 Fño’s theory of polymers.
The early folding events of an 18 kDa \(\alpha/\beta\)-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 \(\mu\)s to 10 ms after the refolding reaction initiation. The analysis of the \(R_g\) values and the behaviour of the Kratky plots at higher angles (\(s^2I(s)\) versus \(s\)) indicated that significant compaction (about 50\% of the total change between the unfolded and folded states) occurs within 300 \(\mu\)s of refolding. DHFR at 300 \(\mu\)s represents an intermediate globule and not a linear combination of the native and the unfolded states. The protein conformation stayed unchanged from 300 \(\mu\)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.

---