REVIEW

Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS)

Michal Hammel

Abstract The dynamics of macromolecular conformations are critical to the action of cellular networks. Solution X-ray scattering studies, in combination with macromolecular X-ray crystallography (MX) and nuclear magnetic resonance (NMR), strive to determine complete and accurate states of macromolecules, providing novel insights describing allosteric mechanisms, supramolecular complexes, and dynamic molecular machines. This review addresses theoretical and practical concepts, concerns, and considerations for using these techniques in conjunction with computational methods to productively combine solution-scattering data with high-resolution structures. I discuss the principal means of direct identification of macromolecular flexibility from SAXS data followed by critical concerns about the methods used to calculate theoretical SAXS profiles from high-resolution structures. The SAXS profile is a direct interrogation of the thermodynamic ensemble and techniques such as, for example, minimal ensemble search (MES), enhance interpretation of SAXS experiments by describing the SAXS profiles as population-weighted thermodynamic ensembles. I discuss recent developments in computational techniques used for conformational sampling, and how these techniques provide a basis for assessing the level of the flexibility within a sample. Although these approaches sacrifice atomic detail, the knowledge gained from ensemble analysis is often appropriate for developing hypotheses and guiding biochemical experiments. Examples of the use of SAXS and combined approaches with X-ray crystallography, NMR, and computational methods to characterize dynamic assemblies are presented.

Keywords Small-angle X-ray scattering (SAXS) ·
Macromolecular flexibility · Rigid-body modeling ·
Ensemble analysis

M. Hammel (✉)
Lawrence Berkeley National Laboratory, Physical Biosciences Division, Berkeley, CA 94720, USA
e-mail: mhammel@lbl.gov

Introduction

Current structure-based research has used high-resolution MX and NMR-derived structures to guide hypothesis-driven research. This has been effective for well-folded, compact enzymes, and has enabled atomic-level dissection of an enzyme's active site. Nevertheless, estimates suggest that over 50 % of eukaryotic proteins contain significant functional unstructured regions (Vucetic et al. 2003) that are intractable to current structure-based model. Macromolecular flexibility is an important aspect of the regulatory mechanisms of biological systems (Henzler-Wildman and Kern 2007; Perry et al. 2010). MX, NMR, and electron microscopy (EM) are regarded as the most reliable methods for determination of structure; nonetheless, these techniques are limited by macromolecules with functional flexibility and intrinsic disorder (Fink 2005). Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS) has recently become a central tool in the new area of characterizing multi-state systems within structural biology (Bernado et al. 2007). Combining data from solution scattering with atomic resolution structures has the potential to address how specific complexes and flexibility drive biological processes (Putnam et al. 2007; Rambo and Tainer 2010a). Although SAXS has some inherent limitations, there is sufficient information

within the one-dimensional scattering profile to distinguish between well-defined conformations and the conformational space occupied by a flexible assembly (Fig. 1). The theoretical basis for solution scattering has been the subject of an excellent review (Koch et al. 2003). Previously, I authored a review providing a general framework for experimental design, data processing, and data interpretation that combined SAXS with atomic-resolution structures from crystallography (Putnam et al. 2007). The purpose of this review is to discuss different tools and methods that have recently been developed for SAXS analysis of flexible multidomain assemblies.

SAXS profile as a indicator of flexibility

Recently Rambo et al. described the use of the Porod–Debye law as a powerful tool for distinguishing between rigid and flexible particles (Rambo and Tainer 2011). In particular, it was shown that for comparative SAXS experiments, application of the law can distinguish between discrete conformational changes and localized flexibility relevant to molecular recognition (Devarakonda et al. 2011; Williams et al. 2011). This approach aids insightful analysis of fully and partly flexible macromolecules that is more robust than traditional Kratky analysis (Porod 1982). Kratky analysis relies on visual inspection of the Kratky plot, which can be confounded by a limited observational q range ($q < 0.2$ Å^{-1}), the presence of high experimental noise, or by non-ideal buffer subtraction (Fig. 1a). Intensity measurements at high scattering angles are exponentially more sensitive to the buffer – blank subtraction than measurements near the Guinier region. Therefore, small errors during the buffer – blank subtraction may confound the baseline convergence necessary for assessing flexibility by
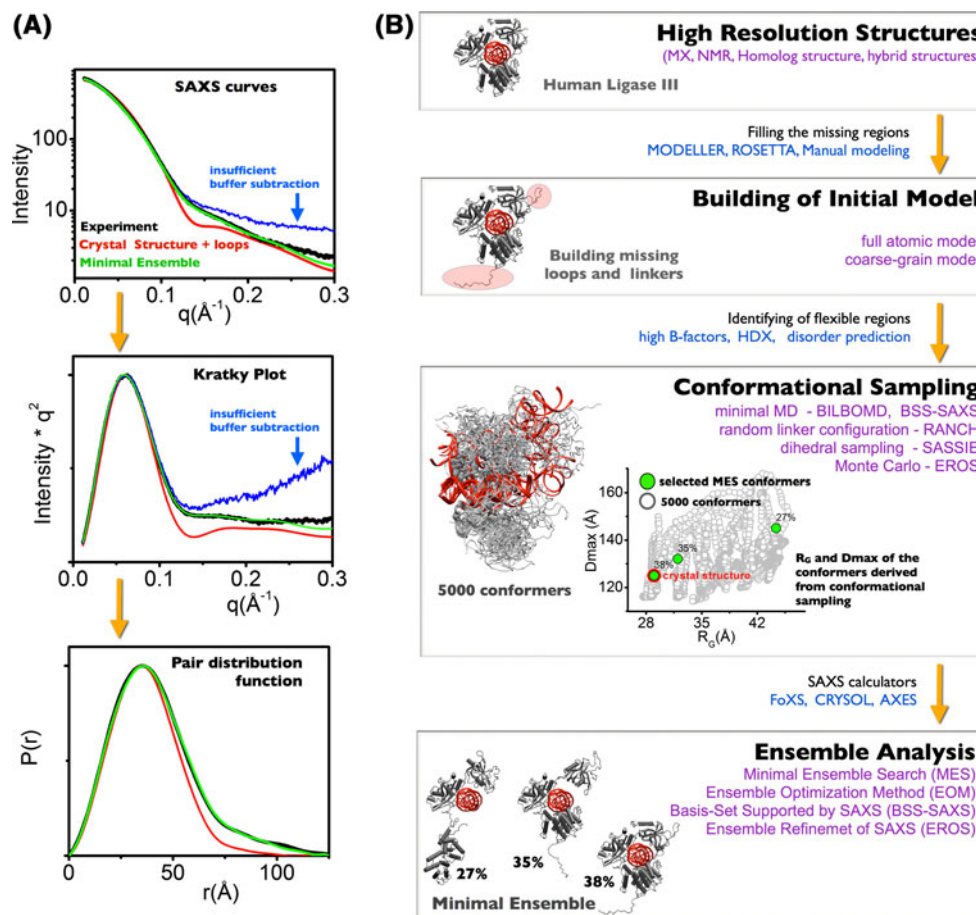

Fig. 1 Validation of flexibility using SAXS curve (**a**) and rigid-body modeling (**b**). **a** Experimental SAXS profiles (*black* and *blue*) for the human DNA Ligase III (Cotner-Gohara et al. 2010) in a match with theoretical profiles calculated for the crystal structure (*red*) (Cotner-Gohara et al. 2010) and its dynamic model (*green*) obtained by BILBOMD and MES (Pelikan et al. 2009). The Kratky plot is used as the initial indicator of the flexibility. Baseline convergence necessary for assessing flexibility is misleading for the SAXS curve with insufficient buffer subtraction (*gray*). Pair distribution $P(r)$ function calculated for the experimental (*black*) and the theoretical SAXS (*red*, *cyan*). Crystal structure, full-length and ensemble models used to calculate theoretical SAXS profiles are shown in the panel **a** (data adapted from Cotner-Gohara et al. 2010). **b** Schematic diagram of typical rigid-body modeling performing building of initial model, conformational sampling, and ensemble analysis

Kratky analysis (Fig. 1a). However, it has been shown that the Porod–Debye law resides within the low-resolution region of the SAXS profile, typically $q < 0.15$ Å^{-1}, that is routinely well measured and not prone to buffer – blank subtraction issues. For example, Kratky analysis of the SAXS data collected for the ATP-free and bound forms of Mre11–Rad50 (Williams et al. 2011) did not clearly identify flexibility of the ATP-free state and rather led to the hypothesis that the particle is switching between two distinct conformational states, similar to PYR1(Nishimura et al. 2009). However, inspection of the Porod plot suggests a fundamentally different mechanism. In the presence of ATP, the complex forms a distinct particle with a sharp scattering contrast, as evidenced by the Porod plateau (Fig. 2b), and in the absence of ATP the particle becomes

Fig. 2 Detecting conformational flexibility. **a** SAXS data for the Mre11–Rad50 complex in both the presence (*black*), and absence (*red*) of ATP (Williams et al. 2011), and an exemplary intrinsically disordered domain Rad51 AP1 (*blue*). *Inset* Comparison of the Kratky plots for Mre11–Rad50 complexes does not confidently demonstrate flexibility of the complex in the absence of ATP (*black* and *red*). However, the Kratky plot of Rad51 AP1 (*blue*) is hyperbolic in shape, clearly demonstrating the full unfolded particle. **b** Porod–Debye plot illustrating changes in the Porod–Debye region. Loss of the plateau suggests Mre11–Rad50 becomes more flexible in the absence of ATP. Rigid and flexible states of Mre11–Rad50 are presented with crystal structure of Mre11–Rad50-ATPγS (Lim et al. 2011) and dynamic model of Mre11–Rad50 (Williams et al. 2011) (*left panel*). Data were adapted from Williams et al. (2011) and Rambo and Tainer (2011). Data for Rad51 AP1 were kindly provided by Gareth Williams at the Lawrence Berkeley National Laboratory

more flexible. In fact, inspection of the Porod–Debye region demonstrates a loss of the plateau, supporting the hypothesis that Mre11–Rad50 is flexible in the absence of ATP. These types of analysis provide qualitative information about conformational states that give credence to modeling the solution state as an ensemble of conformers.

SAXS profiles provide more accurate atomic-level information about structures in solution without crystallographic constraints

Methods of analysis based on the concept of a single conformer cannot provide a complete three-dimensional model of dynamic proteins. Using a single "best" conformer to represent the ensemble at most provides a model representing an average of the conformations that exist in solution. Such a "best" single model of the macromolecular state can still be informative by helping guide a hypothesis regarding the macroscopic conformational state (Hammel et al. 2002; Iyer et al. 2008; Jain et al. 2009; Pascal et al. 2004; Williams et al. 2009). For example, if the crystal structure of a macromolecular assembly is known, a theoretical scattering profile can be calculated from the atomic coordinates. This provides the opportunity to evaluate several user-generated models (Fig. 1). If an extended conformer fits SAXS data better than a compact crystal structure, then an opening of the assembly in solution may be assumed (Nagar et al. 2006; Pascal et al. 2004; Yamagata and Tainer 2007).

Crystal packing forces are a selective pressure on a ensemble that typically promote a single conformer within the crystal lattice. Differences between crystal and solution states often reflect the presence of crystal packing forces (Cotner-Gohara et al. 2010; Datta et al. 2009; Duda et al. 2008; Nishimura et al. 2009; Stoddard et al. 2010) that can be used to gain new insights into a protein's flexibility (Nishimura et al. 2009). Direct comparisons of different conformational states with model SAXS profiles calculated from atomic-resolution structures have been quite successful in identifying and decomposing the relative fractions of conformers of a sample in solution, such as with the archaeal secretion ATPase GspE. The MX structure of the hexameric ring revealed a mixture of open and closed states of the individual subunits (Yamagata and Tainer 2007). In contrast, SAXS studies of GspE suggested a much different conformational state in solution. In the presence of the transition state ATP analogue, AMP-PNP, SAXS experiments suggest the enzyme's subunits assume an all-closed state. In the next step of the catalytic cycle, the ADP-bound state, SAXS experiments suggest GspE exists as a mixture of all-closed and all-open states. The original crystal structure of alternating open–closed states
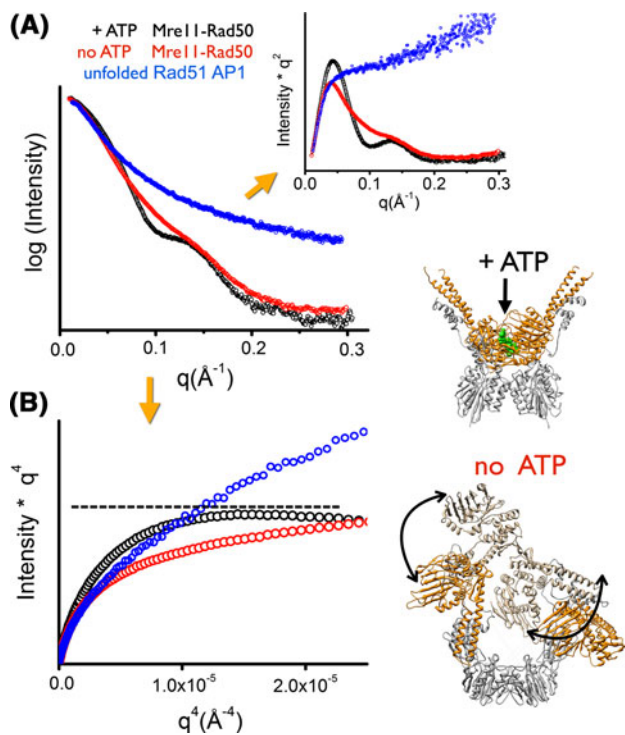
in a ring failed to explain the SAXS experiments and raises significant questions regarding the proper biological state of the crystallized GspE. Crystal packing forces are structurally selective (Nishimura et al. 2009; Stoddard et al. 2010); consequently, a structural biology approach solely dependent on MX will be limited in scope.

Accurate computation of SAXS profiles

High-quality SAXS experiments from advanced instrumentation (Hura et al. 2009) lead to more precise data and confident assignment of the conformational state(s) of a given sample. Notwithstanding instrumentation developments, accurate calculation of a SAXS profile is essential for the accuracy of solution structure modeling. Several methods are available to calculate SAXS profiles from atomic models, and differ in the use of the inter-atomic distances, estimation of excluded volume, treatment of the hydration layer, or background adjustment (Grishaev et al. 2010). Calculation of an SAXS profile from atomic coordinates requires spherical averaging that can be efficiently

accomplished by representing a macromolecule in terms of inter-atomic distances (Schneidman-Duhovny et al. 2010; Zuo et al. 2006) or by using spherical harmonic reconstructions (Grishaev et al. 2010; Liu et al. 2012; Svergun et al. 1995). Explicit calculation of inter-atomic distances with solX software (Zuo et al. 2006) requires more intensive computation, but results in good agreement throughout the large q range with experimental scattering profiles (Putnam et al. 2007). Calculating profiles for anisometric shapes or unfolded regions is also more problematic for spherical harmonic reconstructions (reviewed by Putnam et al. 2007) and inaccuracies in fitting can be compensated by over-adjustment of excluded volume or the density of the hydration layer. As the data quality becomes extraordinary good, full atomistic models are required for accurate interpretation of the experimental SAXS profiles (Fig. 3). In this example of a high-resolution experimental SAXS of the cellulase Cel5A catalytic domain, explicit calculations using inter-atomic distances of several models demonstrate that the calculation of accurate profiles may detect small unfolded regions (Fig. 3). SAXS can detect these unstructured regions only because they affect the overall/globular
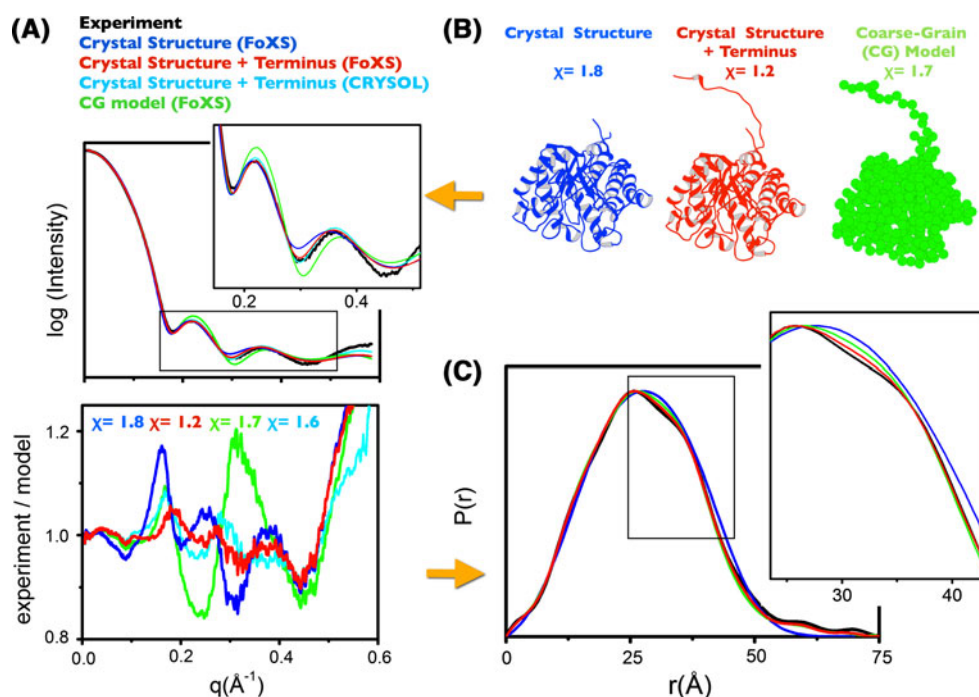

Fig. 3 Accuracy of SAXS-profile calculations. **a** Comparison of the experimental scattering curves of cellulase Cel5A catalytic domain (*black*) with the theoretical curves for Cel5A crystal structure missing the C-terminal unfolded region (PDB 1EDG) (*blue*), full-atomistic model (*red*), and coarse-grain (CG) model (*green*), shown in panel **b**. *Bottom panel* The discrepancy between theoretical and experimental profiles is calculated as Intensity$_{(experiment)}$/Intensity$_{(model)}$. Please note the large discrepancy for the CG model ($\chi = 1.7$) and crystal structure ($\chi = 1.8$) in comparison with the full-atomistic model calculated by FoXS ($\chi = 1.2$). Better profile matches are

obtained by calculating explicit atom distances (FoXS $\chi = 1.2$) in comparison with the SAXS profile calculated by spherical harmonics using CRYSOL—Linux version 2.7 ($\chi = 1.6$). **c** $P(r)$ functions calculated for SAXS profiles shown in **a** have been calculated by use of the software GNOM (Svergun 1992). The production and purification of the cellulase Cel5A catalytic domain has been described elsewhere (Fierobe et al. 2002). SAXS experiments were performed at the European Synchrotron Radiation Facility (Grenoble, France) on beamline ID02 as described by (Hammel et al. 2004a)

shape of the protein. However, the example presented clearly shows the kind of information content stored in the SAXS profiles or its $P(r)$ functions derived from them (Fig. 3b). The fact that the full-atomistic model is important to match experimental data has been further shown by analysis of 19 proteins containing a 19-residue His tag (Hura et al. 2009). His tags increase D_{max}, and should be modeled explicitly with available core atomic models.

Fitting theoretical models to SAXS profiles requires that a measure be established for determining the agreement between two scattering curves. I am not convinced that a "best" measure of assessing agreement between experimental and theoretical curves has been adequately developed. The standard χ clearly weighs the lowest resolution data most strongly. The χ values become less informative as the high resolution SAXS profiles with "low-noise" are used to fit atomistic models. For additional assessment of the quality of model-data agreements I suggest displaying the discrepancy by using the ratio calculated as $I_{experiment}/I_{model}$. This residual-ratio clearly displays discrepancies in the important small q region whereas the standard $\log_{10}$-based presentation of $\log(I)$ versus q frequently does not (Figs. 3a and 5).

Better quality experimental data promotes the need for increased accuracy and computations of SAXS profiles. By using explicit-all atom distances (Schneidman-Duhovny et al. 2010) and water models to account for the effect of solvent (Grishaev et al. 2010) superior fits between experimental high resolution structures and SAXS data are obtained (Fig. 3a). The explicit representation of the molecule is particularly useful for multidomain-flexible assemblies, which frequently adopt highly anisometric shapes (Grishaev et al. 2010). The FoXS algorithm explicitly computes all inter-atomic distances that include the first solvation layer based on the atomic solvent accessible areas (Fig. 3). As FoXS is available through a web server, it enables uploading and simultaneous analysis of a collection of atomic coordinate input files against experimental data. In combination with the MES (Pelikan et al. 2009) that is also part of the suite, the user is provided with powerful tools to identify the heterogeneity or flexibility of the experimental system. These powerful analytical techniques, together with advanced instrumentation, have been the basis for visualizing minimum conformational changes in human complement C3b (Chen et al. 2010) (Fig. 4).

Modeling of the conformational space

Although comparison of model SAXS profiles with the experimental data is one of the most straightforward applications of SAXS, the uniqueness of arrangements of atomic resolution structures that fit SAXS data must also be evaluated. The determination of multidomain or subunit assemblies using rigid-body modeling in conjunction with SAXS data involves preparing a large number of possible atomic models and comparing them with experimental data. The models can either be refined directly against experimental data (Petoukhov and Svergun 2005) or prepared independently using the SAXS data as a filter to select the "best fit" model(s) (Boehm et al. 1999; Forster et al. 2008). The biggest challenge in trying to model flexible multidomain systems using SAXS data is to avoid over-fitting. Most commonly, over-fitting can be detected by visually inspecting the selected models and examining for large unfolded regions or unrealistic inter-domain distances. Extremely elongated or partially unfolded structures may contribute to inappropriate "successful fits" of experimental data derived from aggregated or heterogeneous samples (reviewed by Putnam et al. 2007). For example, studies of mammalian lipoxygenase illustrate the need for establishing monodispersity of sample in cases where domain flexibility is proposed (Dainese et al. 2005; Hammel et al. 2004b; Shang et al. 2011). In early studies the discrepancy between the experimental curve of mammalian lipoxygenase and the profile calculated from the atomic coordinates were interpreted in terms of a very large movement of the N-terminal domain (Hammel et al. 2004b). In a recent study, however, Shang et al. (2011) found that mammalian lipoxygenase, besides its flexible N-terminal domain, forms a transient dimer that also leads to an elongated SAXS signal. Therefore, samples that are suspected of possessing intrinsic flexibility must be carefully characterized to ensure monodispersity before SAXS modeling (Rambo and Tainer 2010b).

A number of techniques have been used to generate realistic atomic models that sample conformational space of multi-modular proteins. Monte Carlo simulation (Forster et al. 2008; Rozycki et al. 2011) based on exploration of the dihedral angles in connection regions (Akiyama et al. 2004; Curtis et al. 2012), torsion/Cartesian simulated annealing (Schwieters et al. 2010), and minimal molecular dynamics (minimal MD) (Boehm et al. 1999; Hammel et al. 2005; Yang et al. 2010) may all be used. In the early years of rigid-body modeling the Perkins group developed constrain molecular modeling. This approach was applied to solution structure determinations of human and chimeric antibodies (reviewed by Perkins and Bonner 2008). The technique uses a large number of conformers that are built with directed MD computations applied only to the inter-domain connections. These models are filtered on the basis of their agreement with properties extracted from experimental SAXS curves, for example the radius of gyration, radius of gyration of cross sections, and the overall fit of the theoretical scattering from the model to the experimental data (Abe et al. 2010; Aslam and Perkins 2001; Gilbert et al.
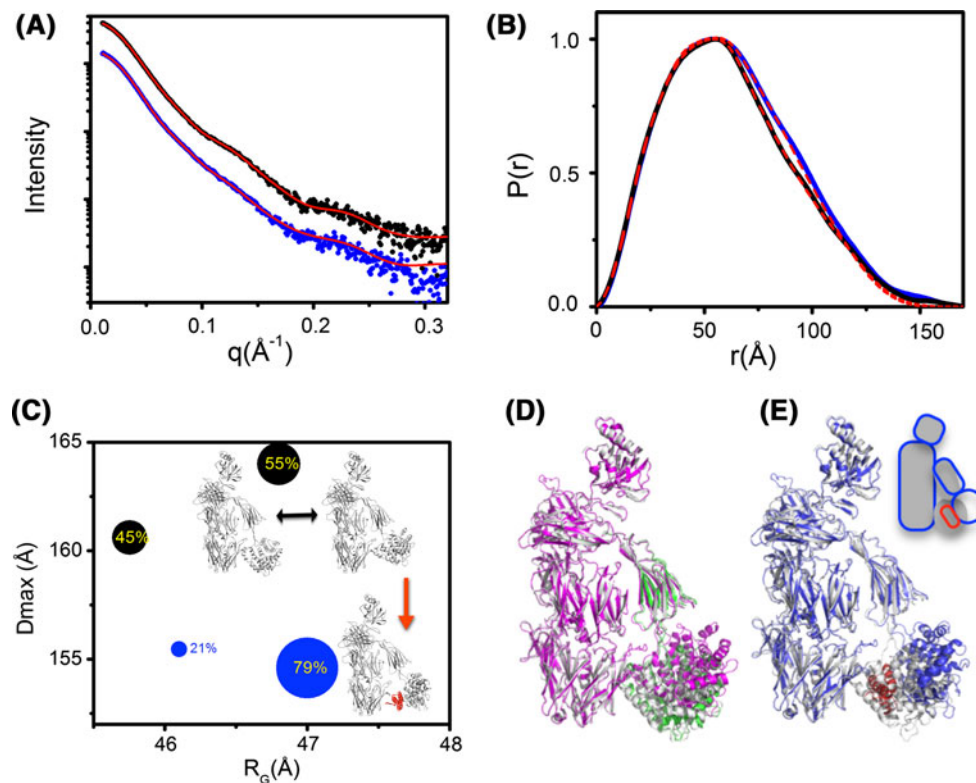
Fig. 4 Efb-induced conformational changes in human complement C3b as revealed by SAXS. **a** Experimental scattering curves for free C3b (*black*) and in the complex with extracellular fibrinogen-binding protein (Efb) from *Staphylococcus aureus* (C3b/Efb) (*blue*) were fit to MES model (*red line*). **b** $P(r)$ functions indicate conformational changes between C3b (*black*) and C3b/Efb (*blue*), where broadening of $P(r)$ for C3b/Efb-C is consistent with reorientation of the CUB-TED domain. $P(r)$ from the atomic MES models is shown as a *red dashed line*. **c** Comparison of R_G for the two predominant MES conformers of either C3b (*black*) or C3b/Efb (*blue*) as obtained by BILBOMD sampling with their maximum dimensions (D_{max}). Dot sizes represent the fraction ratio of the two conformers in each group. Rigid-body modeling-derived C3b conformers are shown in *gray* with Efb highlighted in *red*. (**d**, **e**) Superposition of the BILBOMD-MES-derived conformers of free C3b (**d**, *magenta* and *green*) and C3b/Efb (**e**, *blue/red*) with the crystal structure of C3b (*gray*). The *inset* shows a schematic representation of the proposed domain rearrangements. Data were adapted from Chen et al. (2010)

2006; Khan et al. 2010; Li et al. 2010). Constrained modeling confirms the experimental data analysis and produces families of best-fit models. Although these molecules are most likely an ensemble with a wide range of conformations, the selected best fit conformers are sufficient to reveal conformational switching or flexibility. The recently developed BILBOMD approach uses a similar minimal MD strategy and describes the final model as a population-weighted ensemble selected from the entire pool of conformers (Pelikan et al. 2009) (Figs. 4 and 5).

Conformational sampling may also be performed with simplified coarse-grain (CG) models, where amino-acid residues are presented as spherical beads centered at corresponding $C\alpha$ atom positions (Rozycki et al. 2011; Yang et al. 2010). Although extremely simplified, CG incorporates the main generic features and folding data of the protein under investigation. The CG models are used to not only speed up the production phase of conformational sampling but also to speed up the SAXS calculation. However, CG models are coarse representations, and it has

been shown that full atomistic models are required for accurate calculation of SAXS profiles (Grishaev et al. 2010) (Fig. 3). Particularly for modeling flexible assemblies, the atomistic representation is essential for accurate representation in solution when the particles deviate from a canonical globular shape (Fig. 3).

Distance constraints in rigid-body modeling

Accurate assignment of the flexible regions is crucial to realistic conformational sampling. In most cases, analysis of high-resolution structures can indicate plausible regions of structural flexibility (Chen et al. 2010). Missing electron density (Bernstein et al. 2009; Biersmith et al. 2011; Hammel et al. 2007a; Hammel et al. 2010b) or regions with a high isotropic atomic displacement factor (ADF also called the B-factor) (Duda et al. 2008; Williams et al. 2011) are useful indicators of flexible regions. Empirical determination of flexible regions can be achieved by
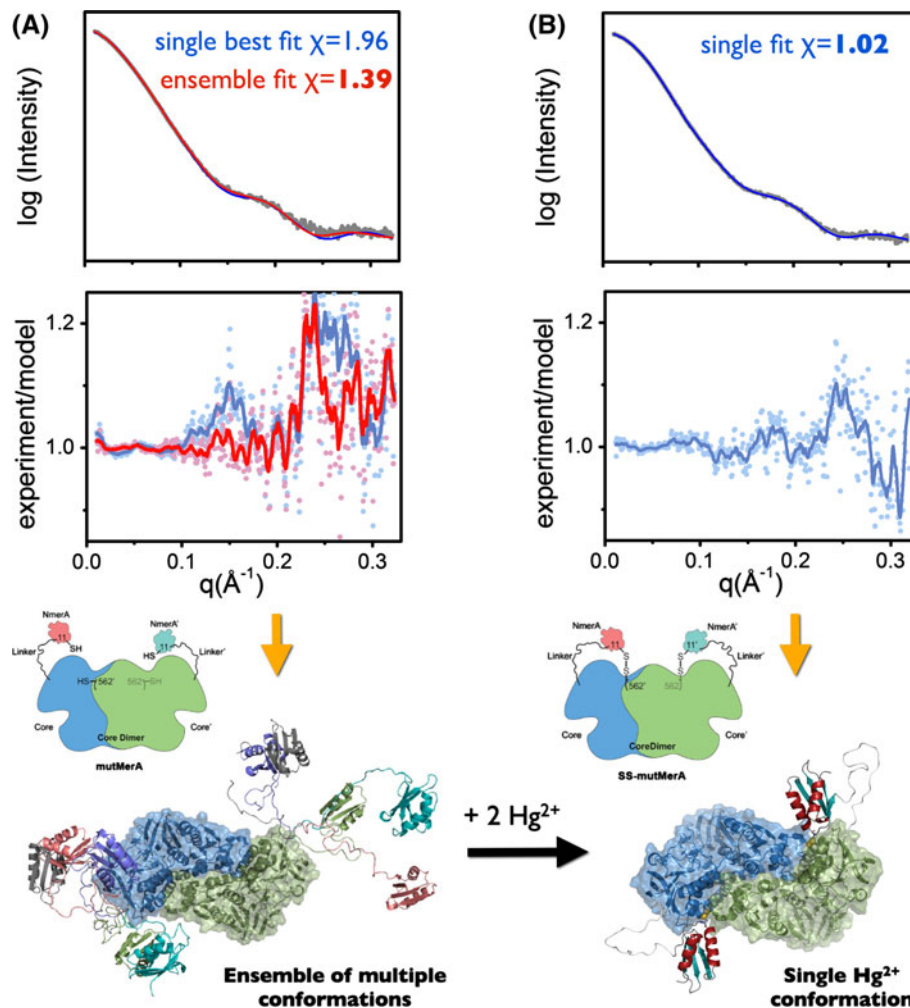
Fig. 5 Solution structure modeling of intramolecular Hg^{2+} transfer between flexibly linked domains of mercuric ion reductase (MerA). **a** Comparison of experimental and calculated scattering profiles for full-length MerA (mutMerA). Experimental SAXS data (*gray*), single best-fit conformation to the experimental scattering profile with $\chi = 1.96$ (*blue line*), and combined profile from five contributing conformations identified by MES (*red line*) with $\chi = 1.39$. Residuals calculated as $I_{experiment}/I_{model}$ are shown at the *bottom*. Superposition of the five models identified by MES with the metallochaperone-like N-terminal domains in a different color weighted by the factors 0.40 (*pink*), 0.29 (*green*), 0.16 (*cyan*), 0.08 (*purple*), and 0.07 (*gray*). **b** Experimental SAXS data for the disulfide-cross-linked handoff complex (SS–mutMerA) (*gray*) and calculated scattering data for the single best-fit conformation $\chi = 1.02$ (*blue line*). Residuals $I_{experiment}/I_{model}$ are shown as *blue dots* and as a *blue line* for smooth residuals. *Inset* shows the schematic representation of mutMerA and S–S-mutMerA. Data were adapted from Johs et al. (2011)

hydrogen–deuterium exchange mass spectrometry (HDX) which specifically follows changes in conformational states of proteins. For example, HDX clearly assigned the flexible region in the complement C3b molecule after its activation (Hammel et al. 2007b). This HDX experiment guided SAXS based rigid-body modeling used to visualize the C3b molecule as a highly dynamic system. SAXS modeling also revealed that C3b flexibility may be effected by an allosteric inhibitor, for example the extracellular fibrinogen-binding protein (Efb) from *Staphylococcus aureus*. This is the first reported evidence that the system is controlled by allosteric inhibitors and supports new views in which modulators may stabilize preexisting intrinsic conformations rather than inducing completely new domain arrangements (Chen et al. 2010) (Fig. 4).

Furthermore, realistic models may by derived by incorporating additional information about the system in question, for example known distance constraints. Techniques that provide local distance and angle information, for example Förster resonance energy transfer (FRET) (Rochel et al. 2011) and NMR (Bertini et al. 2008; Mareuil et al. 2007) may provide useful restriction in inter-domain movement and guide conformational sampling. The rigid body/torsion/Cartesian simulated annealing strategy developed by Grishaev et al. (Grishaev et al. 2005; Mittag et al. 2010) integrated both NMR and SAXS observations into a

unique synergistic method for atomistic modeling. From NMR, residual dipolar coupling (RDC) data were used to orient the symmetrically related protein domains relative to the symmetry axis of the protein core whereas translational, shape, and size information was provided by SAXS (Schwieters et al. 2010). FRET in combination with SAXS guided rigid-body modeling to aid elucidation of the structural basis of the role of DNA in the spatial organization of nuclear hormone receptors in complex co-activators (Rochel et al. 2011). Distance restraints may also be generated from simple biochemical techniques, for example site-direct mutagenesis. For example, integrated site-directed mutagenesis and SAXS combined with conformational sampling of DNA binding sites were used to determine the DNA-binding properties of mPNK (Bernstein et al. 2009) and reveal the intramolecular metal ion transfer between flexibly-linked domains of mercury ion reductase (Johs et al. 2011) (Fig. 5).

The conformational ensemble

Although exhaustive conformational sampling significantly increases the number of realistic models to be used for modeling experimental SAXS data, a single best-fit conformation may be incapable of explaining the observed SAXS profile. The lack of convergence of a single best-fit conformation has been shown to correlate with conformational disorder rather than a limitation of the search space algorithm (Pelikan et al. 2009). In the case of scattering from a heterogeneous population, the measured scattering is derived from the population-weighted thermodynamic ensemble, and the interpretation of dynamic systems requires analysis beyond "best fit" conformations (Figs. 4 and 5). In recent years, new SAXS modeling techniques have been developed to describe dynamic systems in terms of ensembles of structures (Bernado et al. 2007; Pelikan et al. 2009; Rozycki et al. 2011; Yang et al. 2010). Four promising approaches for modeling the ensemble are pushing SAXS into an exciting new direction, the ensemble optimization method (EOM) (Bernado et al. 2007), minimal ensemble search (MES) (Pelikan et al. 2009), ensemble refinement of SAXS (EROS) (Rozycki et al. 2011), and basis-set supported by SAXS (BSS-SAXS) (Yang et al. 2010). Because of the nearly infinite number of conformations that can be adopted by flexible proteins in silico, obtaining meaningful models requires the development of robust statistical approaches that determine the probability a particular multi-conformational equilibrium will exist (Bertini et al. 2010). Again, a common problem with multi-conformational analysis is over-fitting, which occurs when an ensemble model describes noise or aggregation in the experimental system, rather than the desired underlying

relationship. MES avoids over-fitting by asserting the minimum number of states that could be distinguished from SAXS data. In addition, to avoid over-fitting the data with the multiple conformations (Bernado et al. 2007), a quantitative description of the ensemble also requires the weighting of each conformer's distribution (Pelikan et al. 2009; Yang et al. 2010). For the purpose of avoiding over-fitting of raw data, Rozycki et al. constructed a pseudo free energy scheme to refine the statistical weights attributed to configurations generated by simulation (Rozycki et al. 2011). These SAXS ensemble methods seem enormously successful on the basis of analysis of several key biological systems: identification of the correct subunit positions for full-length Ku (Hammel et al. 2010b), demonstration of the flexibility in full-length polynucleotide kinase (Bernstein et al. 2009), establishment of the configurational space of Lys-63 linked tetraubiquitin (Datta et al. 2009), elucidation of the flexibility mode in a Ubiquitin-PCNA complex involved in DNA replication and repair (Tsutakawa et al. 2011), and describing the partially unfolded state of XRCC4 (Hammel et al. 2010a) and XRCC4-likes proteins (Hammel et al. 2011).

Conclusions and prospects

Structural biology now recognizes that partially populated states are crucial to biological function. The single conformation description of a macromolecule is only a snapshot of a macromolecular ensemble. We have seen that integrative methods that utilize NMR and MX with SAXS are proving to be essential for providing a larger description of the macromolecular ensemble. Using SAXS data as a source of experimental restraints for modeling macromolecular flexibility is an exciting and relatively underdeveloped discipline. SAXS data can provide important experimental feedback, and can be extended to include dynamic conformational changes characterized by time-resolved experiments. Time-resolved measurements require very high X-ray flux and fast detectors designed for rapid electronic shuttering. Both are now available, and SAXS, unlike traditional NMR and fluorescence experiments, is not affected by molecular rotation times, so time-resolved SAXS can be performed in an equivalent manner to the traditional static experiments. The development of the approaches for characterizing highly fluctuating conformational equilibria on the basis of traditional static experiments are becoming essential in the description of intrinsic dynamic biomolecular systems (Bernado and Blackledge 2010). Macromolecular machines with flexible and unstructured regions are now tractable to direct structural investigation (Bernado 2010; Bernado and Svergun 2012). These are some of the reasons why SAXS-based

solution structure modeling of flexible macromolecular assemblies are gaining popularity and will be used in the future to elucidate the roles of dynamic equilibrium in biological processes (Rambo and Tainer 2011). A natural complement to the global shape and conformation from SAXS will be residue-level information from advancing techniques of enhanced hydrogen–deuterium exchange mass spectrometry, which can approach single-residue resolution as shown for the photocycle changes of photoactive yellow proteins (Brudler et al. 2006). Thus, SAXS is well positioned to become an important technique, with new weak-field aligned NMR and fluorescence experiments that can probe samples in the biologically interesting millisecond time frame. With appropriate resources for directed efforts, SAXS can provide complementary experimental data on flexibility in macromolecular interactions with widespread effects.

References

Abe Y, Gor J, Bracewell DG, Perkins SJ, Dalby PA (2010) Masking of the Fc region in human IgG4 by constrained X-ray scattering modelling: implications for antibody function and therapy. Biochem J 432:101–111

Akiyama S, Fujisawa T, Ishimori K, Morishima I, Aono S (2004) Activation mechanisms of transcriptional regulator CooA revealed by small-angle X-ray scattering. J Mol Biol 341:651–668

Aslam M, Perkins SJ (2001) Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. J Mol Biol 309:1117–1138

Bernado P (2010) Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. Eur Biophys J 39:769–780

Bernado P, Blackledge M (2010) Structural biology: proteins in dynamic equilibrium. Nature 468:1046–1048

Bernado P, Svergun DI (2012) Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Mol BioSyst 8:151–167

Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129:5656–5664

Bernstein NK, Hammel M, Mani RS, Weinfeld M, Pelikan M, Tainer JA, Glover JN (2009) Mechanism of DNA substrate recognition by the mammalian DNA repair enzyme, Polynucleotide Kinase. Nucleic Acids Res 37:6161–6173

Bertini I, Calderone V, Fragai M, Jaiswal R, Luchinat C, Melikian M, Mylonas E, Svergun DI (2008) Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12. J Am Chem Soc 130:7011–7021

Bertini I, Giachetti A, Luchinat C, Parigi G, Petoukhov MV, Pierattelli R, Ravera E, Svergun DI (2010) Conformational space of flexible biological macromolecules from average data. J Am Chem Soc 132:13553–13558

Biersmith BH, Hammel M, Geisbrecht ER, Bouyain S (2011) The immunoglobulin-like domains 1 and 2 of the protein tyrosine phosphatase LAR adopt an unusual horseshoe-like conformation. J Mol Biol 408:616–627

Boehm MK, Woof JM, Kerr MA, Perkins SJ (1999) The Fab and Fc fragments of IgA1 exhibit a different arrangement from that in IgG: a study by X-ray and neutron solution scattering and homology modelling. J Mol Biol 286:1421–1447

Brudler R, Gessner CR, Li S, Tyndall S, Getzoff ED, Woods VL Jr (2006) PAS domain allostery and light-induced conformational changes in photoactive yellow protein upon I2 intermediate formation, probed with enhanced hydrogen/deuterium exchange mass spectrometry. J Mol Biol 363:148–160

Chen H, Ricklin D, Hammel M, Garcia BL, McWhorter WJ, Sfyroera G, Wu YQ, Tzekou A, Li S, Geisbrecht BV, Woods VL Jr, Lambris JD (2010) Allosteric inhibition of complement function by a staphylococcal immune evasion protein. Proc Natl Acad Sci USA 107:17621–17626

Cotner-Gohara E, Kim IK, Hammel M, Tainer JA, Tomkinson AE, Ellenberger T (2010) Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. Biochemistry 49:6165–6176

Curtis JE, Raghunandan S, Nanda H, Krueger S (2012) SASSIE: a program to study intrinsically disordered biological molecules and macromolecular ensembles using experimental scattering restraints. Comput Phys Commun 183:382–389

Dainese E, Sabatucci A, van Zadelhoff G, Angelucci CB, Vachette P, Veldink GA, Agro AF, Maccarrone M (2005) Structural stability of soybean lipoxygenase-1 in solution as probed by small angle X-ray scattering. J Mol Biol 349:143–152

Datta AB, Hura GL, Wolberger C (2009) The structure and conformation of Lys63-linked tetraubiquitin. J Mol Biol 392:1117–1124

Devarakonda S, Gupta K, Chalmers MJ, Hunt JF, Griffin PR, Van Duyne GD, Spiegelman BM (2011) Disorder-to-order transition underlies the structural basis for the assembly of a transcriptionally active PGC-1alpha/ERRgamma complex. Proc Natl Acad Sci USA 108:18678–18683

Duda DM, Borg LA, Scott DC, Hunt HW, Hammel M, Schulman BA (2008) Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. Cell 134:995–1006

Fierobe HP, Bayer EA, Tardif C, Czjzek M, Mechaly A, Belaich A, Lamed R, Shoham Y, Belaich JP (2002) Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. J Biol Chem 277:49621–49630

Fink AL (2005) Natively unfolded proteins. Curr Opin Struct Biol 15:35–41

Forster F, Webb B, Krukenberg KA, Tsuruta H, Agard DA, Sali A (2008) Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies. J Mol Biol 382:1089–1106

Gilbert HE, Asokan R, Holers VM, Perkins SJ (2006) The 15 SCR flexible extracellular domains of human complement receptor

type 2 can mediate multiple ligand and antigen interactions. J Mol Biol 362:1132–1147

Grishaev A, Wu J, Trewhella J, Bax A (2005) Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. J Am Chem Soc 127:16621–16628

Grishaev A, Guo L, Irving T, Bax A (2010) Improved fitting of solution X-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. J Am Chem Soc 132:15484–15486

Hammel M, Kriechbaum M, Gries A, Kostner GM, Laggner P, Prassl R (2002) Solution structure of human and bovine beta(2)-glycoprotein I revealed by small-angle X-ray scattering. J Mol Biol 321:85–97

Hammel M, Fierobe HP, Czjzek M, Finet S, Receveur-Brechot V (2004a) Structural insights into the mechanism of formation of cellulosomes probed by small angle X-ray scattering. J Biol Chem 279:55985–55994

Hammel M, Walther M, Prassl R, Kuhn H (2004b) Structural flexibility of the N-terminal beta-barrel domain of 15-lipoxygenase-1 probed by small angle X-ray scattering. Functional consequences for activity regulation and membrane binding. J Mol Biol 343:917–929

Hammel M, Fierobe HP, Czjzek M, Kurkal V, Smith JC, Bayer EA, Finet S, Receveur-Brechot V (2005) Structural basis of cellulosome efficiency explored by small angle X-ray scattering. J Biol Chem 280:38562–38568

Hammel M, Nemecek D, Keightley JA, Thomas GJ Jr, Geisbrecht BV (2007a) The Staphylococcus aureus extracellular adherence protein (Eap) adopts an elongated but structured conformation in solution. Protein Sci 16:2605–2617

Hammel M, Sfyroera G, Ricklin D, Magotti P, Lambris JD, Geisbrecht BV (2007b) A structural basis for complement inhibition by Staphylococcus aureus. Nat Immunol 8:430–437

Hammel M, Yu Y, Fang S, Lees-Miller SP, Tainer JA (2010a) XLF regulates filament architecture of the XRCC4.ligase IV complex. Structure 18:1431–1442

Hammel M, Yu Y, Mahaney BL, Cai B, Ye R, Phipps BM, Rambo RP, Hura GL, Pelikan M, So S, Abolfath RM, Chen DJ, Lees-Miller SP, Tainer JA (2010b) Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. J Biol Chem 285:1414–1423

Hammel M, Rey M, Yu Y, Mani RS, Classen S, Liu M, Pique ME, Fang S, Mahaney BL, Weinfeld M, Schriemer DC, Lees-Miller SP, Tainer JA (2011) XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. J Biol Chem 286: 32638–32650

Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. Nature 450:964–972

Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL 2nd, Tsutakawa SE, Jenney FE Jr, Classen S, Frankel KA, Hopkins RC, Yang SJ, Scott JW, Dillard BD, Adams MW, Tainer JA (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods 6:606–612

Iyer RR, Pohlhaus TJ, Chen S, Hura GL, Dzantiev L, Beese LS, Modrich P (2008) The MutSalpha-proliferating cell nuclear antigen interaction in human DNA mismatch repair. J Biol Chem 283:13310–13319

Jain R, Hammel M, Johnson RE, Prakash L, Prakash S, Aggarwal AK (2009) Structural insights into yeast DNA polymerase delta by small angle X-ray scattering. J Mol Biol 394:377–382

Johs A, Harwood IM, Parks JM, Nauss RE, Smith JC, Liang L, Miller SM (2011) Structural characterization of intramolecular Hg(2+) transfer between flexibly linked domains of mercuric ion reductase. J Mol Biol 413:639–656

Khan S, Gor J, Mulloy B, Perkins SJ (2010) Semi-rigid solution structures of heparin by constrained X-ray scattering modelling: new insight into heparin-protein complexes. J Mol Biol 395:504–521

Koch MH, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys 36: 147–227

Li K, Gor J, Perkins SJ (2010) Self-association and domain rearrangements between complement C3 and C3u provide insight into the activation mechanism of C3. Biochem J 431: 63–72

Lim HS, Kim JS, Park YB, Gwon GH, Cho Y (2011) Crystal structure of the Mre11–Rad50-ATPgammaS complex: understanding the interplay between Mre11 and Rad50. Genes Dev 25:1091–1104

Liu H, Morris RJ, Hexemer A, Grandison S, Zwart PH (2012) Computation of small-angle scattering profiles with three-dimensional Zernike polynomials. Acta Crystallogr A 68:278–285

Mareuil F, Sizun C, Perez J, Schoenauer M, Lallemand JY, Bontems F (2007) A simple genetic algorithm for the optimization of multidomain protein homology models driven by NMR residual dipolar coupling and small angle X-ray scattering data. Eur Biophys J 37:95–104

Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD (2010) Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. Structure 18:494–506

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

Nishimura N, Hitomi K, Arvai AS, Rambo RP, Hitomi C, Cutler SR, Schroeder JI, Getzoff ED (2009) Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. Science 326:1373–1379

Pascal JM, O'Brien PJ, Tomkinson AE, Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. Nature 432:473–478

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

Perkins SJ, Bonner A (2008) Structure determinations of human and chimaeric antibodies by solution scattering and constrained molecular modelling. Biochem Soc Trans 36:37–42

Perry JJ, Cotner-Gohara E, Ellenberger T, Tainer JA (2010) Structural dynamics in DNA damage signaling and repair. Curr Opin Struct Biol 20:283–294

Petoukhov MV, Svergun DI (2005) Global rigid-body modeling of macromolecular complexes against small-angle scattering data. Biophys J 89:1237–1250

Porod G (1982) General Theory. In: Glatter O, Kratky O (eds) Small angle X-ray scattering. Academic Press, London, pp 17–51

Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40:191–285

Rambo RP, Tainer JA (2010a) Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. Curr Opin Struct Biol 20:128–137

Rambo RP, Tainer JA (2010b) Improving small-angle X-ray scattering data for structural analyses of the RNA world. RNA 16:638–646

Rambo RP, Tainer JA (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod–Debye law. Biopolymers 95:559–571

Rochel N, Ciesielski F, Godet J, Moman E, Roessle M, Peluso-Iltis C, Moulin M, Haertlein M, Callow P, Mely Y, Svergun DI, Moras D (2011) Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. Nat Struct Mol Biol 18:564–570

Rozycki B, Kim YC, Hummer G (2011) SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. Structure 19:109–116

Schneidman-Duhovny D, Hammel M, Sali A (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. Nucleic Acids Res 38:W540–W544

Schwieters CD, Suh JY, Grishaev A, Ghirlando R, Takayama Y, Clore GM (2010) Solution structure of the 128 kDa enzyme I dimer from Escherichia coli and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. J Am Chem Soc 132:13026–13045

Shang W, Ivanov I, Svergun DI, Borbulevych OY, Aleem AM, Stehling S, Jankun J, Kuhn H, Skrzypczak-Jankun E (2011) Probing dimerization and structural flexibility of mammalian lipoxygenases by small-angle X-ray scattering. J Mol Biol 409:654–668

Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY, Batey RT (2010) Free state conformational sampling of the SAM-I riboswitch aptamer domain. Structure 18:787–797

Svergun D (1992) Determination of the regularization Parameter in Indirect-Transform Methods using perceptual criteria. J. Appl. Cryst. 25:495–503

Svergun D, Barberato C, Koch MHJ (1995) CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 28:768–773

Tsutakawa SE, Van Wynsberghe AW, Freudenthal BD, Weinacht CP, Gakhar L, Washington MT, Zhuang Z, Tainer JA, Ivanov I (2011) Solution X-ray scattering combined with computational modeling reveals multiple conformations of covalently bound ubiquitin on PCNA. Proc Natl Acad Sci USA 108:17672–17677

Vucetic S, Brown CJ, Dunker AK, Obradovic Z (2003) Flavors of protein disorder. Proteins 52:573–584

Williams RS, Dodson GE, Limbo O, Yamada Y, Williams JS, Guenther G, Classen S, Glover JN, Iwasaki H, Russell P, Tainer JA (2009) Nbs1 flexibly tethers Ctp1 and Mre11–Rad50 to coordinate DNA double-strand break processing and repair. Cell 139:87–99

Williams GJ, Williams RS, Williams JS, Moncalian G, Arvai AS, Limbo O, Guenther G, SilDas S, Hammel M, Russell P, Tainer JA (2011) ABC ATPase signature helices in Rad50 link nucleotide state to Mre11 interface for DNA repair. Nat Struct Mol Biol 18:423–431

Yamagata A, Tainer JA (2007) Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism. EMBO J 26:878–890

Yang S, Blachowicz L, Makowski L, Roux B (2010) Multidomain assembled states of Hck tyrosine kinase in solution. Proc Natl Acad Sci USA 107:15757–15762

Zuo X, Cui G, Merz KM Jr, Zhang L, Lewis FD, Tiede DM (2006) X-ray diffraction "fingerprinting" of DNA structure in solution for quantitative evaluation of molecular dynamics simulation. Proc Natl Acad Sci USA 103:3534–3539