Reconstruction of quaternary structure from X-ray scattering by equilibrium mixtures of biological macromolecules

Funding Sources: This work was supported by the European Commission (the 7th Framework Programme) BioStruct-X project (contract number 283570), the BMBF research grant BIOSCAT (contract number 05K12YE1), the CNRS, INSERM, the Association pour la Recherche sur le Cancer (ARC), the Karo Bio Research Foundation, the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01 and INSTRUCT as part of the European Strategy Forum on Research Infrastructures (ESFRI). M.A.G is supported by a fellowship from the European Molecular Biology Laboratory Interdisciplinary Postdocs programme (EIPOD).

Maxim V. Petoukhov[1)]

Isabelle M.L. Billas[2)]

Maria Takacs[2)]

Melissa A. Graewert[1)]

Dino Moras[2)]

Dmitri I. Svergun[1)*]

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

[2)] Department of Integrative Structural Biology, Institute of Genetics and Molecular and Cellular Biology (IGBMC), Centre National de la Recherche Scientifique (CNRS), UMR 7104, Institut National de la Santé et de la Recherche Médicale (INSERM) U964, Université de Strasbourg (UdS), Illkirch, 67404, France

*Corresponding Author: Dmitri I. Svergun, European Molecular Biology Laboratory, Hamburg Outstation, Notkestrasse 85, 22607, Hamburg, Germany. Phone: +49 40 89902 125, Fax: +49 40 89902 149, e-mail: svergun@embl-hamburg.de.

Keywords: Solution scattering, molecular modeling, transient complex, oligomeric equilibrium, estrogen-related receptors

Abbreviations

SAXS, small-angle X-ray scattering; 3D, three-dimensional; ERR, estrogen-related receptor; DR, dummy residue; D_{max}, maximum size (diameter); R_g, radius of gyration; MX, macromolecular crystallography; NMR, nuclear magnetic resonance; SA, simulated annealing; FRET, fluorescence resonance energy transfer; HPLC, high-performance liquid chromatography; NSD, normalized spatial discrepancy; r.m.s.d., root mean square deviation; BSA, bovine serum albumin; SEC, size exclusion chromatography; TDA, triple detector array; RALS, right angle light scattering; Ri, refractive index; LBD, ligand binding domain; DBD, DNA binding domain; FL, full-length; GTS, glutamyl-tRNA synthetase; NGF, nerve growth factor; ER, estrogen receptor; IR, inverted repeat; RE, response element; bp, basepair.

Abstract: Recent renaissance of small-angle X-ray scattering (SAXS) made this technique a streamline tool for low-resolution structural characterization of biological macromolecules in solution. The major limitation of existing methods for reconstructing three-dimensional (3D) models from SAXS is imposed by the requirement the solute monodispersity. We present a novel approach that couples low resolution 3D SAXS reconstruction with composition analysis of mixtures. The approach is applicable to polydisperse and difficult to purify systems, including weakly associated oligomers and transient complexes. *Ab initio* shape analysis is possible for symmetric homo-oligomers, while rigid body modeling is applied also to dissociating complexes when atomic structures of the individual subunits are available. In both approaches, the sample is considered as an equilibrium mixture of intact complexes/oligomers with their dissociation products or free subunits. The algorithms provide the 3D low resolution model (for *ab initio* modeling, also the shape of the monomer) and the volume fractions of the bound and free state(s). Simultaneous fitting of multiple scattering data sets collected at different conditions allows one to further restrain the modeling. The possibilities of the approach are illustrated in simulated and experimental SAXS data from protein oligomers and multisubunit complexes including nucleoproteins. Using this approach, new structural insights are provided in the association behavior and conformations of estrogen-related receptors ERRα and ERRγ. The possibility of 3D modeling from the scattering by mixtures significantly widens the range of applicability of SAXS and opens novel avenues in the analysis of oligomeric mixtures and assembly/dissociation processes.

Introduction

Small-angle X-ray scattering (SAXS) is an established tool in structural molecular biology providing valuable information about overall structure and conformational changes of individual proteins and functional complexes in solution (*1, 2*). SAXS also enables quantitative characterization of flexible macromolecules and is suitable for studies of processes e.g. assembly/dissociation. Thanks to the progress in instrumentation and to the development of novel data analysis approaches enhancing the reliability of reconstructions (*3*), this technique is increasingly demanded by the biological community.

The methods to reconstruct three-dimensional (3D) models from the solution SAXS data utilize the concept of a dilute monodisperse solution containing non-interacting identical chaotically oriented particles. In this case, the measured scattering intensity is proportional to that from a single particle averaged over all orientations (*4*). It was demonstrated (*5-7*), that a low resolution particle shape can be reconstructed *ab initio* from the scattering data. The *ab initio* methods representing the shape as a collection of interconnected beads on a hexagonal grid are now routinely employed allowing one to reconstruct low resolution models within minutes (e.g. (*8*)). For proteins, an alternative reconstruction approach is available utilizing a dummy residues (DR) model. Here, each amino acid is represented by a DR with an effective average scattering form factor. *Ab initio* program GASBOR (*9*) performs a Monte-Carlo based search of the spatial distribution of DRs inside a spherical search volume with diameter D_{max} equal to the maximum particle size. The resulting DR assembly is required to be compatible with a typical distribution of backbone C_α atoms in proteins and the scattering computed from this assembly should fit the experimental data. In contrast to the bead modelling, the DR approach requires some a priori information (the number of residues has to be specified) but provides better resolution models.

Many *ab initio* approaches allow one to take particle symmetry into account, and this is often useful for the reconstruction of oligomeric proteins.

SAXS has also become an integral part of hybrid approaches, where the technique is employed together with high resolution methods such as macromolecular crystallography (MX) or NMR. For macromolecular assemblies consisting of several subunits, rigid body modelling can be applied if high resolution models of individual subunits are available. Assuming that these do not change significantly upon binding, the quaternary structure of the entire complex can be described by mutual arrangement of the subunits. Several rigid body modelling algorithms are available to model various types of macromolecular systems from multidomain proteins to homo-oligomeric assemblies to multimeric complexes (*3, 10*). In many cases, these algorithms utilize restrains provided by other techniques, e.g. NMR (*11, 12*).

Among the rigid body modelling programs, SASREF (*10*) is one of the most general algorithms employed to perform a global search of optimal positions and orientations of individual fragments in an assembly using simulated annealing (SA). Similarly to *ab initio* approaches from ATSAS package (*7-9*), SASREF allows one to utilize point symmetry of the assembly so that the equivalent subunits are relocated in symmetric manner during the modelling. Various addition restraints (e.g. from NMR chemical shifts and residual dipolar couplings, FRET distances and interfaces proved by mutagenesis) can be incorporated into the reconstruction procedure, so that the resulting model not only fits the experimental SAXS profile but also is compatible with the data from complementary methods.

All the existing methods require sample monodispersity for a reliable 3D model building from solution scattering data. In general, monodispersity can be achieved by proper preparation and handling of the sample prior to the scattering experiment or by using on-line HPLC purification (*13*) coupled with the SAXS measurement. There are, however, numerous cases where sample monodispersity cannot be achieved despite valiant efforts, and this makes the structural

interpretation rather difficult. Typical examples of such polydisperse systems are oligomeric equilibrium mixtures and low affinity transient complexes. In the former case, higher order oligomers are in dynamic association/dissociation leading to an equilibrium with monomers or other lower oligomeric states (e.g. tetramer-dimer equilibrium). For transient complexes, the components may permanently undergo transitions between bound and free states; moreover, an excess of one of the components is often required for the complex formation. In all these cases, the mixtures may rapidly embark into the dynamic exchange immediately after size-exclusion chromatography such that even on-line purification is not feasible. Here, we propose a general approach applicable to such notoriously polydisperse systems.

The scattering profile from mixtures without interparticle interactions is a linear combination of the scattering intensities of individual components $I_k(s)$, weighted by their volume fractions v_k (*14*):

$$I(s) = \sum_k v_k I_k(s) . \quad (1)$$

Here, $I_k(s)$ are normalized by the concentration, i.e. proportional to the molecular weight of the component. The values of v_k can be directly computed from the SAXS data, using a simple least squares fitting (*15*) if the scattering patterns of all the components are available (or if their models are known). If both, v_k and $I_k(s)$, are not known, meaningful analysis becomes difficult. Apart from the trivial cases where 3D modelling is performed solely for the dominating component while neglecting the presence of the minor species (which inevitably leads to systematic errors and possible misinterpretations), only few approaches are available dealing with polydispersity.

For flexible systems (e.g. multidomain proteins with potentially disordered linkers) ensemble fitting techniques can be applied (*16-19*). These approaches allowing for co-existence of multiple conformers in solution are able to describe the degree of flexibility in a system and in some cases

also to depict more populated conformations. For oligomerization processes, scattering patterns from intermediates may be extracted, as demonstrated e.g. in the study of insulin fibrillation (*20*). In a recent approach (*21*) the oligomerization data was treated using a multivariate curve resolution with alternating least squares. In the two latter cases, multiple scattering curves recorded at different conditions were used to extract the contribution from individual species.

While the existing methods are useful in the analysis of the scattering data from polydisperse systems, they are usually not able to directly provide structural models. Here, a novel general methodology to analyze SAXS data from equilibrium mixtures is presented combining 3D modelling with the analysis of oligomeric composition. We describe the methods extending the capabilities of *ab initio* and rigid body modelling, validate them on simulated data sets and provide examples of practical application.

Theory and methods

General idea of the method, its possibilities and limitations

Instead of trying to extract the scattering intensities of the individual components, the method primarily aims at the construction of the largest species (*i.e.* of the complete oligomer or of the entire complex). While fitting the experimental data by this largest assembly, the algorithms also take the presence of dissociation products into account, *e.g.* individual subunits (monomers) or (in case of a complex) subcomplexes consisting of fewer number of subunits. In the latter case the same arrangement of subunits is assumed in the partial constructs as found in the entire complex. For a given configuration of the entire assembly, the scattering intensity and those of the dissociation products are evaluated. The fit to the experimental data is computed as a linear combination of these intensities, and the volume fractions of the components are obtained by linear least squares minimization using Eq. (1).

The types of the oligomeric systems, which can be studied are schematically depicted in Fig. 1. In case of oligomeric mixtures, dissociation into a single type of lower molecular weight particles is allowed (Fig. 1A). These are typically monomers but dissociation into other lower multimers is also possible (e.g. in case of tetramer-dimer equilibrium). However, dissociation intermediates of the oligomer (Fig. 1B) are assumed to be absent. For the rigid body modelling of the complexes, excess of one of the components (Fig. 1C) or formation of a partial subcomplex (Fig. 1D) are supported as well as full dissociation into individual subunits (Fig. 1E) but not the presence of several partial constructs in one sample (Fig. 1F). Given that (i) the dissociation scenario is selected by the User prior to the modeling, (ii) the algorithms do not attempt to extract the scattering intensities of individual components of the mixture and (iii) the structures of dissociation products are defined by the structure of the intact assembly, the presented approach has just a few additional parameters (the volume fractions of the components) as compared to an equivalent monodisperse case and the level of ambiguity due to polydispersity is not increased significantly.

Both for the *ab initio* and the rigid body cases, simultaneous fitting of multiple data sets is possible to monitor the changes of the volume fractions of the components with time, with the solute concentration or by changing other environmental parameters. This option not only allows for more reliable models to be obtained, but is also useful in the studies of allosteric regulation, medium based structural transitions and other biologically relevant processes.

Accounting for polydispersity by ab initio 3D reconstructions

The DR approach implemented in the modelling program GASBOR (*9*) can be extended to *ab inito* shape reconstruction of oligomeric equilibrium mixtures, where a monomer represents the dissociation product of a symmetric multimer. Here, the shape of the monomer, represented as a gas of DRs, is considered to be an asymmetric part of the entire multimer. The multimer is then constructed using the symmetry group defined by the oligomeric state, e.g. P2 for a dimer, P3 for

a trimer *etc*. In DR gas condensation procedure, the scattering intensities of the multimer and of the monomer are computed from the appropriate sets of DRs and the resulting profile from the mixture is computed as their linear combination following Eq. (1). The coefficients of the two scattering patterns (corresponding to the volume fractions of the two components) are in the general case defined by the least square fitting of the experimental curve. If, however, the volume fraction of the monomer is known *a priori* (*e.g.* from the analysis of the forward scattering intensity), the value can be fixed in the course of modelling. The algorithm called GASBORMX utilizes the principle of the original GASBOR method, but there are also significant differences. In the case of GASBORMX, interconnectivity of the reconstructed model is required not only for the entire DR assembly (as in GASBOR) but also for the DR portion representing the monomer. This ensures that the latter is an interconnected entity. Special care has to be taken to properly determine the maximum size of the solute D_{max} in the case where the monomer is the dominating species, while keeping in mind that the search volume has to accommodate the entire multimer. Obviously, the results provided by GASBORMX critically depend on the proper choice of the degree of oligomerisation (and hence the symmetry of the multimer). The input for GASBORMX consists of the regularized scattering intensity generated by GNOM (*22*) (or a set of GNOM files from concentration series), the number of amino acids in the monomer and the symmetry of the multimer. Schematic comparison of GASBOR and GASBORMX algorithms is presented in Supplementary Fig. S1.

Rigid body modelling of transient complexes and weak oligomers

When conducting rigid body modelling, polydispersity can be taken into account in a more general way than for the *ab initio* case. For the rigid body analysis, arbitrary subsets of subunits (dissociation products) can be selected as additional components of the mixture (Fig. 1E-1F). During the SA-driven modification of the subunit arrangement, the experimental scattering curve composed of the intensities computed for the entire multisubunit complex and for the specified

subcomplex(es). As an example, if a ternary complex is able to partially dissociate into a binary part and a free subunit, the experimental data can be fitted by a combination of the three scattering profiles. The volume fractions of the bound and the dissociated states in the mixture are determined by linear least square fitting. In contrast to the *ab initio* case, the atomic models of the monomers are known and do not change in the course of rigid body reconstruction and thus, the option of volume fraction fixation is not provided. The equilibrium rigid body modelling implemented in the program SASREFMX is generally more robust and reliable than the equilibrium *ab initio* analysis, since the complexed and free states are made of the building blocks (subunits) with predefined structures, (i.e. the modelling bears more information). Furthermore, in SASREFMX there is no limitation on the search volume as it is the case with GASBORMX. The results are thus not biased by the overall size of the complex a value which may be not straightforward to estimate if the complex is present in small quantities in the mixture. The input to SASREFMX consists of the sets of scattering intensities at different conditions (can be just a single profile), the set of PDB files of rigid bodies and a table describing the relationship between the subunits and the scattering profiles (i.e. if the subunit is present in the given construct) as well as the dissociation products. That means that the modeling scenario (as in Fig. 1) is predefined by the User. Schematic comparison of SASREF and SASREFMX algorithms is presented in Supplementary Fig. S2.

Testing ab initio and rigid body modelling reconstructions on simulated mixtures

The proposed approaches have been validated in numerous test examples, and some of them are presented below. In all cases, simulated scattering profiles from mixtures obtained by the averaging of the scattering intensities of the components computed by CRYSOL (*23*) with default parameters. To emulate the experimental conditions, the intensities were randomized to yield 3% relative noise. All reconstructions were made multiple (typically, ten) times and the most typical models were selected using DAMAVER (*24*) taking the symmetry into account

where applicable. For the GASBORMX runs, comparisons between the target and reconstructed models were done using SUPCOMB (*25*). This program aligns two arbitrary low or high resolution models represented by ensembles of points by minimizing a dissimilarity measure called normalized spatial discrepancy (NSD). NSD values around or below 1.0 indicate that the models are similar at low resolution, for NSD > 2.0 the models are significantly different. For SASREFMX reconstructions, r.m.s.d. between the initial and reconstructed models was also calculated.

Sample preparation

Purification and molecular mass measurement of bovine serum albumin (BSA) with SEC-TDA. For separation of BSA (Sigma, A7906) in its monomeric and higher oligomeric components, the sample was passed through a Superdex 200 column (10/300, GE Healthcare) connected to a triple detector array (TDA) model 305 (Viscotek Ltd., Houston, TX) which consisted of a static light scattering cell with a photodiode dectector at 90° for right angle light scattering (RALS), a deflection refractometer (Ri), as well as a photometer. The use of TDA allows molecular weight determination of the monomeric and dimeric components as well as their weight fractions.

Buffer A (50 mM Hepes, 150 mM NaCl, ph 7.5) was used as the mobile phase and the sample (20 mg/ml in Buffer A) was filtered through 0.2μm filters before injection (100 μl). Size exclusion chromatography (SEC) was performed at a flow rate of 0.5 ml/min and at room temperature. All data was acquired using the Omnisec software (Viscotek Ltd., Houston, TX). Ribonuclease (GE Healthcare) was used for TDA internal constants calibration. The refractive index increments, dn/dc, was set to 0.185. The RALS data in combination with the concentration as determined with the deflection refractometer provided an estimation of the molecular mass.

Eluting protein fractions were collected manually and the concentrations of the individual samples were obtained by absorption at 280 nm using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and an extinction coefficient of

6.9 (1% BSA solution). In addition, a BSA sample was prepared for measurements, which was solely centrifuged (Beckman-Coulter Allegra X-22R, 14.000 rpm, 10 minutes, 4°C) to remove large aggregates but still contained dimeric compounds.

Cloning, protein expression and purification of ERR samples

Full lengths ERRα (2-422), A/B domain truncated ERRα (70-422) and ERRγ (120-458) were cloned into pDEST8 vector containing a C-terminal hexahistidine tag. The recombinant proteins were expressed in Sf9 or Sf21 insect cells using the baculovirus technology. ERRα LBD (189-423) and ERRγ LBD (229-458) were cloned in pET24a and pET15b, respectively, and were expressed in *E. coli* BL21 (DE3) strain using Luria Bertani (ERRγ) and terrific broth (ERRα) media.

For purification, harvested cells were re-suspended in lysis buffer (20 mM Tris-HCl pH 8, 400 mM NaCl, 10% (v/v) glycerol, 2 mM CHAPS, 5 mM imidazole, cOmplete, EDTA-free protease inhibitor cocktail tablet (Roche Applied Science)), sonicated and centrifuged. The supernatant was loaded on 5 ml HisTrap FF crude column (GE Healthcare). The proteins were eluted at 250 mM imidazole and further purified by SEC on Superdex S75 (16/60 and 10/300, GE Healthcare) and Superdex S200 (16/60 and 10/300, GE Healthcare) columns. The His_6-tag of ERRα LBD and ERRγ LBD was removed using thrombin digestion, followed by a further SEC step.

Complexes of FL or ΔA/B ERR and response elements were formed by adding a 1.3 fold molar excess of DNA to the receptor dimer followed by an additional gel filtration step in case of direct sample injection. For HPLC coupled SAXS cell measurements the samples were concentrated to 4-19 mg/ml and injected into a SEC column (SHODEX KW402.5). The composition of the final buffer was 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 100 mM KCl, 5 mM $MgCl_2$, 1% (v/v) glycerol, 1 mM CHAPS for the ERR α samples and 20 mM Tris-HCl pH 8.0, 60 mM NaCl, 60 mM KCl, 4 mM $MgCl_2$, 1% (v/v) glycerol, 1 mM CHAPS for the ERRγ samples. TCEP was

added to the purified proteins at a 10 mM final concentration. Protein samples were concentrated using Amicon-Ultra centrifugal filter units (Millipore). Purity and homogeneity of the proteins were assessed by SDS-PAGE and complex formation was monitored by native PAGE.

SAXS experiments

Solution scattering experiments were performed at the X33 beamline (EMBL@ DESY, Hamburg) (*26*) and at the SWING beamline at the SOLEIL Synchrotron (Gif-sur-Yvette, France). REF SAXS data from BSA and ERRα samples at X33 beamline were collected using PILATUS detector and automated filling (*27*). Data acquisition at X33 beamline was performed at a sample-detector distance of 2.7 m, covering the range of momentum transfer $0.01 < q < 0.6$ $Å^{-1}$ ($q = 4\pi \sin(\theta)/\lambda$ where 2θ is the scattering angle and $\lambda = 0.15$ nm is the X-ray wavelength) in eight frames (15 seconds each) to check for possible radiation damage. All scattering measurements were carried out at 10°C and the data were processed using standard procedures (*28*). SAXS data for ERRα and ERRγ were also collected at the SWING beamline, using a 17 x 17 cm^2 low-noise Aviex CCD detector positioned at a distance of 1,82 m from the sample. Sample solutions were circulated in a thermostated Quartz capillary with a diameter of 1.5 mm and 10 µm wall thickness, positioned within a vacuum chamber. About 20-40 µl protein were injected at relatively high concentration (up to 19 mg/ml) into a SEC Column (Agilent Bio SEC-3), using an Agilent© HPLC system and eluted directly into the SAXS flow-through capillary cell at a flow rate of 150 µl/min (*13*). SAXS data were collected online throughout the whole elution time, with 1.5 s frame duration and 0.5 s dead time between frames. In total 250 frames were collected, normalized to transmitted intensity. The frames corresponding to the elution peak of the complex were averaged using “Foxtrot”, a dedicated home-made application.

Results and discussion

To demonstrate the capabilities but also to indicate the limitations of the proposed approaches, the methods were first extensively tested on simulated data. After these tests, GASBORMX and SASREFMX have been applied to structural analysis of experimental SAXS profiles collected from polydisperse systems at the EMBL X33 beamline.

Artificial lysozyme dimer – The first simulated example is given by an artificial lysozyme dimer. The most tight dimeric interface in the crystallographic model of lysozyme (6lyz, (*29*)) according to the PISA server (*30*) has the area of 548.3 A^2, whereby the two monomers are related by a twofold symmetry axis (Fig. 2A). A simulated scattering profile from an equimolar mixture of monomers and such dimers (volume fractions of 33% and 67%, respectively) was generated and *ab initio* and rigid body models were reconstructed for a monomer-dimer equilibrium case (the corresponding overall parameters and distance distribution p(r) function are presented in Supplementary Table 1 and Supplementary Fig. S3). In multiple runs of GASBORMX, the dimer was always correctly reconstructed. A typical model displayed in Fig. 2A yields the fit with discrepancy χ=1.25 and NSD of 0.93 to the initial dimer The volume fraction of the monomer is found to be 39%. SASREFMX was capable to find the solution very close to the correct one (Fig. 2B) with the *r.m.s.d.* of the reconstructed dimer to the original coordinates of 1.9 Å, whereby the corresponding volume fraction of the monomer was 35%. Not surprisingly, the reconstruction yields a very good fit to the simulated data with χ=1.15 (Fig. 2C). Similar results were obtained with both approaches in the range of the monomer volume fractions between 20 and 80 %.

Hexamer - monomer equilibrium of RuvB – The methods were also tested for oligomers with higher symmetry and for lower volume fractions of either of the components. As a test example

for these conditions, a single hexameric ring of RuvB ATPase recently studied by SAXS (*31*) has been selected. The crystal structure of the hexamer (pdb code 2XSZ) is shown in Fig. 3A and the simulated profiles from hexamer-monomer mixtures with very low (15%) and very high (85%) monomeric components are presented in Fig. 3B (the corresponding overall parameters and p(r) function are presented in Supplementary Table 1 and Supplementary Fig. S3). For both scenarios reconstructions made by GASBORMX and SASREFMX represent the overall architecture of the entire hexamer rather well (Fig. 3, C-F) with typical NSD value around 1.1. The obtained volume fractions were close to the expected ones with 5%-10% excess for the monomer in some cases and the resulting intensities always neatly fitted the simulated target data (Fig. 3B). The asymmetric parts of the *ab initio* models (Fig. 3 C,E) provide very good agreement with the shape of the monomer (NSD below 1.0).

Simulated and experimental data from mixtures of glutamyl-tRNA synthetase – To explore the capabilities of the approaches in extreme conditions, GASBORMX and SASREFMX were applied to a protein with a very anisometric shape of the monomer (aspect ratio ~ 3:1). This rather complicated example concerns a dimer of glutamyl-tRNA synthetase (GTS) (PDB code 1G59), with a calculated interface of 754 A^2 according to PISA. The profile from an equimolar mixture of monomers and dimers was evaluated in the same way as for lysozyme. Using GASBORMX it was possible to restore the low resolution model of the GTS dimer very well from the simulated scattering by an equimolar mixture of monomers and dimers (NSD about 1.3). However, in none of the modelling attempts (even when fixing the correct volume fractions), was the monomer shape correctly reconstructed; the program always divided the dimer into the two compact rather than extended monomers. A typical reconstruction in Fig. 4A presents such a separation of the DRs into two compact monomers within the correct dimeric shape. This artifact can be avoided using *ab initio* modelling with simultaneous fitting of multiple scattering profiles including those with high monomer content. As an example, fitting

two curves with 15% and 90% monomeric content provides correct reconstruction of both monomers and dimers. The typical NSD values obtained for monomer and dimer in this case were 1.16 and 1.37, respectively (comparison of the DR model with the GTS crystal structure is presented in Supplementary Fig. S4).

With SASREFMX, the modelling was done against the simulated curve, as well as against the available experimental data from a polydisperse GTS sample (*32*). In the latter case, the experimental curve was used whose corresponding experimental conditions yielded volume fractions of monomers and dimers of about 30% and 70%, respectively (i.e. close to the ones in the simulated profile). For both simulated and experimental data fitting, the overall shape of the dimer and the position of the monomer were reconstructed (Fig. 4B) and the accuracy of the volume fraction estimate was about 10%. The orientation of the monomer around the long axis was however not sufficiently defined, and the monomer was able to rotate this axis in individual runs. It should be mentioned that such ambiguity may even be observed when modelling the dimeric structure in the monodisperse case by SASREF as such rotations of the monomers have little influence on the overall appearance of the dimer. Typically, additional information allows one to remove the ambiguity, and, indeed, SASREFMX reconstructions using contacts conditions (in terms of a 7A distance restraint between one pair of residues, GLU287 and SER323) leads to an unambiguous reconstruction (Fig. 4C). The use of the restraint yielded models with the overall *r.m.s.d.* between the rigid body model and the target structure to within 8A, and a typical fit to the experimental profile with χ=1.16 is shown in Fig. 4D.

BSA solution as a case study of monomer-dimer equilibrium. This example describes the application of GASBORMX and SASREFMX to the analysis of dissolved BSA, a protein frequently used as a standard calibration sample in SAXS studies. BSA powders are cheap and readily available from various suppliers, but the protein has a known propensity to form dimers in solution. Fig. 5A shows the comparison of the experimental profiles of freshly made samples

obtained from the same batch, where one sample was purified on a size-exclusion chromatography (SEC) column immediately before the data were taken (Supplementary Fig. S5), while the other one was just centrifuged after suspension. These intensities appear similar but the SEC purified BSA yields somewhat lower R_g value as can be seen from the deviation between the two curves at very low angles (Fig. 5A, inset). As a consequence, the scattering computed from the crystallographic BSA monomer (PDB code 3V03, (*33*)) fits well the SEC purified sample data but not those of resuspended and centrifuged sample (Fig. 5A). This observation suggests that a minor fraction of the protein may exist in dimeric state. The experimental scattering intensity of the resuspended BSA was employed for *ab initio* and rigid body modelling reconstructions of BSA dimer assuming monomer-dimer equilibrium in the mixture (the overall parameters and the p(r) function are presented in Supplementary Table 1 and Supplementary Fig. S3). The typical models of the dimeric BSA reconstructed by GASBORMX and SASREFMX using P2 symmetry (shown in Fig. 5B,C and Supplementary Fig. S6) allow one to neatly fit the SAXS data from the mixture with χ=1.13 and 0.98, respectively (Fig. 5A). The volume fraction of the dimer was about 20% for both approaches, and the monomer (i.e. the asymmetric part) of the *ab initio* dimer agreed well with the crystallographic BSA monomer. The experimentally determined volume fractions obtained from the analysis of the elution profiles on a size exclusion column using TDA are 83% for the momomeric and 17% for the dimeric components, (see Supplementary Table 2) and, thus in good agreement with the SAXS derived values. Interestingly, the *ab initio* and rigid body models displayed the overall quaternary structures similar to the dimers within the crystal structure of BSA, as suggested by PISA server (*30*).

Partially dissociated αD11 Fab–NGF complex. Human nerve growth factor (NGF) with bound αD11 Fab fragments serves as an example of rigid body modeling against polydisperse data from partially dissociated complex. The system has been previously examined by SAXS in combination with computational docking (*34*). That study revealed that an NGF dimer can bind

up to two αD11 monoclonal antibodies, but the complex is not fully associated and there are dissociation products in solution. A computational docking model of the 2:1 Fab–NGF complex based on epitope mapping was validated by experimental scattering data. Several possible complexes were tested and the fitting was performed by equilibrium mixtures of the intact and partially dissociated complexs. We employed SASREFMX to build independently a rigid body model of 2:1 Fab–NGF from the polydisperse SAXS profile (Fig. 6A). In SASREFMX modeling, the long axis of the NGF dimer has been aligned with Z and the molecule was fixed in this orientation, whereas the two αD11 Fab fragments were moving and rotating in a symmetric way (two-fold axis coincided with Z). The typical rigid body reconstruction (Fig. 6B) obtained by SASREFMX without any contact restraints agrees well (NSD of 1.3) with the best model from the computational docking studies reported in [REF] . The volume fraction of the intact Fab-NGF assembly in the individual runs of SASREFMX was about 50% neatly fitting the polydisperse SAXS data (Fig. 6A); the typical rigid body reconstruction yields the χ value of 1.33. This example demonstrates the capability of SASREFMX approach to model not only homo-oligomers but also the assemblies made of distinct subunits.

Study of ERRα - IR3 DNA complex partially assembled in higher order multimers. Nuclear Receptors (NRs) are transcription factors involved in the regulation of crucial biological processes, such as cell growth, differentiation and energy homeostasis. NRs consists of functional modules that include a strongly variable N-terminal A/B domain, a highly conserved DNA binding domain (DBD) which is connected to the C-terminal ligand-binding domain (LBD) through the hinge region (*35*). NRs act as transcriptional regulators by binding to response elements (REs) present in the enhancer or in the promoter regions of their target genes, by recruiting transcriptional regulators with various enzymatic activities, thus allowing the opening of chromatin and the initiation of transcription. The transcriptional regulation of the mitochondrial and metabolic functions is highly dependent on the orphan estrogen-related

receptors (ERRα and ERRγ).The ERRs were originally identified based on the sequence homology of their DBD with that of the estrogen receptors (ERs). Like ERs, the ERRs behave as homodimers and recognize inverted repeat (IR3) REs, composed of two half-sites of 6 bp each arranged in a tail-to-tail manner and separated by 3 bp (IR3) (see Supplementary Fig. S7) (*36*). However, this type of REs represent only a small fraction of the natural ERR binding sites, which are extended half-site REs composed of a single 6 bp half-site DNA sequence preceded at its 5'-end by a 3 bp extension. In the study presented here, we will only focused on the ERR receptors bound to IR3 REs.

In order to facilitate biophysical and structural studies of the ERRs, ΔA/B constructs were first considered where the flexible A/B domain of the receptor was truncated (ERRΔA/B). Solution studies of ERRγΔA/B bound to DNA indicated a 2:1 stoichiometry, whereby a ERRγ homodimer is bound to one double-stranded DNA bearing the IR3 RE (Takacs et al, in preparation). Solution studies of ERRΔA/B-IR3 complexes indicated significant differences between the solution scattering profiles of ERRαΔA/B-IR3 and ERRγΔA/B-IR3 (Supplementary Fig. S8). Similarly, the scattering profiles of the isolated LBD of ERRα and ERRγ differ from each other (Supplementary Fig. S8). This observation is unexpected, since the corresponding homodimeric LBD crystal structures are identical at the (low) resolution of SAXS measurements (*37-39*). A hint to resolve this issue is given by the crystal structure of ERRα LBD bound to a synthetic ligand (PDB code 2PJL). A peculiar crystal packing was reported (*40*) where three ERRα LBD homodimers are arranged around a 3-fold symmetry axis forming a hexameric arrangement. This observation led us to hypothesize that in solution a fraction of ERRα LBD might form hexameric assemblies. The scattering profile of a mixture containing about 80% dimers and 20% hexamers (the latter as seen in the crystal packing) does indeed result in a rather good fit to the data (Supplementary Fig. S9). To support our data, additional experiments were performed at the

SWING beamline at SOLEIL synchrotron, where the SAXS measurements are carried out on-line on a sample eluted from a size exclusion chromatography (SEC) column (*13*). In the case of ERRα LBD, two SEC elution peaks were observed. The scattering profile measured for the peak of the largest eluted molecules was well fitted by the scattering computed for the crystallographic hexamer, while the scattering profile corresponding to the smaller species agreed well with the scattering curve calculated for the dimer (Supplementary Fig. S9). Given that the LBDs of ERRα, but not of ERRγ, can form hexamers suggests that ERRαΔA/B bound to DNA may behave similarly with a small fraction of ERR-IR3 molecules having a 6:3 stoichiometry in addition to those possessing the expected 2:1 stoichiometry. SASREFMX has therefore been applied to the ERRαΔA/B-IR3 RE sample in order to build a model for the 6:3 complex from the experimental SAXS data of the potentially polydisperse ERRαΔA/B-IR3 sample. Given the concentration effect observed for this complex, several SAXS profiles recorded at different concentration values *c* ranging from 1 mg/ml to 3mg/ml were simultaneously fitted (Fig. 7A). In the modelling, it was assumed that the solute is a mixture of hexamer- and dimer-based complexes with the concentration-dependent volume fractions. Additional constraints were applied to the system, namely P3 symmetry was imposed and the structure of the hexameric LBD part was fixed. Three copies of DBD dimers with bound DNA were then moved and rotated in a symmetric manner to adjust their arrangement with respect to the LBD hexamer. The connectivity of the individual ERRα polypeptide chains was ensured by placing tentative linkers (generated at random), bridging the appropriate termini of the corresponding DBDs and LBDs using the proximity requirement. A typical model reconstructed by SASREFMX (Fig. 7B) indicates a flat shape of the 6:3 assembly, with the three DBD-DNA moieties located on the periphery close to the plane of the LBD hexamer. The scattering profiles of ERRαΔA/B-IR3 are neatly fitted for the entire range of concentrations (Fig. 7A, Table 1), whereby the volume

fraction of the 6:3 assembly reaches 20% for the largest concentration, a value which is consistent to that of the isolated ERRα LBD system. It is interesting to note that the structure of the ERRαΔA/B-IR3 2:1 subcomplex, extracted from the 6:3 model is highly similar to that calculated independently for ERRγΔA/B-IR3 (with NSD = 1.2). Moreover, the 6:3 rigid body model of ERRαΔA/B-IR3 can be docked into the *ab initio* envelope of the full-length ERRα in complex with IR3 RE calculated by DAMMIF (*8*) (Supplementary Fig. S10). In this case, the scattering data collected using the HPLC-coupled SAXS cell (SWING beamline, SOLEIL synchrotron) corresponded to a monodisperse population of complexes. This cross-validation lends further support to the results obtained by SASREFMX.

Conclusions

We presented here a novel approach combining structure modelling with oligomeric composition analysis. *Ab initio* (GASBORMX) and rigid body modelling (SASREFMX) methods provide structural models from the scattering data by oligomeric mixtures and partially dissociating complexes. As demonstrated in the simulated and practical examples, these methods allow one to meaningfully analyze the scattering data and in most cases reconstruct 3D models. These examples also revealed limitations of the approach, especially in cases when the volume fraction(s) of some of the components are small.

The test calculations indicated that GASBORMX could be employed for the mixtures with the volume fraction of monomers ranging from about 15% to about 85%, independently whether the volume fractions were fixed at the target values during the reconstruction or not. The lower limit arises simply due to too small contribution of the monomeric intensity at the lower volume fractions of the monomer (note that the intensity of a given component in Eq. 1 is proportional to

its molecular weight). The upper limit is explained by the need to estimate the maximum size of the oligomer which is rather problematic at higher monomeric content. SASREFMX appears to be applicable for somewhat lower affinity complexes with the volume fraction of the dissociated species within 90%-15%.

The accuracy of the estimated volume fractions for all the tests was typically within about 10-15 %, but, interestingly, the observed deviations in the volume fractions had little effect on the 3D structure restorations. The glutamyl-tRNA synthetase example revealed the limitation of the *ab initio* modelling for highly extended monomers in polydisperse scenario. The *ab initio* reconstruction tends to provide more compact shapes, and, although the shape of the entire oligomer is reconstructed correctly, it may be divided into compact monomers. Fitting of multiple scattering data sets, including those with higher content of monomeric species, helps to overcome this problem. In any case, if multiple data sets corresponding to different equilibrium points are available, simultaneous fitting provides more information and improved the fidelity of the models.

The presented approach significantly extends the possibility of modelling of oligomeric mixtures and we expect that it will find broad applications in the studies of notoriously polydisperse systems. One should however not forget that accounting for polydispersity further adds to the intrinsic ambiguity of the SAS data interpretation in terms of 3D models. The above methods must therefore be never utilized as an alternative to the purification procedures to obtain monodisperse solutions, for which the 3D analysis is in any case more reliable. The provided tools should only be applied in cases where the monodisperity is impossible to achieve (e.g. low affinity complexes or oligomers).

Both GASBORMX and SASREFMX are included in the ATSAS package which is freely available for the academic users and can be downloaded from URL http://www.embl-hamburg.de/biosaxs/download.html. The access to GASBORMX and SASREFMX is also

provided via ATSAS-online interface for remote calculations at the EMBL BioSAXS cluster (http://www.embl-hamburg.de/biosaxs/atsas-online/).

Acknowledgements

We thank the staff of the SWING beamline (SOLEIL synchrotron, Gif-sur-Yvette, France) for assistance during synchrotron data collection.

Supporting Information Available

Supplementary Tables S1 with the overall parameters of the test examples, S2 with SEC derived estimation of the molecular mass, figures displaying the schemes of GASBORMX and SASREFMX algorithms, p(r) functions of the test examples, comparison of DR model with the GTS crystal structure, Sec profile of BSA, comparison of *ab initio* and rigid body models of BSA dimer, scheme of ERR-IR3 architecture, comparison of ERRα and ERRγ scattering profiles as well as ERRα LBD and full-length ERRα-IR3 models and fitting (Supplementary Figures S1-S10). This material is available free of charge via the Internet at http://pubs.acs.org.

References

1. Mertens, H. D., and Svergun, D. I. (2010) Structural characterization of proteins and complexes using small-angle X-ray solution scattering, *J Struct Biol 172*, 128-141.
2. Jacques, D. A., and Trewhella, J. (2010) Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls, *Protein Sci 19*, 642-657.
3. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis, *J Appl Crystallogr 45*, 342-350.
4. Feigin, L. A., and Svergun, D. I. (1987) *Structure analysis by small-angle x-ray and neutron scattering*, Plenum Press, New York.
5. Svergun, D. I., Volkov, V. V., Kozin, M. B., and Stuhrmann, H. B. (1996) New developments in direct shape determination from small-angle scattering 2. Uniqueness., *Acta Crystallogr. A52*, 419-426.
6. Chacon, P., Moran, F., Diaz, J. F., Pantos, E., and 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.
7. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophys J 76*, 2879-2886.
8. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering, *J. Appl. Cryst. 42*, 342-346.
9. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. J. (2001) Determination of domain structure of proteins from X-ray solution scattering, *Biophys J 80*, 2946-2953.

10. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data, *Biophys J 89*, 1237-1250.

11. Evrard, G., Mareuil, F., Bontems, F., Sizun, C., and Perez, J. (2011) DADIMODO: a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints, *J Appl Crystallogr 44*, 1264-1271.

12. Gabel, F., Simon, B., Nilges, M., Petoukhov, M., Svergun, D., and Sattler, M. (2008) A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints, *Journal of Biomolecular Nmr 41*, 199-208.

13. David, G., and 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, *J Appl Crystallogr 42*, 892-900.

14. Koch, M. H., Vachette, P., and Svergun, D. I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution, *Q Rev Biophys 36*, 147-227.

15. Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun, D. I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis, *J. Appl. Crystallogr. 39*, 277-286.

16. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering, *J Am Chem Soc 129*, 5656-5664.

17. Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering, *Gen Physiol Biophys 28*, 174-189.

18. Makowski, L., Gore, D., Mandava, S., Minh, D., Park, S., Rodi, D. J., and Fischetti, R. F. (2011) X-ray solution scattering studies of the structural diversity intrinsic to protein ensembles, *Biopolymers 95*, 531-542.

19. Yang, S., Blachowicz, L., Makowski, L., and Roux, B. (2010) Multidomain assembled states of Hck tyrosine kinase in solution, *Proc Natl Acad Sci U S A 107*, 15757-15762.

20. Vestergaard, B., Groenning, M., Roessle, M., Kastrup, J. S., van de Weert, M., Flink, J. M., Frokjaer, S., Gajhede, M., and Svergun, D. I. (2007) A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils, *Plos Biol 5*, e134.

21. Blobel, J., Bernado, P., Svergun, D. I., Tauler, R., and Pons, M. (2009) Low-resolution structures of transient protein-protein complexes using small-angle X-ray scattering, *J Am Chem Soc 131*, 4378-4386.

22. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J. Appl. Crystallogr. 25*, 495-503.

23. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, *J. Appl. Crystallogr. 28*, 768-773.

24. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small angle scattering, *J. Appl. Crystallogr. 36*, 860-864.

25. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models, *J. Appl. Crystallogr. 34*, 33-41.

26. Roessle, M. W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrmann, T., Konarev, P., Round, A., Fiedler, S., Hermes, C., and 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.

27. Round, A. R., Franke, D., Moritz, S., Huchler, R., Fritsche, M., Malthan, D., Klaering, R., Svergun, D. I., and Roessle, M. (2008) Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33, *J Appl Crystallogr 41*, 913-917.

28. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS - a Windows-PC based system for small-angle scattering data analysis, *J. Appl. Crystallogr. 36*, 1277-1282.

29. Diamond, R. (1974) Real-space refinement of the structure of hen egg-white lysozyme, *J Mol Biol 82*, 371-391.

30. Krissinel, E., and Henrick, K. (2005) Detection of Protein Assemblies in Crystals, In *CompLife 2005* (Berthold, M. R., Ed.), pp 163-174, Springer-Verlag, Berlin Heidelberg.

31. Gorynia, S., Bandeiras, T. M., Pinho, F. G., McVey, C. E., Vonrhein, C., Round, A., Svergun, D. I., Donner, P., Matias, P. M., and Carrondo, M. A. (2011) Structural and functional insights into a dodecameric molecular machine - the RuvBL1/RuvBL2 complex, *J Struct Biol 176*, 279-291.

32. Paravisi, S., Fumagalli, G., Riva, M., Morandi, P., Morosi, R., Konarev, P. V., Petoukhov, M. V., Bernier, S., Chenevert, R., Svergun, D. I., Curti, B., and Vanoni, M. A. (2009) Kinetic and mechanistic characterization of Mycobacterium tuberculosis glutamyl-tRNA synthetase and determination of its oligomeric structure in solution, *Febs J 276*, 1398-1417.

33. Majorek, K. A., Porebski, P. J., Dayal, A., Zimmerman, M. D., Jablonska, K., Stewart, A. J., Chruszcz, M., and Minor, W. (2012) Structural and immunologic characterization of bovine, horse, and rabbit serum albumins, *Mol Immunol 52*, 174-182.

34. Covaceuszach, S., Cassetta, A., Konarev, P. V., Gonfloni, S., Rudolph, R., Svergun, D. I., Lamba, D., and Cattaneo, A. (2008) Dissecting NGF interactions with TrkA and p75

receptors by structural and functional studies of an anti-NGF neutralizing antibody, *Journal of Molecular Biology 381*, 881-896.

35. Laudet, V., and Gronemeyer, H., (Eds.) (2002) *The Nuclear Receptor Factsbook*, Academic Press.

36. Giguere, V. (2008) Transcriptional control of energy homeostasis by the estrogen-related receptors, *Endocr Rev 29*, 677-696.

37. Greschik, H., Wurtz, J. M., Sanglier, S., Bourguet, W., van Dorsselaer, A., Moras, D., and Renaud, J. P. (2002) Structural and functional evidence for ligand-independent transcriptional activation by the estrogen-related receptor 3, *Mol Cell 9*, 303-313.

38. Kallen, J., Schlaeppi, J. M., Bitsch, F., Filipuzzi, I., Schilb, A., Riou, V., Graham, A., Strauss, A., Geiser, M., and Fournier, B. (2004) Evidence for ligand-independent transcriptional activation of the human estrogen-related receptor alpha (ERRalpha): crystal structure of ERRalpha ligand binding domain in complex with peroxisome proliferator-activated receptor coactivator-1alpha, *J Biol Chem 279*, 49330-49337.

39. Greschik, H., Althage, M., Flaig, R., Sato, Y., Chavant, V., Peluso-Iltis, C., Choulier, L., Cronet, P., Rochel, N., Schule, R., Stromstedt, P. E., and Moras, D. (2008) Communication between the ERRalpha homodimer interface and the PGC-1alpha binding surface via the helix 8-9 loop, *J Biol Chem 283*, 20220-20230.

40. Kallen, J., Lattmann, R., Beerli, R., Blechschmidt, A., Blommers, M. J., Geiser, M., Ottl, J., Schlaeppi, J. M., Strauss, A., and Fournier, B. (2007) Crystal structure of human estrogen-related receptor alpha in complex with a synthetic inverse agonist reveals its novel molecular mechanism, *J Biol Chem 282*, 23231-23239.

Table 1. Modelling of ERRα-IR3 complex from polydisperse data

Concentration, mg/ml	Volume fraction of 6:3 complex, %	χ
2.9	21±2	1.05
1.9	16±2	0.98
1.0	11±2	0.96

Figure captions

Figure 1. Applicability of the new *ab initio* and rigid body modelling algorithms for different mixture scenario. (A) Oligomer-monomer equilibrium with full dissociation – supported by GASBORMX and SASREFMX; (B) Oligomer-monomer equilibrium with intermediates – not supported; (C) Complex with excess of one component – supported by SASREFMX; (D) Complex with excess of a subcomplex – supported by SASREFMX; (E) Complex with individual components (full dissociation) – supported by SASREFMX; (F) Complex and multiple subcomplexes (intermediates) – not supported.

Figure 2. Modelling of dimeric lysozyme interface versus simulated data from equimolar mixture of monomers and dimers. (A) The *ab initio* model built by GASBORMX (blue and cyan beads represent two monomers) is compared with the atomic structure (monomers shown as red and orange backbones). (B) Rigid-body model of the dimer reconstructed by SASREFMX (blue and cyan backbones) overlayed to the original structure (as in panel A). (C) Fit from the rigid body model to the simulated data.

Figure 3. Modelling of RuvB ATPase from simulated data of hexamer – monomer equilibrium. (A) Crystallographic model of the RuvB hexamer. (B) Scattering profiles of the mixtures containing 15% (1) and 85% (2) of the monomer. Simulated data are shown as black circles, fits from reconstructed models as red solid lines. Profiles are displaced in logarithmic scale for better visualization. (C) and (D) present typical hexamer reconstructions from the curve containig 15% of the monomer, (E) and (F) – for 85% monomeric content. *Ab initio* models in (C) and (E) are shown in spacefilling mode, rigid body models in (D) and (F) are represented by the backbones. Right views in (A), (C)-(F) are rotated by 90 degrees about vertical axis.

Figure 4. Modelling of dimeric GTS from polydisperse solutions. (A) *Ab initio* model (blue and cyan beads represent two monomers) built by GASBORMX versus simulated data from equimolar mixture of monomers and dimers. (B) Rigid-body model of the dimer (blue and cyan backbones) reconstructed by SASREFMX versus simulated data from equimolar mixture of monomers and dimers. (C) Rigid body model of the dimer reconstructed using contact restraints from the experimental SAXS data containing 70% dimers and 30% monomers (by volume). (D) Fit of the rigid-body model to the experimental data. In (A)-(C) crystallographic interface is shown for comparison (monomers are represented by red and orange backbones).

Figure 5. Modelling oligomeric equilibrium of BSA. (A) Scattering profiles of SEC-purified (1) and solely centrifuged (2) BSA. Black dots represent experimental data, red solid lines are the intensities computed from the crystallographic monomer. Green and blue dashed lines are the fits to the data of the centrifuged sample, assuming an oligomeric mixture of dimers built using GASBORMX and SASREFMX, respectively. The profiles are displaced in logarithmic scale for better visualization. The inset shows the comparison of the experimental data at low angles, where the data of the SEC purifed BSA sample are shown by open circles. (B) Typical *ab initio* dimeric model obtained by GASBORMX (in space filling mode) yielding NSD=1.6 to the closest interface from the BSA crystallographic lattice. (C) Typical dimer (shown as backbone)

built by rigid-body modelling using SASREFMX with *r.m.s.d.* of 2A to the closest crystallographic interface. In (B)-(C), distinct monomers are displayed in blue and cyan. Bottom views are rotated by 90° around the horizontal axis.

Figure 6. Modelling of partially dissociated 2:1 Fab-NGF complex. (A) Fit of the rigid-body model (red line) to the experimental data from the polydisperse Fab-NGF sample (black dots). The best fit from the monodisperse solution of 2:1 complex generated by computational docking is shown as blue dashed line. (B) Comparison of the best computational docking model published in [REF] (top) with the rigid body reconstruction by SASREFMX (bottom). NGF dimers are shown in blue and light blue and the two bound Fab molecules as green/red and light green/orange.

Figure 7. Modelling of ERRα-IR3 complex from polydisperse data. (A) Scattering patterns at different concentrations: (1) – 3 mg/ml, (2) – 2mg/ml, (3) – 1mg/ml. Experimental data are shown by black dots, the fits corresponding to various ratios of 6:3 and 2:1 complexes are shown with red solid lines. The corresponding volume fractions are given in Table 1. The profiles are displaced in logarithmic scale for better visualization. (B) Rigid-body model of the 6:3 complex reconstructed using SASREFMX against the concentration series from equilibrium mixtures. Distinct ERRα LBD dimers of the hexameric core are shown in red orange and yellow, the DBD dimers are depicted in blue shades, three copies of IR3 RE are displayed in green. The bottom view is rotated by 90° about the horizontal axis.

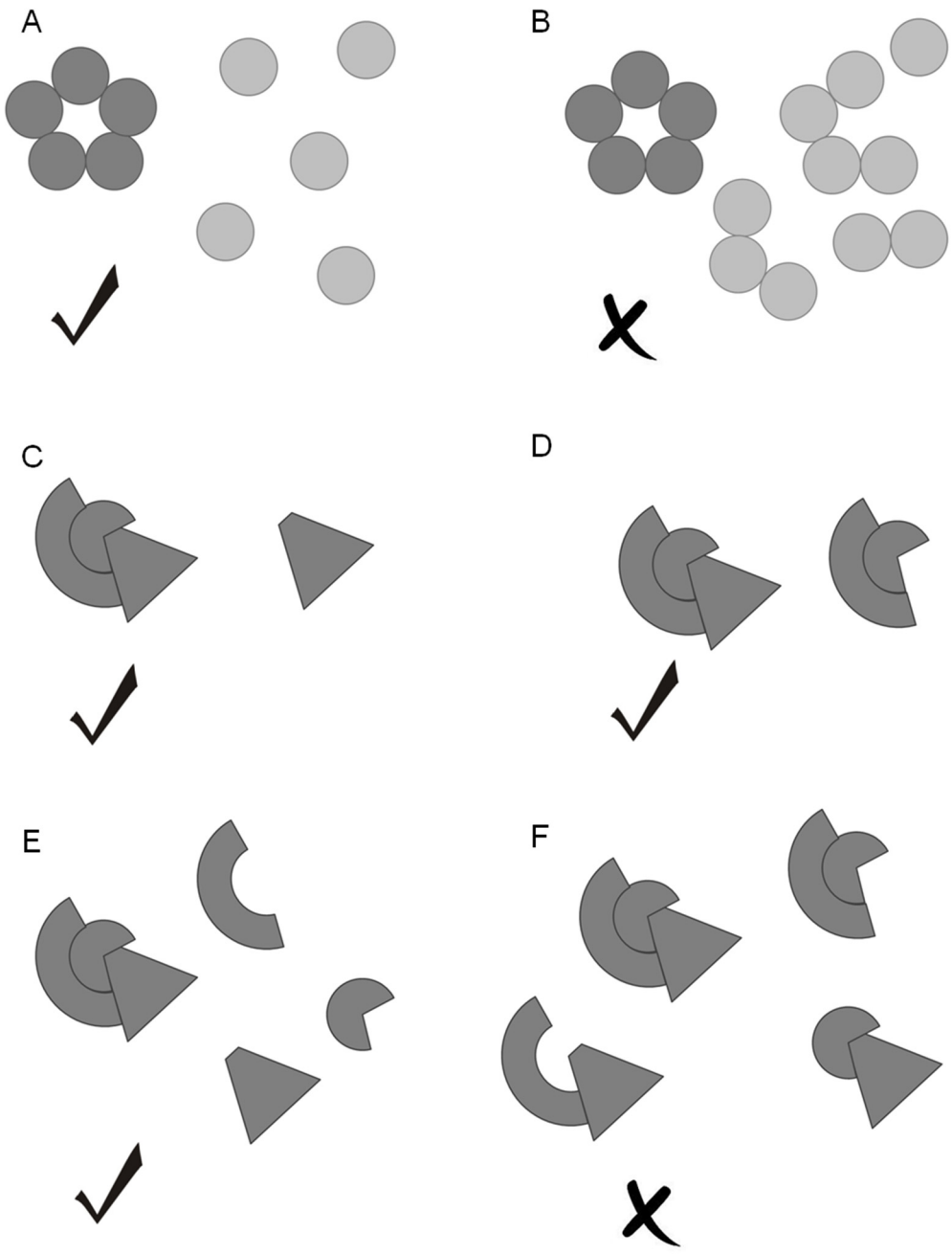

Figure 1.

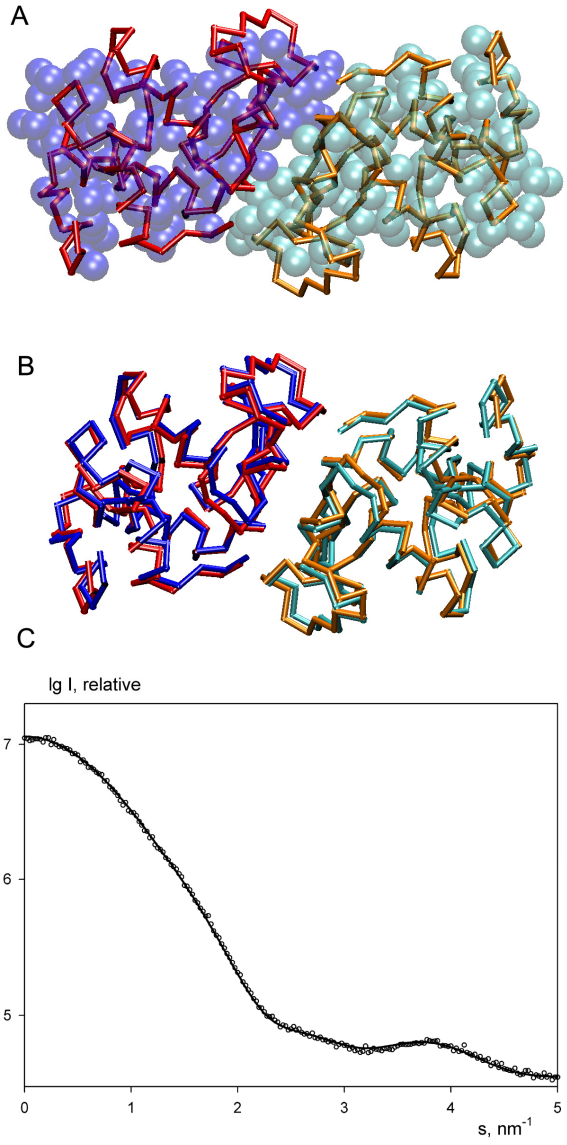

Figure 2.

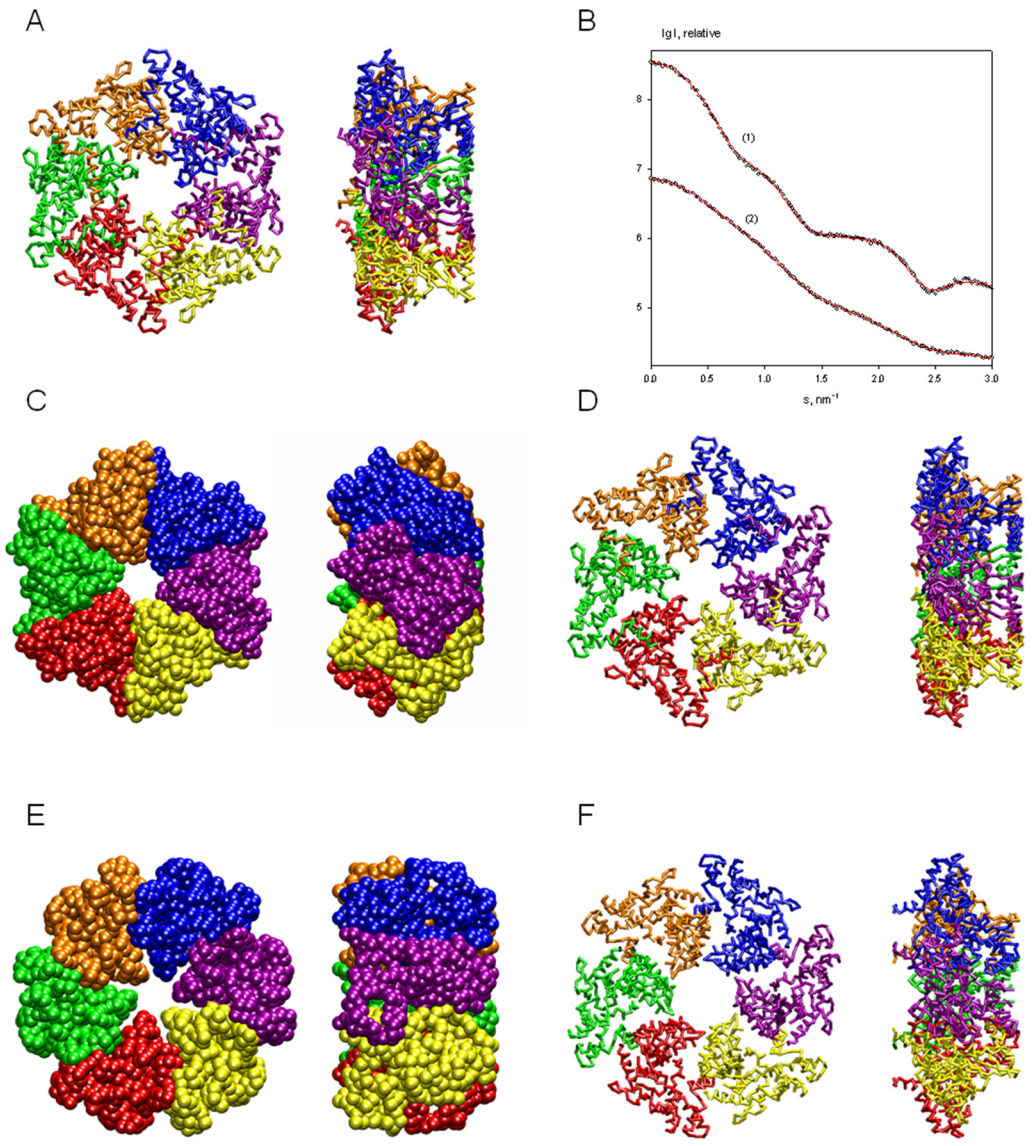

Figure 3.

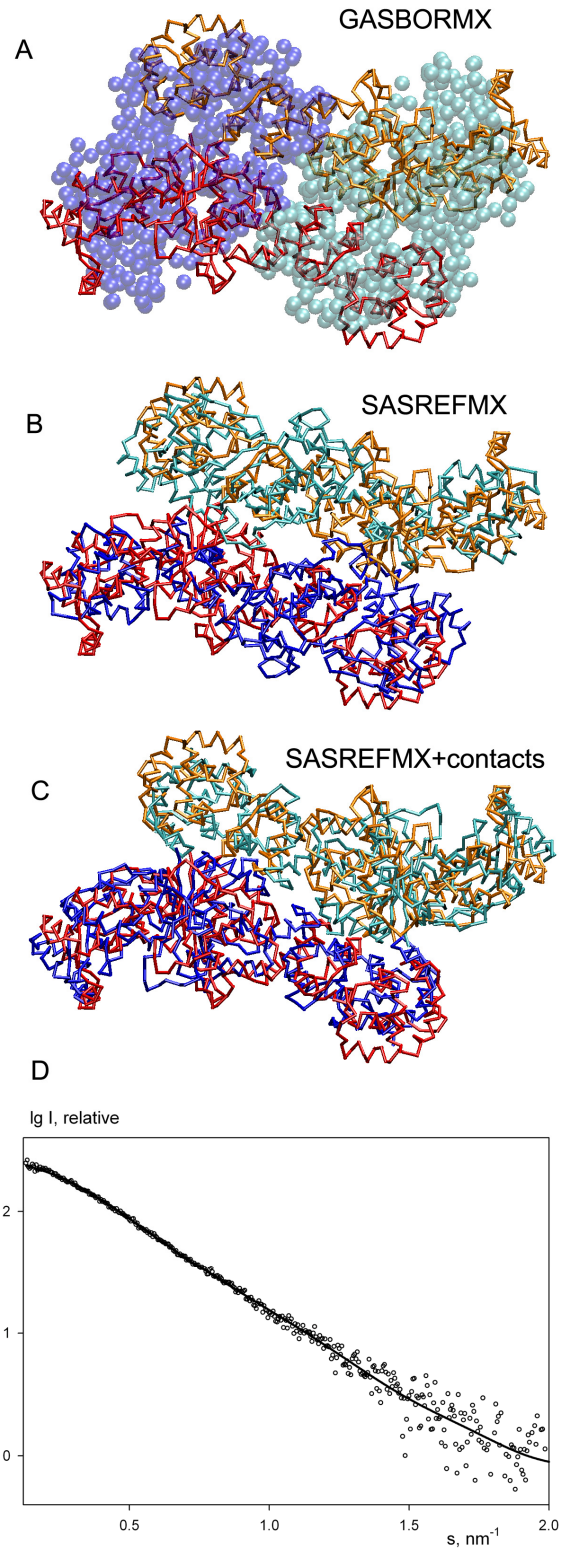
A
GASBORMX
B
SASREFMX
C
SASREFMX+contacts
D
lg I, relative
2
1
0
0.5
1.0
1.5
2.0
s, nm-1

Figure 4.

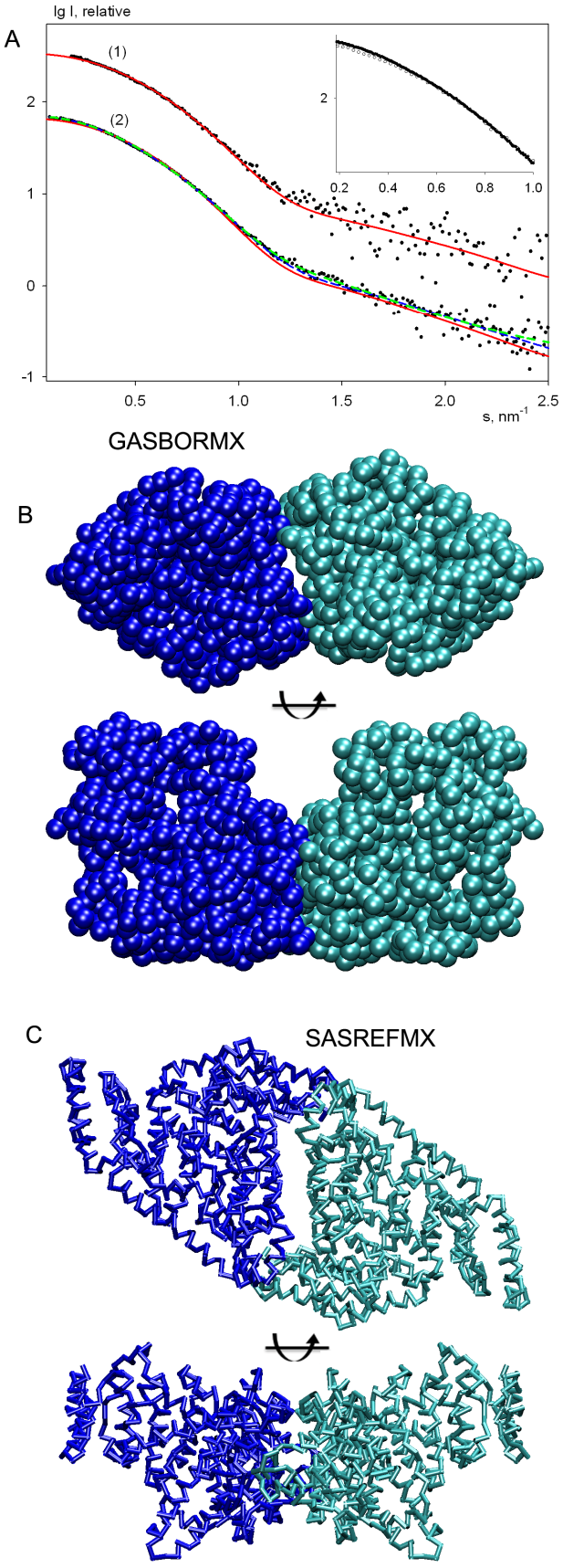
A
lg I, relative
(1)
(2)
2
1
0
-1
0.5
1.0
1.5
2.0
2.5
s, nm-1
0.2
0.4
0.6
0.8
1.0
GASBORMX
B
C
SASREFMX

Figure 5.

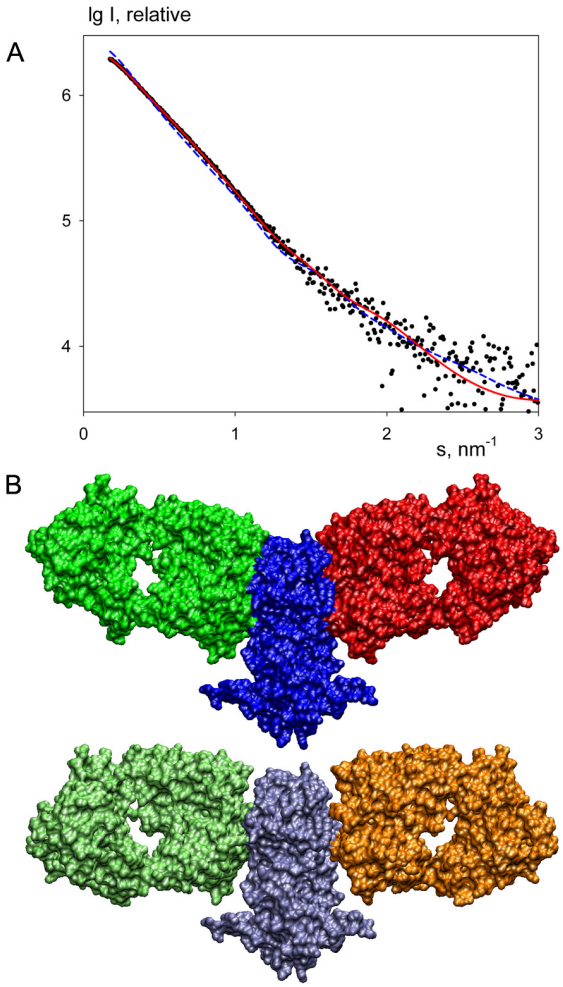

Figure 6.

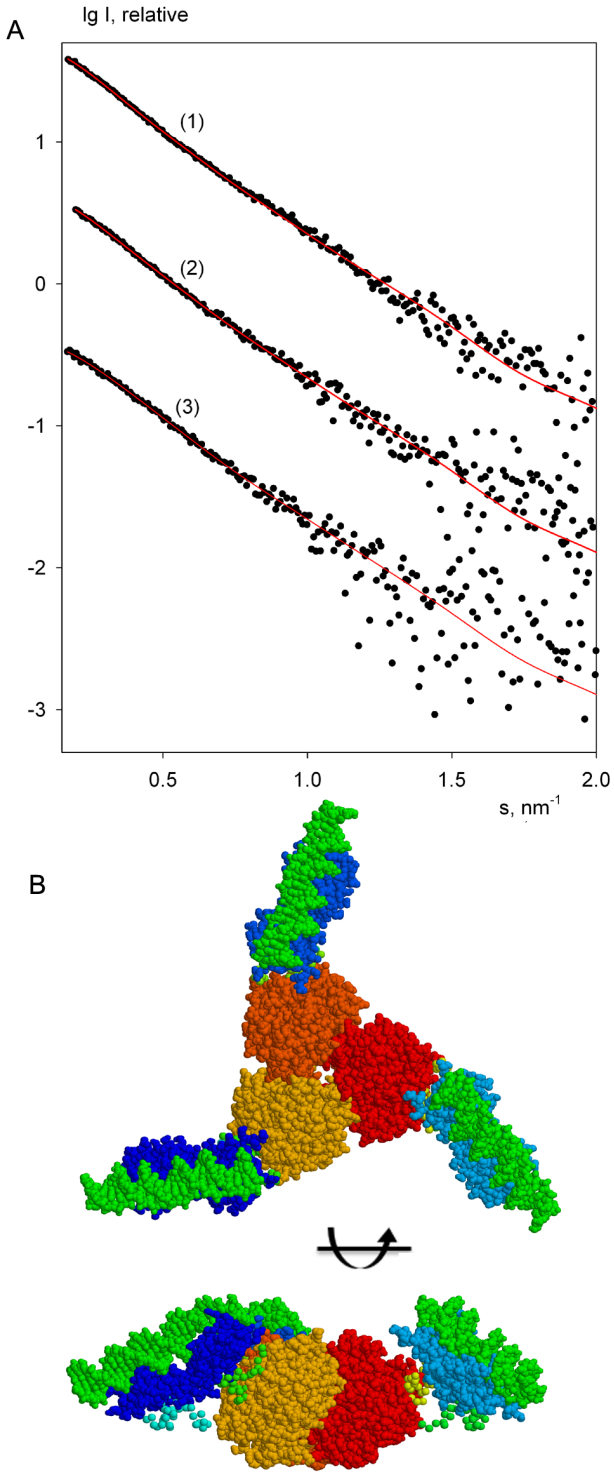

Figure 7.

For Table of Contents Use Only

"Reconstruction of quaternary structure from X-ray scattering by equilibrium mixtures of biological macromolecules"

by Maxim V. Petoukhov[1)], Isabelle M.L. Billas[2)], Maria Takacs[2)], Melissa A. Graewert[1)], Dino Moras[2)] and Dmitri I. Svergun[1)]

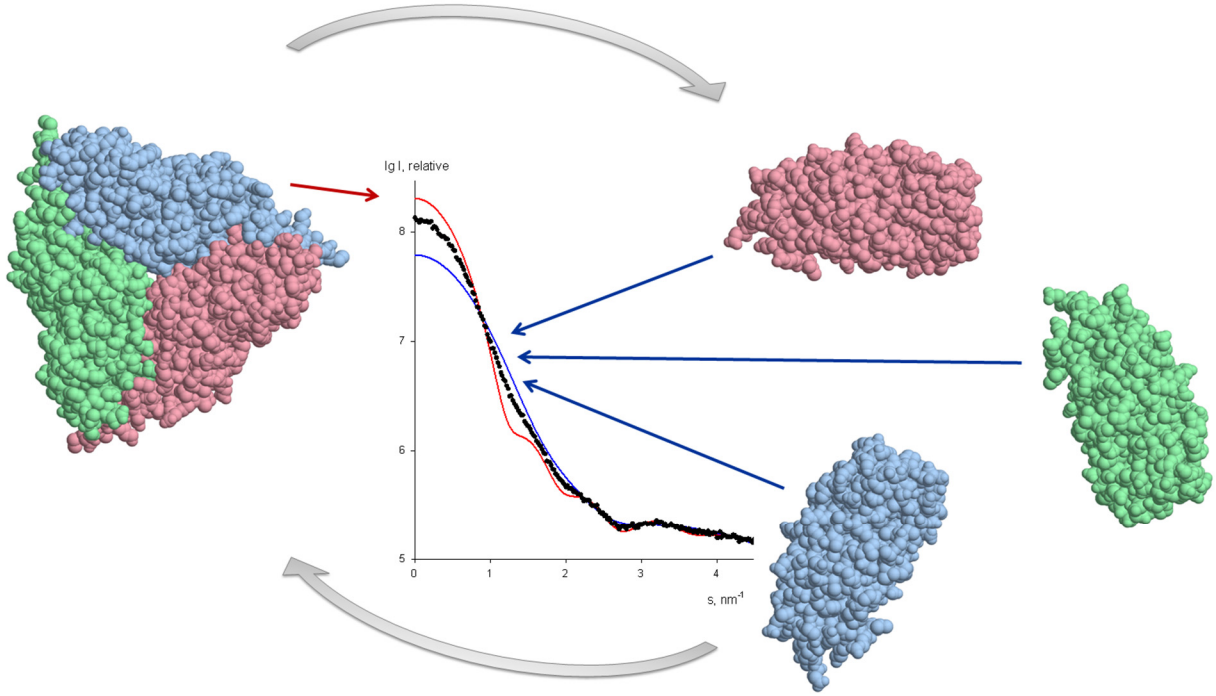

Table of Contents Graphic. 3D modelling algorithms (GASBORMX & SASREFMX) provide means for the structural characterisation of transient complexes and weak oligomers based on the SAXS profiles of these typical polydisperse systems.