

Applications of Small-Angle X-ray Scattering on Biomacromolecular Solutions

Running title (maximum 50 characters): X-ray Scattering from Biological Solutions

Maxim V. Petoukhov¹⁾ & Dmitri I. Svergun^{1)*}

¹⁾ European Molecular Biology Laboratory, Hamburg Outstation. Notkestrasse 85, 22607 Hamburg, Germany

*Corresponding Author: Svergun@EMBL-Hamburg.DE

Keywords (5): solution scattering, data analysis, low resolution structure, biological macromolecules, hybrid modeling

Abstract (250 words max): Small-angle scattering of X-rays (SAXS) is an established method for low-resolution structural characterization of biological macromolecules in solution. Being highly complementary to the high resolution methods (X-ray crystallography and NMR), SAXS is often used in combination with them. The technique provides overall three-dimensional structures using *ab initio* reconstructions and hybrid modeling, and allows one to quantitatively characterize equilibrium mixtures and flexible systems. Recent progress in SAXS instrumentation, most notably, high brilliance synchrotron sources, pave the way for high throughput automated SAXS studies including screening of external conditions (pH, temperature, ligand binding etc). The modern approaches for SAXS data analysis are presented in this review including rapid characterization of macromolecular solutions in terms of low-resolution shapes, validation of high-resolution models in close-to-native conditions, quaternary structure analysis of complexes and quantitative description of the oligomeric composition in mixtures. Practical aspects of SAXS as a standalone tool and its combinations with other structural, biophysical or bioinformatics methods are reviewed. The capabilities of the technique are illustrated by recent selected applications for the studies of biological objects. The future perspectives of SAXS and their potential impact for structural molecular biology are discussed.

Introduction

Systemic approach dominating the field of modern structural biology sets the focus of the structural studies on complex objects, molecular machines and processes (Gavin et al., 2006, Abrahams et al., 2011). Here, structural analysis of interactions within biological systems and prediction of their behavior based on their composition aim to elucidate the relationship between structural mechanisms and functionality. Hybrid methods are often applied combining various bioinformatics and experimental techniques covering a broad range of temporal and spatial resolution to provide quantitative understanding of the entire system.

Small-angle X-ray scattering (SAXS) is a universal structure probe for a wide variety of non-crystalline objects (e.g. macromolecular solutions, detergents, nanocomposites, alloys, synthetic and bio-polymers, biomaterials, organic/inorganic films etc). SAXS enables structural studies in an extremely broad range of molecular sizes from kDa to GDa and in experimental environments from extreme (e.g. high pressure or cryo-frozen) to nearly native. This allows one to simultaneously monitor the interactions in hierarchical systems at different levels of structural organization. SAXS is also able to operate *in situ* and in a time-resolved manner yielding unique information about the kinetics and dynamics of processes.

In a typical biological SAXS experiment, a monochromatic and collimated X-ray beam is illuminating macromolecules chaotically oriented in solution and their scattering pattern is recorded by an X-ray detector. For dilute solutions, the scattered intensity I is isotropic and depends only on the scattering angle 2θ (Fig. 1). Measurements of the sample containing the macromolecules are complemented by those of the blank solvent (typically dialysis buffer) and the latter scattering is subtracted from the sample intensity. The resulting scattering profile usually expressed as a function of the momentum transfer $s = 4\pi\sin\theta/\lambda$ (here, λ is the wavelength) is proportional to the scattering from a single particle averaged over all orientations and to the solute concentration.

A scattering profile from monodisperse solution of non-interacting identical particles carries information about the major geometrical parameters of the particle. Particularly, molecular mass (MM) of the solute and its radius of gyration (R_g) are derived from the slope of the Guinier plot

$\ln(I(s))$ versus s^2 (which is linear at low s) (Guinier and Fournet, 1955), the values of the hydrated particle volume (V) and its specific surface (S) can be obtained using the so-called Porod invariant (Porod, 1982). MM and V provide estimates of the possible oligomeric state or may give a hint on (partial) dissociation in case of complexes. Here, MM is calculated from the forward scattering by comparison with a reference sample (i.e. standard protein), which requires precise information about the solute concentration. The use of volume, although somewhat less accurate, does not depend on concentration (Petoukhov et al., 2012). The maximum distance within the particle (D_{max}) is also readily available from the Fourier transformed SAXS pattern (Svergun, 1992). The folding state of the particle can be assessed utilizing the Kratky plot ($I(s)s^2$ versus s): a bell-shaped profile is typical for globular proteins whereby a monotonously increasing function points to unfolded ones (Mertens and Svergun, 2010).

In addition, to the overall parameters, one dimensional (1D) solution scattering profiles $I(s)$ allow one to meaningfully analyze three dimensional (3D) structures (Fig. 2). First of all, low resolution particle shapes can be reconstructed *ab initio* (i.e. without any prior information) from the scattering patterns. For the cases when a high resolution model is available, it can be validated against the experimental SAXS data and possible conformations or oligomeric states can be rapidly screened to identify the biologically relevant structure. In case of mixtures and flexible systems, solution scattering provides means for quantitative description of the sample composition. For the multisubunit particles, rigid body analysis can be applied based on the models of individual subunits to reconstruct the quaternary structure of the complex. The major SAXS-based modelling methods and some applications will be presented below.

Ab initio Shape Reconstruction

Although the reconstruction of a 3D model from 1D scattering pattern is an ill-posed problem potentially yielding multiple shapes compatible with one and the same scattering profile, still, the scattering intensity can be considered to be a particle shape's footprint and the ambiguity of the reconstruction can be reduced by imposing the restraints to the model. These restrictions in case of *ab initio* shape determination consist in representing the macromolecule at low resolution as a homogeneous and compact particle. Such a concept is implemented in an approach representing the particle as an ensemble of (several thousands of) densely packed beads on a regular grid (Chacon et al., 1998, Svergun, 1999). Each bead on the grid belongs either to the particle (index

= 1) or to the solvent (index = 0) so that the shape is described by a binary configuration vector. A widely used bead modelling algorithm implemented in the programs DAMMIN (Svergun, 1999) and DAMMIF (Franke and Svergun, 2009) applies simulated annealing (SA) to build a compact interconnected configuration of beads that fits the experimental data by minimizing discrepancy between the experimental data $I_{exp}(s)$ and a scattering profile computed from the model $I_{calc}(s)$:

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[\frac{I_{exp}(s_j) - cI_{calc}(s_j)}{\sigma(s_j)} \right]^2, \quad (1)$$

where c is a scaling factor, N is the number of points and σ denotes the experimental errors. The search can be done either in a confined volume (DAMMIN) or on an unlimited grid (DAMMIF). In the former case, the search volume can be defined by the maximum size of the particle D_{max} deduced from the SAXS data or is taken from complementary structural methods *e.g.* electron microscopy maps. In the latter case, the available space is expanded as necessary in the course of modelling.

For oligomeric particles consisting of identical subunits, symmetry can be taken into account as a rigid constraint so that the symmetrically related beads are added/removed to the model simultaneously. For multicomponent particles (*e.g.* nucleo- and lipo-proteins or protein complexes) there is a possibility to extend the approach to a multiphase bead modeling where the phase index varies up to the number of components describing not only the overall shape but also the internal organization of the composite molecule. The prerequisite for this is availability of the scattering profiles from the individual components (or their combinations) so that they are fitted simultaneously with the data of the intact particle (Svergun and Nierhaus, 2000).

Ab initio modelling is now routinely used in various SAXS-based studies providing a very useful first glance onto the molecule at a low (about 1.5-2 nm) resolution. Moreover, the bead models could be utilized in the complementary techniques, *e.g.* as starting points for image classification in electron microscopy.

Validation and Screening of Concurrent High-Resolution Models

Availability of atomic models from crystallography or NMR still keeps ample room for SAXS-based model evaluation and validation. Thus, NMR models might display several alternative conformations (Nicastro et al., 2006) whereas crystallographic packing forces can affect the quaternary structure of macromolecule (Svergun et al., 2000). Also, crystallographic contacts often suggest several possible oligomeric states and interfaces between the individual subunits in solution. SAXS comes into play to identify the biologically active conformation and oligomeric composition and/or to validate structural similarity between crystal and solution by comparison of the experimental scattering profile with the curve computed from the model(s). Evaluation of the theoretical curve from high resolution model in solution requires taking into account not only the scattering from atomic structure *per se* but also the solvent effects. These include the scattering contributions from the excluded volume inaccessible to the solvent and from a hydration shell surrounding the macromolecule (with an averaged density about 10% higher than that of the bulk solvent). A program CRY SOL (Svergun et al., 1995) computes the X-ray scattering profile from atomic coordinates and can either fit the provided experimental data by adjusting the hydration shell density and the excluded particle volume or predict the theoretical curve with the default parameters. Recently, other programs have become available which, in particular, allow to accurately compute the wide-angle (WAXS) scattering patterns (Bardhan et al., 2009, Virtanen et al., 2011).

Given the atomic model, the theoretical profiles can be rapidly computed even for large macromolecular assemblies (especially fast are computations using spherical harmonics (Stuhrmann, 1970)). This opens a possibility of fast screening by computations of multiple concurrent models (e.g. predicted using bioinformatic methods) against the SAXS data (Zheng and Doniach, 2002, Pons et al., 2010).

Rigid Body Modeling of Complex Particles and Multidomain Proteins

The strength of SAXS is further revealed in hybrid approaches when the technique is used in integrated manner with other structural methods. Large macromolecular complexes responsible for accomplishment the most important cellular functions are also rather challenging for crystallization and subsequent structure determination. Other techniques also have limitations, thus, structural NMR is employed to relatively small proteins, whereas cryo-EM is rather time demanding and can be used for objects starting from ca 200 kDa. Rigid body modeling approach

in SAXS utilizes atomic models of individual subunits or domains deduced by high resolution methods to analyze the quaternary structure of a complex or multidomain protein in solution. Here, structural changes in response to external conditions as well as time-resolved studies are quite feasible.

In SAXS-based rigid body modeling, the atomic building blocks (subunits or domains) are assumed to keep their tertiary structures upon formation of the assembly. They are moved and rotated with respect to each other to form the entire particle whose scattering intensity is directly compared with the experimental data (as opposite e.g. to the EM-based docking, where models are fitted into density maps (Wriggers et al., 1999)). Location of each rigid body within the composite model is described by six roto-translational parameters: three Euler angles define the orientation in respect to the reference one and three Cartesian shifts give the position of the subunit's center in respect to the origin. There are several algorithms (Konarev et al., 2001, Petoukhov and Svergun, 2005, Petoukhov et al., 2012) available for various modeling scenarios, from local refinement of homo and heterodimers to automated global optimization of multisubunit complexes (SASREF), multidomain proteins (BUNCH) and their combination (CORAL). Apart from providing good fits to the experimental data, the resulting models also must have physical sense, particularly steric clashes should be absent and the entire assembly must be interconnected. An SA-based approach is typically employed that allows one to incorporate these restraints and potential additional information from other methods. The target function is minimized, which not only contains discrepancy between the experimental SAXS curve and the fit but also a set of penalties formulating additional requirements to the model. In particular, desired inter-residue contacts based on the complementary data from NMR chemical shifts (Xu et al., 2008), mutagenesis (Gherardi et al., 2006), fluorescence resonance energy transfer (FRET) or bioinformatics predictions can be introduced as spring force potentials. NMR residual dipolar couplings (RDCs) data, if available, are useful to enforce the mutual subunit orientations by constraining the rotations of rigid bodies (Marino et al., 2006). This constraint significantly increases the reliability of rigid body models as SAXS is generally more sensitive to translations than to rotations. For oligomeric assemblies, symmetry may be taken into account which significantly reduces ambiguity of the reconstruction. If multiple scattering patterns from subcomplexes or deletion mutants with presumably same subunit arrangement as in the intact

assembly are available, they are typically fitted simultaneously to yield better description of the internal organization of the complex or multidomain protein (Petoukhov et al., 2006).

In the specific case of multidomain proteins with large structurally unassigned portions (*i.e.* disordered linkers between the rigid domains), a combined approach is applied which represents the missing portions as self-avoiding dummy residues (DR) chains (Svergun et al., 2001, Petoukhov and Svergun, 2005) of appropriate length connecting the termini of the subsequent domains. The combined algorithm searches for the optimal positions and orientations of the domains and probable configuration of DR linkers taking care about proper bond and dihedral angle distributions to simulate native-like conformations. The use of DR chains not only provides more accurate scattering computation of modular proteins but also serves as a constraint to the inter-domain distances.

Analysis of Polydisperse Systems

The requirement of sample monodispersity is a major prerequisite for reliable 3D reconstructions from SAXS data, but the method is also capable to deal with polydisperse systems. In fact, SAXS is a method of choice in the studies of oligomeric equilibrium, assembly and dissociation processes, analysis of stoichiometry and in case of macromolecules with significant degree of flexibility. For dilute systems without inter-particle interactions, the resulting scattering profile is a linear combination of the individual scattering intensities of its individual components, where the coefficients are their volume fractions, v :

$$I(s) = \sum_k v_k I_k(s) \quad (2)$$

If the scattering intensities of the pure components are available (e.g. if their structures are known or a component is monodisperse at some conditions), the volume fractions in an equilibrium mixture can be readily obtained. The program OLIGOMER (Konarev et al., 2006) finds the volume fractions of the contributing components by nonnegative linear least-squares algorithm (Lawson and Hanson, 1974) minimizing the discrepancy (Eq. 1) between the predicted composite curve and the experimental data. This provides the opportunity to investigate the effect of various external conditions (concentration, temperature, ionic strength, cofactors etc) on the system.

Suitable preparation and handling of the sample and usage of online high-performance liquid chromatography purification in many cases allows one to reach high purity solutions whose scattering profiles are suitable for *ab initio* and rigid body modeling (David and Perez, 2009). Some samples like low affinity transient complexes and weak oligomers in dynamic equilibrium state with its monomers still present a challenge despite all efforts. The ability to account for possible polydispersity was recently added to 3D modeling algorithms from ATSAS package (Petoukhov et al., 2012). Particularly, the rigid body modeling program SASREF_MX is capable to reconstruct multisubunit complex at low resolution in the presence of dissociation. In SA-driven restoration of the mutual subunit arrangement, the experimental scattering curve is fitted by weighted sum of the intensities computed from the entire multisubunit complex and from the specific dissociation products. Here, the volume fractions of the bound and the dissociated states in the mixture are additional minimization parameters of the modeling besides the positions and orientations of individual subunits. This approach is useful for studies of structural transitions and oligomerisation processes.

Disordered proteins represent yet another type of equilibrium mixtures. Even in the absence of mass polydispersity, their intrinsic conformational flexibility makes single-particle based methods barely applicable. An ensemble optimization method (EOM) (Bernado et al., 2007) was proposed to characterize the conformational variability of such systems by fitting the experimental data with the scattering profile calculated from mixtures of conformers. A genetic algorithm is applied to select a small representative ensemble from a large pool of randomly generated structures so that the effective scattering profile of the ensemble members fits the experimental data. The selected models do not necessarily correspond to actual conformations of the flexible molecules in solution, but instead they provide reasonable estimates of the distribution of structural descriptors such as R_g , D_{max} , and anisometry, and allow one to evaluate the degrees of flexibility/rigidity, compactness/extension and folding of the protein. EOM allows one to consider also oligomeric multichain proteins (either with or without symmetry) whereby the original pool contains both monomers and multimers. In this case, the selected ensemble also provides an idea on the volume fraction of the monomeric and assembled states together with their representative conformations. Generally, EOM is a valuable tool for the investigation of intrinsically disordered proteins, studies of protein (un)folding and for structural dynamics.

Examples of SAXS Applications

Below, we shall illustrate recent applications of the advanced SAXS data analysis methods in the studies of different biological objects. The number of publications employing SAXS in the structural studies of macromolecular solutions is constantly and rapidly growing and we can of course not cover the entire field here. We have therefore selected several recently published papers to demonstrate the capacity of SAXS, and especially its synergistic use with complementary methods.

Validation of an Unusual Oligomeric Assembly in the Crystal

An example illustrating the power of SAXS to validate crystallographic models is given by a recent study of dihydrolipoyl acyl-transferase E2 assembly of the catalytic core of an archaeal 2-oxoacid dehydrogenase in solution (Marrott et al., 2012). This enzyme forms homotrimers and further associates into symmetric higher-order complexes. Depending on the source, octahedral (cubic symmetry) or icosahedral particles are typically formed whereby each monomer has virtually the same environment according to quasi-equivalence principles. Surprisingly the crystal structure of E2 core from *Thermoplasma acidophilum* demonstrated a rather unusual 42-mer assembly. To validate such a unique conformation, synchrotron SAXS data from E2 were collected at the EMBL X33 beamline (DORIS III storage ring at DESY, Hamburg) (Roessle et al., 2007, Round et al., 2008) and processed using the standard procedures (Konarev et al., 2003). The resulting scattering profile was typical for a globular hollow particle and the defined forward scattering intensity I_0 and the total excluded volume of the solute yielded molecular weight close to 1Mda in agreement with the 42-meric assembly. The experimental R_g and D_{max} values of 85 Å and 220 Å, respectively, were also consistent with the crystallographic structure. *Ab initio* reconstruction using the bead modeling with DAMMIF produced oblate spheroid shapes well superimposable with the crystal structure (Fig. 3). Finally, the scattering pattern computed from the crystallographic coordinates by CRY SOL has shown perfect agreement with the experimental SAXS profile. The comparison of the scattering curves computed from homological 24-mer and 60-mer assemblies with the experimental data clearly ruled out these two options. These results confirmed that the unique 42-meric assembly is not an artifact caused by the crystallographic packing forces, but instead is a novel type of protein assembly which

opens new ways for manipulation of self-assembling biological complexes in non-identical environments.

Study of Modular Protein Structure by Joint Use of SAXS and NMR

Modular proteins consisting of several well-defined domains separated by unstructured linkers are challenging objects for structural studies due to their inherent flexibility. A new concept combining SAXS and NMR data has recently been proposed (Bertini et al., 2010) for such proteins. The approach called Maximum Occurrence (MO) provides the maximum percent of time that flexible proteins can spend in any given conformation. Similarly to the EOM, a large pool of random structures is first generated. The MO of any conformation randomly picked from the pool is calculated as the maximum weight, which this conformation can have in the ensemble under condition that the ensemble is still able to reproduce the heterogeneous experimental data. When calculating the MO, the given conformation is taken with the fixed weight and all the other conformations can have variable weights in the ensemble to best fit the data. MO was first applied to characterise the conformation of calmodulin (CaM), a protein with a large conformational freedom. CaM consists of two domains, and their mutual arrangement was studied using the SAXS profile, pseudocontact shifts (PCS) and self-orientation residual dipolar couplings (RDC) from NMR. The first two data types (SAXS and PCS) are mostly constraining the interdomain distances whereby RDCs are sensible to their orientations. It was shown that, despite of high conformational variability, mutual arrangement of C- and N-terminal domains is not random and that extended conformations are the most abundant in solution (corresponding MO value reached 35%), whereby the “closed” and “fully extended” conformations trapped in the crystal structure of CaM (Fallon and Quioco, 2003, Babu et al., 1988) have MOs of only 5 and 15%, respectively. The MO approach is an excellent example of a synergistic use of SAXS with NMR and can be accessed via a MaxOcc web portal (<http://py-enmr.cerm.unifi.it/access/index/maxocc>) utilizing the GRID services for computations.

Rigid Body Analysis of Calmodulin-Activated Glutamate Decarboxylase

An example of the quaternary structure analysis by SAXS together with complementary methods is given by the study of plant glutamate decarboxylase (Gad) in complex with CaM. Gad catalyzes conversion of glutamate to γ -aminobutyrate and is regulated by several factors

including CaM. Plant Gad is a symmetric hexamer (MM about 340 kDa) with known crystal structure. SAXS was employed to elucidate the mode of binding and stoichiometry of the complex formed by binding of CaM to the C-terminal domain of Gad. First, hexameric assembly of Gad alone was validated in solution. The scattering data from solution of free Gad yielded structural parameters compatible with crystallographic hexamer. The agreement between the theoretical curve computed from atomic coordinates by CRY SOL and the experimental profile was quite reasonable. The fit could be further improved by adding six copies of the flexible C-terminal missing in the crystallographic model by BUNCH. The increase of MM of the complex compared to free Gad, as estimated from the forward scattering intensity, pointed to binding of three CaM molecules to one Gad hexamer. This result was seconded by a number of independent experiments including mass determination of the Gad–CaM complex using size-exclusion chromatography and static light-scattering as well as by fluorimetric and UV-vis titrations of Gad with CaM. All these methods confirmed the 1:3 mode of binding. Significant increase of the SAXS-based R_g and D_{max} values from Gad to the Gad–CaM complex indicated that the CaM molecules were located at the periphery of the complex. The comparison of the distance distribution functions suggested that Gad is located in the centre of the complex and preserves its overall structure. This finding was further supported by *ab initio* bead reconstructions which produced models consisting of triangular cores with quasi-P32 symmetry corresponding to the Gad hexamer and three protruding clusters of beads representing the CaM molecules. Available high-resolution structural models were utilized for yet more detailed analysis of the quaternary structure. Given the established 1:3 stoichiometry of the complex, one CaM molecule is bound to two C-terminal domains of the adjacent Gad monomers. An NMR model of homologous *X. laevis* CaM bound to two short peptides of petunia Gad (PDB code 1NWD) provided a structural template for the rigid body modeling. Three copies of this homology model were arranged in respect to the Gad crystallographic hexamer to fit the experimental SAXS curve. The reconstruction was done by BUNCH, which also provided idea on the conformation of the six linkers connecting Gad core monomers with the peptides representing its C-terminal. Modelling with the imposed 3-fold symmetry axis yielded reasonable fit to the experimental data. In the P3 models the three CaM moieties were positioned approximately on the three 2-fold axes of Gad hexamer and were displaced by about 20 Å from the hexameric core. Given the loose packing of Gad and CaM in respect to each other, some degree of flexibility was assumed for CaM and the

modeling was also performed without symmetry constraints so that individual CaM subunits may display different orientations and distances from Gad core. The resulting model (Fig. 4) provides slightly better fit and is similar in the overall appearance to symmetric ones with CaM stretching outwards near the equatorial plane of Gad whereby the differences in the arrangement of individual CaM molecules in the complex provided an idea of possible conformational variability. This example shows that SAXS is useful for establishing the stoichiometry of the complexes and for structural analysis of macromolecular assemblies that could not be analyzed by high resolution methods, e.g. due to flexibility or weak interactions between the partners.

Study of Ionic Strength Dependence of Actin Complexes

Another example of multidisciplinary use of SAXS with crystallography and NMR is provided by a comprehensive study of interactions of actin with intrinsically disordered actin-binding peptides (Didry et al., 2012). SAXS data were collected at the SWING beamline (Paris, France) on homogeneous G-actin and its 1:1 complexes with β -Thymosin (β T) domain eluted from a size exclusion chromatography column (David and Perez, 2009). The use of online HPLC allows in many cases to significantly reduce the polydispersity of the sample and is now becoming a standard setup at the modern SAXS beamlines. The measurements of G-actin in non-polymerizing low ionic strength conditions and at physiological ionic strength provided similar scattering profiles indicating that the increase of ionic strength does not result in major conformational change of the free protein. The scattering data from the complexes of G-actin with isolated first β T domain (CibD1) of *Drosophila Ciboulot* and with a CH3 chimera (containing an extra linker) collected at low ionic strength could be fitted by crystallographic model with sequestering conformations of the β T:actin complexes. The sequestering conformation of CibD1 on G-actin at these conditions was somewhat unexpected but fully confirmed by actin assembly assays revealing the sequestering activity of CibD1 at low ionic strength. In contrast, at physiological ionic strength the complexes of CibD1 and CH3 with G-actin, had the most divergent functions and the differences in conformation was captured by SAXS. While the scattering profile of CH3:actin was compatible with the above sequestering conformation, the curve from CibD1:actin pointed to the conformational change of the peptide. Hybrid modeling performed by BUNCH revealed that the interactions of the C-terminal half of CibD1 with the pointed face of G-actin are released. Interestingly, the release of C-terminal

portion is consistent with the disorder of this region observed in the crystal structure. This demonstrates how can SAXS analysis of the conformational changes in response to variation of the external conditions aid in establishing the functional properties of biological systems.

Future Prospects: High Throughput, Full Automation

The variety of tools for solution scattering data analysis from biological macromolecules developed so far allows an experienced researcher to address challenging questions of modern structural biology. As demonstrated in the examples mentioned above, SAXS is capable of fast determination of overall geometrical parameters and low-resolution shape reconstruction in the absence of *a priori* structural information to structure validation, identification of biologically active oligomers and atomic-structure-based molecular modelling. To streamline the future SAXS applications, the efforts are now targeting full automation of the experiment covering the data acquisition and reduction and going to data interpretation and 3D model building (Hura et al., 2009, Franke et al., 2012, Petoukhov et al., 2012, Blanchet et al., 2012). At dedicated biological SAXS synchrotron beamlines, unattended and remote operation as well as automated primary processing has already been implemented (Fischetti et al., 2004, Round et al., 2008, Hura et al., 2009, Pernot et al., 2010). Automated data interpretation and model building can be performed by a dedicated expert system. A prototype of such a system, DANESSA (Data ANalysis Expert System for Small Angles, presented at XIV International Conference on Small-Angle Scattering, SAS 2009 in Oxford, UK, http://www.sas2009.org/dms/sas2009/SAS2009_Abstract_Book.pdf) shall be of significant help in treatment of large amounts of scattering data. The integrated system can perform analysis of single scattering curves looking for the best models, but, even more important, it may simultaneously interpret multiple data sets in project-oriented approach (e.g. when analyzing deletion mutants or sub-complexes). The present prototype may not provide the ultimate answer about the object structure but instead outputs alternative models with the corresponding fits to be further evaluated by a human. Despite the present need in post-assessment, the system is already useful as it automatically considers various possible scenarios saving the researcher efforts and provides a reasonable first glance on the problem. Moreover, the output can be reutilized, if there is a need to switch to a manual mode for fine tuning, so that one is released from preparatory steps.

Progress in the automated analysis systems will first of all assist the researchers new to the field of SAXS, to effectively use the novel modelling algorithms. The intelligent computer system should however also help the experienced users to perform large scale data handling and model building more effectively as it automatically executes the tasks that would otherwise be done manually. This is extremely important in view of a large amount of scattering data acquired at modern synchrotron radiation facilities, and their treatment may present a challenge even for the experienced researchers.

Conclusions

The study of biological systems using solution SAXS became a streamline tool in modern molecular biology, with more and more groups incorporating this technique into their research programs. Most of the SAXS analysis tools reached a mature state, their application is straightforward and can even be performed automatically. From the experimental point, all major synchrotron facilities worldwide offer beamlines capable of biological SAXS. New high brilliance beamlines specifically dedicated to solution studies are in operation or in construction. For many biological SAXS applications, the modern laboratory X-ray instruments (for example built by Rigaku, Bruker or Anton Paar) permit one to collect data of sufficient quality. The laboratory SAXS experiment usually takes much longer compared to synchrotrons (hours instead of seconds). Still, computation of overall parameters, *ab initio* shapes and hybrid modeling are possible based on the laboratory data, and these cameras can also be employed for screening of the sample quality before going to a synchrotron facility.

Most of the computer programs mentioned in this review belong to the ATSAS program suite created to process and analyse the scattering data from isotropic macromolecular systems (Konarev et al., 2006, Petoukhov et al., 2007, Petoukhov et al., 2012). These programs are publicly available for download by academic users from the URL <http://www.embl-hamburg.de/biosaxs/software.html>. Many of these programs, including the prototype of the expert system, are also available for on-line access at the ATSAS-online portal <http://www.embl-hamburg.de/biosaxs/atsas-online>.

In the present review we largely mentioned the SAXS applications, which emerged from the collaborative projects at the EMBL biological SAXS beamlines, largely to illustrate the

capacities of the novel data analysis methods. The number of biological SAXS applications by numerous groups worldwide is rapidly growing and especially exciting results are obtained in hybrid modeling projects, where the method is used together with other structural and computational techniques (Bach et al., 2012, Clerici et al., 2009, Guo et al., 2012, Hilge et al., 2009, Jehle et al., 2010). In this respect, standardization of the reporting of the SAXS data and its treatment is of high importance for adequate assessment by the reviewers and readers. Preliminary guidelines for publication of biomolecular structures from small-angle scattering data have been proposed by the International Union of Crystallography Commission on Small-Angle Scattering and summarized in a recent paper (Jacques et al., 2012).

Figure Captions

Figure 1. Schematic representation of a solution SAXS experiment

Figure 2. Overview of structural analysis options provided by SAXS

Figure 3. Validation of 42-meric assembly of dihydrolipoyl acyl-transferase E2 by SAXS. (A) Comparison of SAXS-based *ab initio* reconstruction (cyan beads) with the crystal structure. Two orientations are shown with viewpoints along the twofold (left) and the threefold (right) symmetry axes of the 42-mer. Corresponding viewpoints of schematic 42-mer polyhedron are included for clarity. (B) Fit computed from the atomic coordinates of the 42-mer (red line) to the experimental SAXS data of E2 core (shown in blue). For comparison (C) and (D) demonstrate that homologous 24-mer and 60-mer assemblies yield poor fits to the E2 scattering curve.

Figure 4. Rigid Body Analysis of Calmodulin-Activated Glutamate Decarboxylase. (A) SAXS scattering pattern of Gad:CaM complex. Experimental data are denoted by black dots and the fit provided by the best model of the complex is shown as red solid line. (B) Rigid body model of the 3:1 complex reconstructed without symmetry restraints is presented in space-filling mode. The hexameric core of Gad and the three CaM molecules are shown in cyan and yellow, respectively. Pink beads represent the DR linkers connecting the core monomers with the C-terminal peptides of Gad (shown in green) bound to CaMs. Right view shows the orientation rotated by 90 degrees about the vertical axis. The model suggests the equatorial positioning of the CaMs whose centers are located close to the three quasi-twofold axes.

References

- Abrahams, J. P., Apweiler, R., Balling, R., Bertero, M. G., Bujnicki, J. M., Chayen, N. E., Chene, P., Corthals, G. L., Dylag, T., Forster, F., Heck, A. J., Henderson, P. J., Herwig, R., Jehenson, P., Kokalj, S. J., Laue, E., Legrain, P., Martens, L., Migliorini, C., Musacchio, A., Podobnik, M., Schertler, G. F., Schreiber, G., Sixma, T. K., Smit, A. B., Stuart, D., Svergun, D. I. & Taussig, M. J. (2011). "4D Biology for health and disease" workshop report. *N Biotechnol*, 28, 291-293.
- Babu, Y. S., Bugg, C. E. & Cook, W. J. (1988). Structure of calmodulin refined at 2.2 Å resolution. *J Mol Biol*, 204, 191-204.
- Bach, A., Clausen, B. H., Moller, M., Vestergaard, B., Chi, C. N., Round, A., Sorensen, P. L., Nissen, K. B., Kastrup, J. S., Gajhede, M., Jemth, P., Kristensen, A. S., Lundstrom, P., Lambertsen, K. L. & Stromgaard, K. (2012). A high-affinity, dimeric inhibitor of PSD-95 bivalently interacts with PDZ1-2 and protects against ischemic brain damage. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 3317-3322.
- Bardhan, J., Park, S. & Makowski, L. (2009). SoftWAXS: a computational tool for modeling wide-angle X-ray solution scattering from biomolecules. *J. Appl. Cryst.*, 42, 932-943.
- Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*, 129, 5656-5664.
- Bertini, I., Giachetti, A., Luchinat, C., Parigi, G., Petoukhov, M. V., Pierattelli, R., Ravera, E. & Svergun, D. I. (2010). Conformational space of flexible biological macromolecules from average data. *J Am Chem Soc*, 132, 13553-13558.
- Blanchet, C. E., Zozulya, A. V., Kikhney, A. G., Franke, D., Konarev, P. V., Shang, W. F., Klaering, R., Robrahn, B., Hermes, C., Cipriani, F., Svergun, D. I. & Roessle, M. (2012). Instrumental setup for high-throughput small- and wide-angle solution scattering at the X33 beamline of EMBL Hamburg. *Journal of Applied Crystallography*, 45, 489-495.
- Chacon, P., Moran, F., Diaz, J. F., Pantos, E. & Andreu, J. M. (1998). Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys J*, 74, 2760-2775.
- Clerici, M., Mourao, A., Gutsche, I., Gehring, N. H., Hentze, M. W., Kulozik, A., Kadlec, J., Sattler, M. & Cusack, S. (2009). Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *Embo Journal*, 28, 2293-2306.
- David, G. & Perez, J. (2009). Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. *Journal of Applied Crystallography*, 42, 892-900.
- Didry, D., Cantrelle, F. X., Husson, C., Roblin, P., Moorthy, A. M. E., Perez, J., Le Clainche, C., Hertzog, M., Guittet, E., Carlier, M. F., van Heijenoort, C. & Renault, L. (2012). How a single residue in individual beta-thymosin/WH2 domains controls their functions in actin assembly. *Embo Journal*, 31, 1000-1013.

- Fallon, J. L. & Quirocho, F. A. (2003). A closed compact structure of native Ca(2+)-calmodulin. *Structure*, *11*, 1303-1307.
- Fischetti, R., Stepanov, S., Rosenbaum, G., Barrea, R., Black, E., Gore, D., Heurich, R., Kondrashkina, E., Kropf, A. J., Wang, S., Zhang, K., Irving, T. C. & Bunker, G. B. (2004). The BioCAT undulator beamline 18ID: a facility for biological non-crystalline diffraction and X-ray absorption spectroscopy at the Advanced Photon Source. *J Synchrotron Radiat*, *11*, 399-405.
- Franke, D., Kikhney, A. G. & Svergun, D. I. (2012). Automated acquisition and analysis of small angle X-ray scattering data. *Nuclear Instruments and Methods in Physics Research A*, *689*, 52-59.
- Franke, D. & Svergun, D. I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.*, *42*, 342-346.
- Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B. & Superti-Furga, G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, *440*, 631-636.
- Gherardi, E., Sandin, S., Petoukhov, M. V., Finch, J., Youles, M. E., Ofverstedt, L. G., Miguel, R. N., Blundell, T. L., Vande Woude, G. F., Skoglund, U. & Svergun, D. I. (2006). Structural basis of hepatocyte growth factor/scatter factor and MET signalling. *Proc Natl Acad Sci U S A*, *103*, 4046-4051.
- Guinier, A. & Fournet, G. (1955). *Small Angle Scattering of X-Rays*, (Editor ed.). New York: Wiley.
- Guo, W. Y., West, J. M., Dutton, A. S., Tsuruta, H. & Kantrowitz, E. R. (2012). Trapping and structure determination of an intermediate in the allosteric transition of aspartate transcarbamoylase. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 7741-7746.
- Hilge, M., Aelen, J., Foarce, A., Perrakis, A. & Vuister, G. W. (2009). Ca2+ regulation in the Na+/Ca2+ exchanger features a dual electrostatic switch mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 14333-14338.
- Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W. & Tainer, J. A. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods*, *6*, 606-612.
- Jacques, D. A., Guss, J. M., Svergun, D. I. & Trehwella, J. (2012). Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution. *Acta Crystallographica Section D*, *68*, 620-626.
- Jehle, S., Rajagopal, P., Bardiaux, B., Markovic, S., Kuhne, R., Stout, J. R., Higman, V. A., Klevit, R. E., van Rossum, B. J. & Oschkinat, H. (2010). Solid-state NMR and SAXS studies provide a structural basis for the activation of alpha B-crystallin oligomers. *Nature Structural & Molecular Biology*, *17*, 1037-U1031.
- Konarev, P. V., Petoukhov, M. V. & Svergun, D. I. (2001). MASSHA - a graphic system for rigid body modelling of macromolecular complexes against solution scattering data. *J. Appl. Crystallogr.*, *34*, 527-532.

- Konarev, P. V., Petoukhov, M. V., Volkov, V. V. & Svergun, D. I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **39**, 277-286.
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. (2003). PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **36**, 1277-1282.
- Lawson, C. L. & Hanson, R. J. (1974). *Solving Least Squares Problems.*, (Editor ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Marino, M., Zou, P., Svergun, D., Garcia, P., Edlich, C., Simon, B., Wilmanns, M., Muhle-Goll, C. & Mayans, O. (2006). The Ig doublet Z1Z2: a model system for the hybrid analysis of conformational dynamics in Ig tandems from titin. *Structure*, **14**, 1437-1447.
- Marrott, N. L., Marshall, J. J., Svergun, D. I., Crennell, S. J., Hough, D. W., Danson, M. J. & van den Elsen, J. M. (2012). The catalytic core of an archaeal 2-oxoacid dehydrogenase multienzyme complex is a 42-mer protein assembly. *FEBS J*, **279**, 713-723.
- Mertens, H. D. & Svergun, D. I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol*, **172**, 128-141.
- Nicastro, G., Habeck, M., Masino, L., Svergun, D. I. & Pastore, A. (2006). Structure validation of the Josephin domain of ataxin-3: Conclusive evidence for an open conformation. *J Biomol NMR*, **36**, 267-277.
- Pernot, P., Theveneau, P., Giraud, T., Fernandes, R. N., Nurizzo, D., Spruce, D., Surr, J., McSweeney, S., Round, A., Felisaz, F., Foedinger, L., Gobbo, A., Huet, J., Villard, C. & Cipriani, F. (2010). New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *Journal of Physics: Conference Series*, **247**, 012009.
- Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, **45**, 342-350.
- Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *J. Appl. Cryst.*, **40**, s223-s228.
- Petoukhov, M. V., Monie, T. P., Allain, F. H., Matthews, S., Curry, S. & Svergun, D. I. (2006). Conformation of polypyrimidine tract binding protein in solution. *Structure*, **14**, 1021-1027.
- Petoukhov, M. V. & Svergun, D. I. (2005). Global rigid body modelling of macromolecular complexes against small-angle scattering data. *Biophys J*, **89**, 1237-1250.
- Pons, C., D'Abramo, M., Svergun, D. I., Orozco, M., Bernado, P. & Fernandez-Recio, J. (2010). Structural characterization of protein-protein complexes by integrating computational docking with small-angle scattering data. *J Mol Biol*, **403**, 217-230.
- Porod, G. (1982). General theory. In O. Glatter & O. Kratky (Eds.) *Small-angle X-ray scattering*. pp. 17-51). London: Academic Press.
- Roessle, M. W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrman, T., Konarev, P., Round, A., Fiedler, S., Hermes, C. & Svergun, D. (2007). Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *J. Appl. Cryst.*, **40**, s190-s194.
- Round, A. R., Franke, D., Moritz, S., Huchler, R., Fritsche, M., Malthan, D., Klaering, R., Svergun, D. I. & Roessle, M. (2008). Automated sample-changing robot for solution

- scattering experiments at the EMBL Hamburg SAXS station X33. *Journal of Applied Crystallography*, 41, 913-917.
- Stuhrmann, H. B. (1970). Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle-scattering function. *Acta Cryst.*, A26, 297-306.
- Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, 25, 495-503.
- Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J*, 76, 2879-2886.
- Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, 28, 768-773.
- Svergun, D. I. & Nierhaus, K. H. (2000). A map of protein-rRNA distribution in the 70 S Escherichia coli ribosome. *J Biol Chem*, 275, 14432-14439.
- Svergun, D. I., Petoukhov, M. V. & Koch, M. H. J. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys J*, 80, 2946-2953.
- Svergun, D. I., Petoukhov, M. V., Koch, M. H. J. & Koenig, S. (2000). Crystal versus solution structures of thiamine diphosphate-dependent enzymes. *J. Biol. Chem.*, 275, 297-302.
- Virtanen, J. J., Makowski, L., Sosnick, T. R. & Freed, K. F. (2011). Modeling the hydration layer around proteins: applications to small- and wide-angle x-ray scattering. *Biophys J*, 101, 2061-2069.
- Wriggers, W., Milligan, R. A. & McCammon, J. A. (1999). Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol*, 125, 185-195.
- Xu, X., Reinle, W., Hannemann, F., Konarev, P. V., Svergun, D. I., Bernhardt, R. & Ubbink, M. (2008). Dynamics in a pure encounter complex of two proteins studied by solution scattering and paramagnetic NMR spectroscopy. *J Am Chem Soc*, 130, 6395-6403.
- Zheng, W. & Doniach, S. (2002). Protein structure prediction constrained by solution X-ray scattering data and structural homology identification. *J Mol Biol*, 316, 173-187.

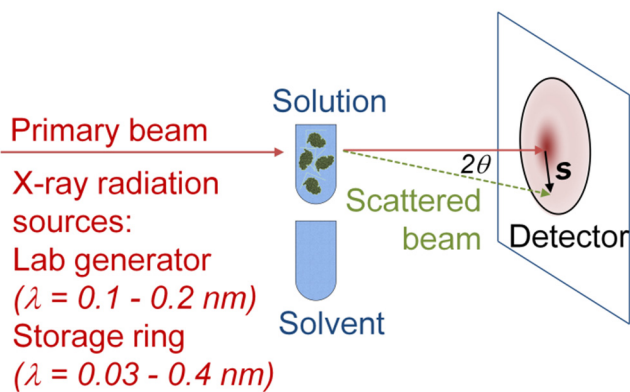


Fig 1

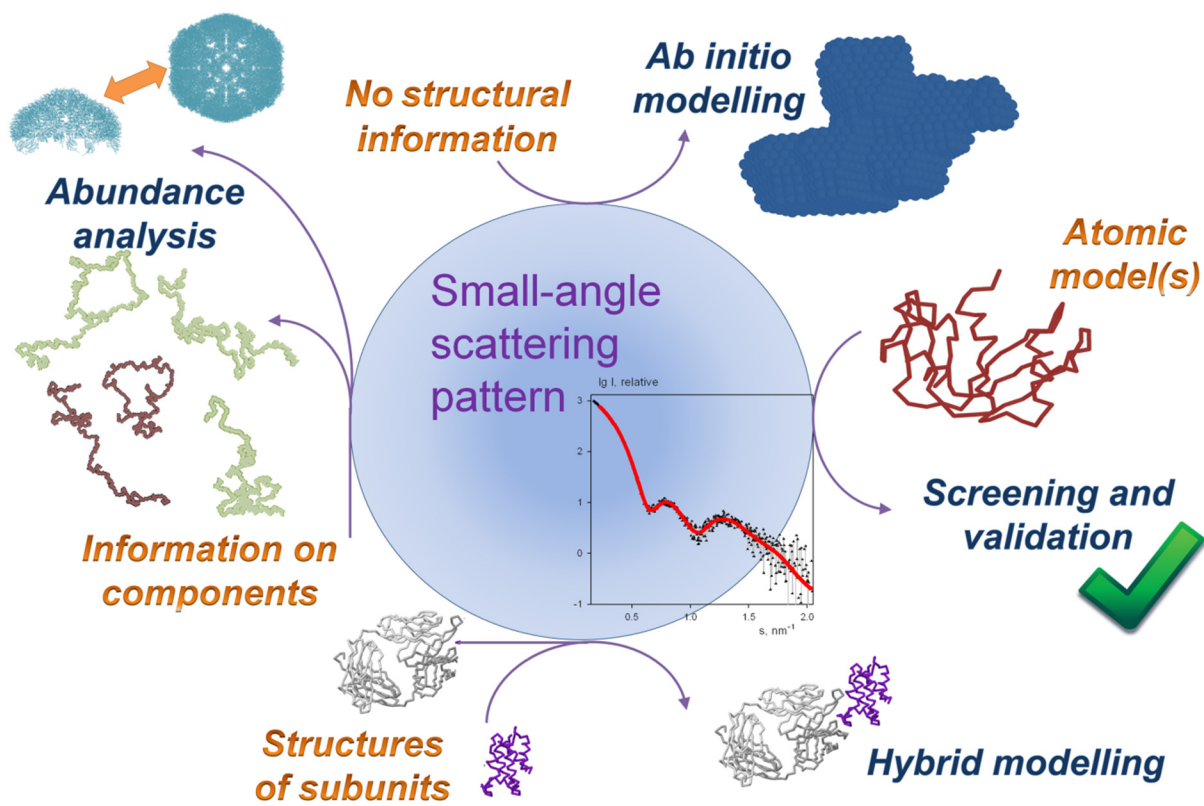


Fig 2

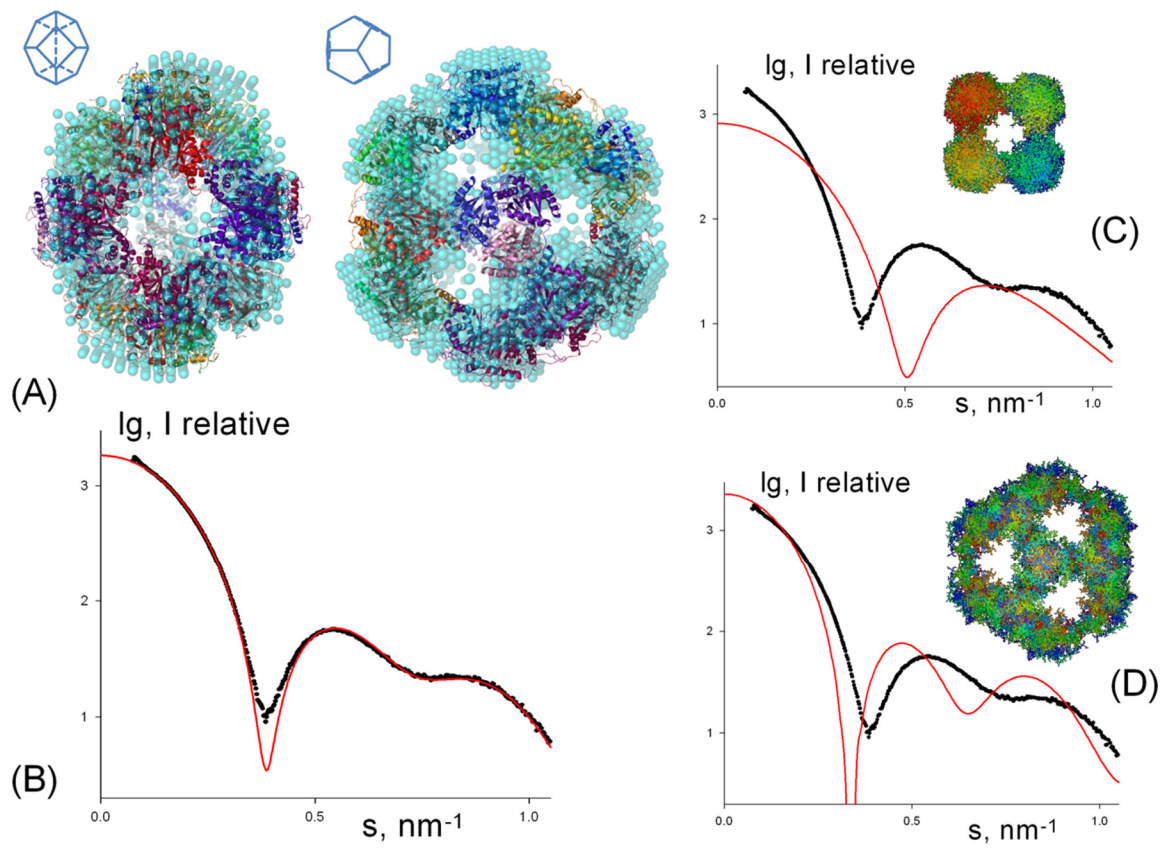


Fig 3

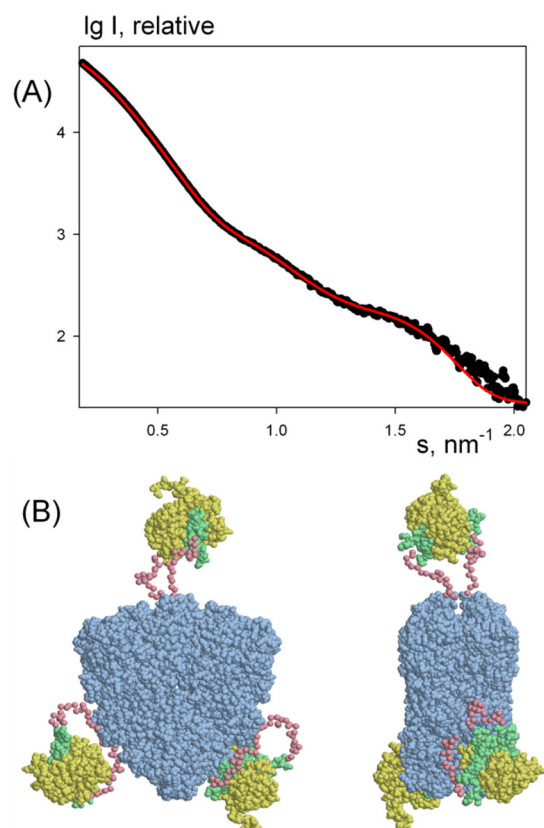


Fig 4