Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering

Robert P Rambo[1] and John A Tainer[1,2]

Small-angle X-ray scattering (SAXS) is changing how we perceive biological structures, because it reveals dynamic macromolecular conformations and assemblies in solution. SAXS information captures thermodynamic ensembles, enhances static structures detailed by high-resolution methods, uncovers commonalities among diverse macromolecules, and helps define biological mechanisms. SAXS-based experiments on RNA riboswitches and ribozymes and on DNA–protein complexes including DNA–PK and p53 discover flexibilities that better define structure–function relationships. Furthermore, SAXS results suggest conformational variation is a general functional feature of macromolecules. Thus, accurate structural analyses will require a comprehensive approach that assesses both flexibility, as seen by SAXS, and detail, as determined by X-ray crystallography and NMR. Here, we review recent SAXS computational tools, technologies, and applications to nucleic acids and related structures.

Addresses
[1] Life Science Division, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[2] Department of Molecular Biology, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Corresponding author: Tainer, John A (jat@scripps.edu)

Current Opinion in Structural Biology 2010, **20**:128–137

This review comes from a themed issue on
Protein-nuclei acid interactions
Edited by Simon Phillips and Karolin Luger

Available online 22nd January 2010

0959-440X/$ – see front matter
Published by Elsevier Ltd.

DOI 10.1016/j.sbi.2009.12.015

Introduction

Dynamic macromolecular assemblies composed of nucleic acids and proteins integrate the environmental and cellular signals into biological actions. Yet as of 2008, only 5% of the total structures deposited in the PDB were of nucleic or protein–nucleic acid composition, and most macromolecules function in complexes with an average of 5 partners per protein. Macromolecular X-ray crystallography (MX), Nuclear Magnetic Resonance (NMR), and electron microscopy (EM) are our most reliable structural tools; nonetheless, these techniques have limitations for macromolecules with functional flexibility and intrinsic disorder, which occurs in functional regions and interfaces [1,2]. Therefore, as structural biology evolves, it will need to provide structural insights into larger macromolecular assemblies that display dynamics, flexibility, and disorder.

Solution structures from X-ray scattering

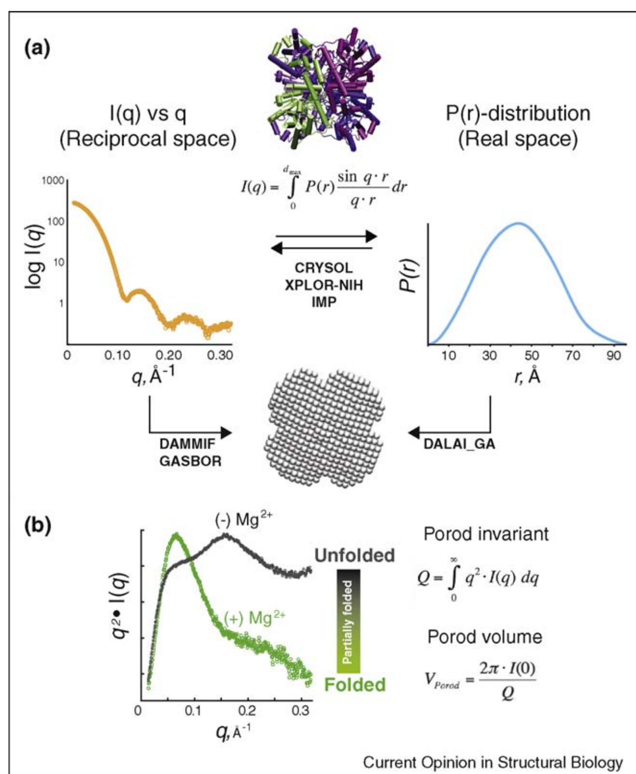
For noncoding functional RNAs and intrinsically disordered proteins, defining their shapes and conformational space in solution marks a critical step toward understanding their functional roles. Facilitating this goal are *ab initio* bead-modeling algorithms for interpreting small-angle X-ray scattering (SAXS) data. These *ab initio* models represent low-resolution shapes and contribute significantly to interpretation of flexible systems in solution, particularly for protein–DNA complexes. SAXS is proving to be increasingly useful for studying RNA and other nucleic acids, which scatter five times more strongly than proteins.

Thus, in recent years, the 'old' but *with advanced computing and synchrotron technologies* 'new' technique of SAXS has developed into a powerful structural tool that complements MX and other high-resolution methods [3,4]. SAXS can be performed under many solution conditions, and is the only structural technique that can be performed under near native conditions with minimal amounts of sample (15 μL at 0.1–1 mg/mL). Advances in X-ray sources, detectors and collection strategies have improved SAXS measurements and reduced the time from days to hundreds of milliseconds at synchrotron facilities supporting biological investigations [5••]. The now efficient combination of SAXS with high-resolution MX and NMR is providing insights into functional conformations of dynamic assemblies in solution and thereby fundamentally changing the way we perceive structure–function relationships [2,3].

Ins and outs of SAXS (basic theory)

Under sufficiently dilute conditions, the SAXS profile, $I(q)$, represents the simultaneous scattering measurement of the macromolecule in all orientations. This inherently reduces the resolution, whereas in MX the scattering from the ordered macromolecules produces diffraction intensities that are subsequently transformed to an electron density map. In SAXS, the transformation of the scattering data, $I(q)$ yields the $P(r)$-distribution, a histogram of the interatomic vectors within the macromolecule

Figure 1



SAXS provides accurate information on folding and conformation in solution for both rigid and flexible macromolecules and their complexes. **(a)** The calculation of a SAXS profile from an atomic model (CRYSOL, XPLOR-NIH, or IMP) is defined by the Debye equation and defines a relationship between the reciprocal space SAXS observations (orange) and the real space $P(r)$-distribution (cyan). Data in either reciprocal or real space can be used as a target for bead modeling with such programs as DAMMIN/F [46•], DALAI_GA [47] or GASBOR [48]. Glucose isomerase is presented as an example above. **(b)** Kratky plot of the lysine riboswitch in the presence and absence of Mg^{2+} [14]. The $q^2 \cdot I(q)$ versus q plot demonstrates, for a folded particle, a parabolic convergence of the data at high q. Such a property fulfills the Porod invariant, Q, where the area under the curve approaches a constant value allowing for the calculation of the Porod volume. This is not true for an unfolded particle, as its scattering curve does not capture a defined area in a Kratky plot.

(Figure 1a). This relationship makes it possible to directly calculate a SAXS profile given an atomic resolution macromolecular structure; however, the reverse is not true. The atomic details of an underlying macromolecule are not uniquely determined from its experimental SAXS profile alone.

The transformation of the SAXS profile, $I(q)$, into the $P(r)$-distribution has been the subject of continued development. The relationship illustrated in Figure 1a takes data measured in reciprocal space (Å^{-1}) and transforms it to real space (Å). Several computational methods have been developed [6–9] where the final result of interest is typically an estimate of the maximum dimension of the particle ($d_{\max}$). Resolution in SAXS is observed as additional features in a $P(r)$-distribution [10], a result of data collected at greater scattering angles, q. In all cases, an inspection of the $P(r)$-distribution helps insure a correct choice of $d_{\max}$ has been made where the distribution is expected to be smooth with minimal oscillations toward $d_{\max}$.

A SAXS profile allows direct estimation of the radius-of-gyration (R_g) within the Guinier region of the data ($q \cdot R_g < 1.3$). R_g describes the mass distribution of a macromolecule about its center of gravity. An increase in R_g is generally consistent with an opening of the macromolecule whereas a decrease in R_g suggests compaction. Furthermore, for well-folded samples, as shown by the Kratky plot (Figure 1b), the SAXS data provide the hydrated volume (Porod Volume) of the scattering particle [11]. These SAXS parameters represent quantitative assessments derived from a SAXS profile; nonetheless, SAXS can monitor structural changes in a semi-quantitative manner where comparative changes in a Kratky plot can reveal flexibility, unfolding or a conformational change (Figure 1b) [3].

Bridging the solution divide

A direct application of SAXS is to test the validity of X-ray crystal structures for functionally important macromolecular conformations and assemblies in solution. Crystal packing forces and subsequent cryogenic temperatures needed for MX experiments constrain the structural ensemble and beg the question as to how well the crystal structure models the solution conformation including possible multiple conformational states. Similarly, a crystal structure may be obtained only with a bound ligand, raising the question as to whether the crystal structure changes in the *apo* form. Moreover, the choice of the assembly state and biologically relevant interfaces cannot be determined by the crystal structure alone. SAXS as a solution technique is ideally suited to address these structural questions and complements the high-resolution information obtained by MX and NMR.

SAXS results showed that a relatively small interface seen in crystal structures of the DNA double strand break repair (DSBR) Mre11 nuclease bound to DNA was a dimer interface in solution rather than simply a crystal contact [12••]. Moreover combined SAXS and mutational analyses showed that this Mre11 dimer is critical to DNA binding and Mre11 function *in vivo* [12••]. Similarly, Vav1 plays an important role in T-cell activation and tumorigenesis for the guanine nucleotide exchange factor (GER) superfamily, but its conformation and domain contacts were different in MX and EM structures. The agreement between the Vav1 crystal structure and *in solution* SAXS data showed that the compact Vav/Rac conformation and unusual contacts among the regulatory

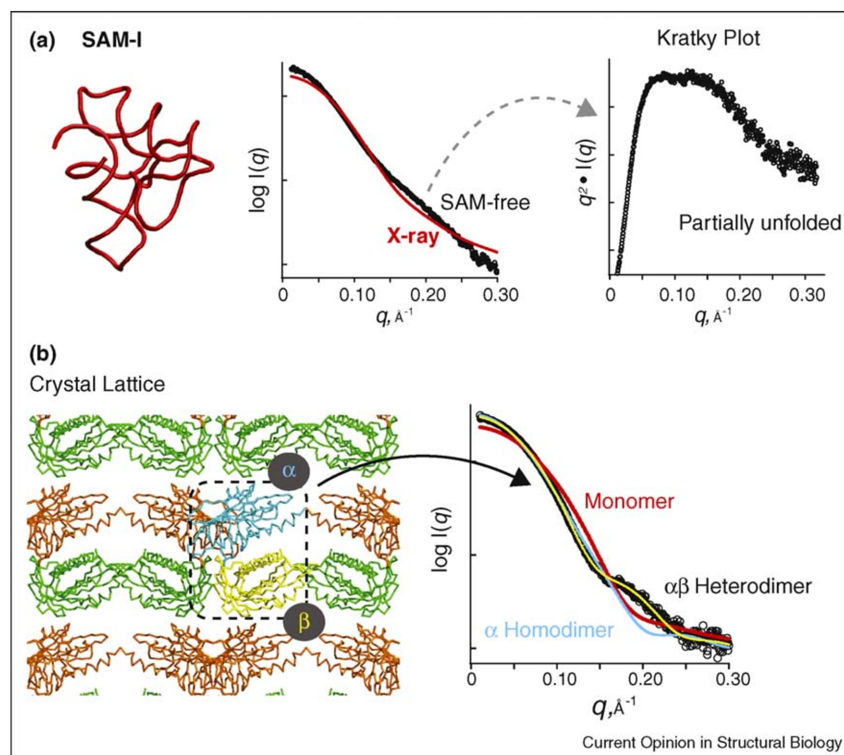
domains are biologically relevant, rather than reflecting crystal packing [13].

Detecting conformational switching as a control of biological function

The *T. maritima* lysine-responsive riboswitch, a 161-nucleotide regulatory RNA element found in the 5′UTR UTR of several lysine biosynthetic genes, binds the amino acid lysine and regulates gene expression. Comparison of the crystal structures with the SAXS data of the bound and free states of the riboswitch demonstrated that the riboswitch does not adopt a different conformation in any state suggesting that the functional mechanism of transcriptional regulation does not occur through large-scale conformational switching of the lysine binding domain [14]. This is in stark contrast to a similar study on the SAM-I riboswitch, where the crystal structure differed significantly from its free form in solution (RT Batey, unpublished data) (Figure 2A). Inspection of the SAM-I Kratky plot indicated an increase in the conformational dynamics of the RNA in the absence of ligand. Here, the solution SAXS studies guided the biochemistry developing a hypothesis reflecting conformational switching as an underlying mechanism for the riboswitch. This conversion of a flexible unbound state that can open and close into a less flexible, closed-bound state is emerging as a common feature for ligand recognition for many macromolecular systems ranging from RNA to proteins.

The SAXS profile is principally determined by the shape of the underlying $P(r)$-distribution, which can determine the correct oligomeric isoform in solution. The plant abscisic acid (ABA) hormone receptor PYR1 crystallized with distinct ‘open-lid’ and ‘closed-lid’ conformations. Crystal contacts show three possible dimer assemblies: symmetric open, symmetric closed, or asymmetric open–closed [15]. While a calibrated size-exclusion column or multiangle light scattering system could easily reject the monomeric form in solution, neither technique could determine the correct dimeric species. Here, SAXS identified the asymmetric dimer as the correct assembly in solution (Figure 2b) [15] and further SAXS studies of the receptor in the presence of ABA revealed a more compact dimer assembly not observed in the crystals.

Figure 2



SAXS results suggest that flexibility in *apo*-receptors is reduced upon ligand binding for many RNA and protein receptors. **(a)** Fit of the calculated (red) SAM-I riboswitch SAXS profile to the experimental (black) SAXS data, SAM-free. The crystal structure of SAM-I [49] was determined in the presence of ligand and the poor fit of the data to the model implies an alternate state of the riboswitch in solution. Further inspection of the Kratky plot demonstrates partial unfolding or increased flexibility. **(b)** Identification of the correct dimer model using SAXS. The three possible dimers inferred from the crystal packing of the PYR1 protein are designated α–α (orange and blue), β–β (yellow and green), and α–β (blue and yellow) [15]. SAXS clearly distinguishes the correct asymmetric (yellow curve) from incorrect symmetric (blue) dimer assembly in solution.

Architectural flexibility and rigidity in biological outcomes

SAXS can be readily combined with biochemical and mutational information to define flexible complexes, despite limited resolution. In MX, flexible regions may contribute to an incomplete structural model because of missing electron density. In SAXS, flexible regions will contribute to the observed intensity giving a complete assessment of the entire macromolecule. Consequently, SAXS measurements from a macromolecule with an incomplete high-resolution structure provide an opportunity to extend the modeling of the NMR-derived or MX-derived model [16•] through computational strategies. Thus, site-directed mutagenesis and SAXS combined with computational approaches characterized the conformational variability and DNA-binding properties of mammalian polynucleotide kinase (mPNK) [17••]. The 5′-kinase and 3′-phosphatase activities of mPNK function to restore 5′-phosphate/3′-hydroxyl termini at sites of DNA damage. The flexible attachment of the FHA domain to the catalytic segment, elucidated by SAXS, enables the interactions of mPNK with diverse DNA substrates and protein partners, as required for the effective orchestration of DNA-end repair [17••].

Objective experimental examination of conformation in solution by SAXS and other means is important for understanding biological outcomes including those from defects associated with human disease. For example, structural analyses of the DNA repair proteins XPD, Mre11, and Nbs1 contributed to hypotheses elucidating how structural defects that alter architectural integrity can cause cancer and degenerative diseases [12••,18,19••].

Envisioning whole macromolecular systems

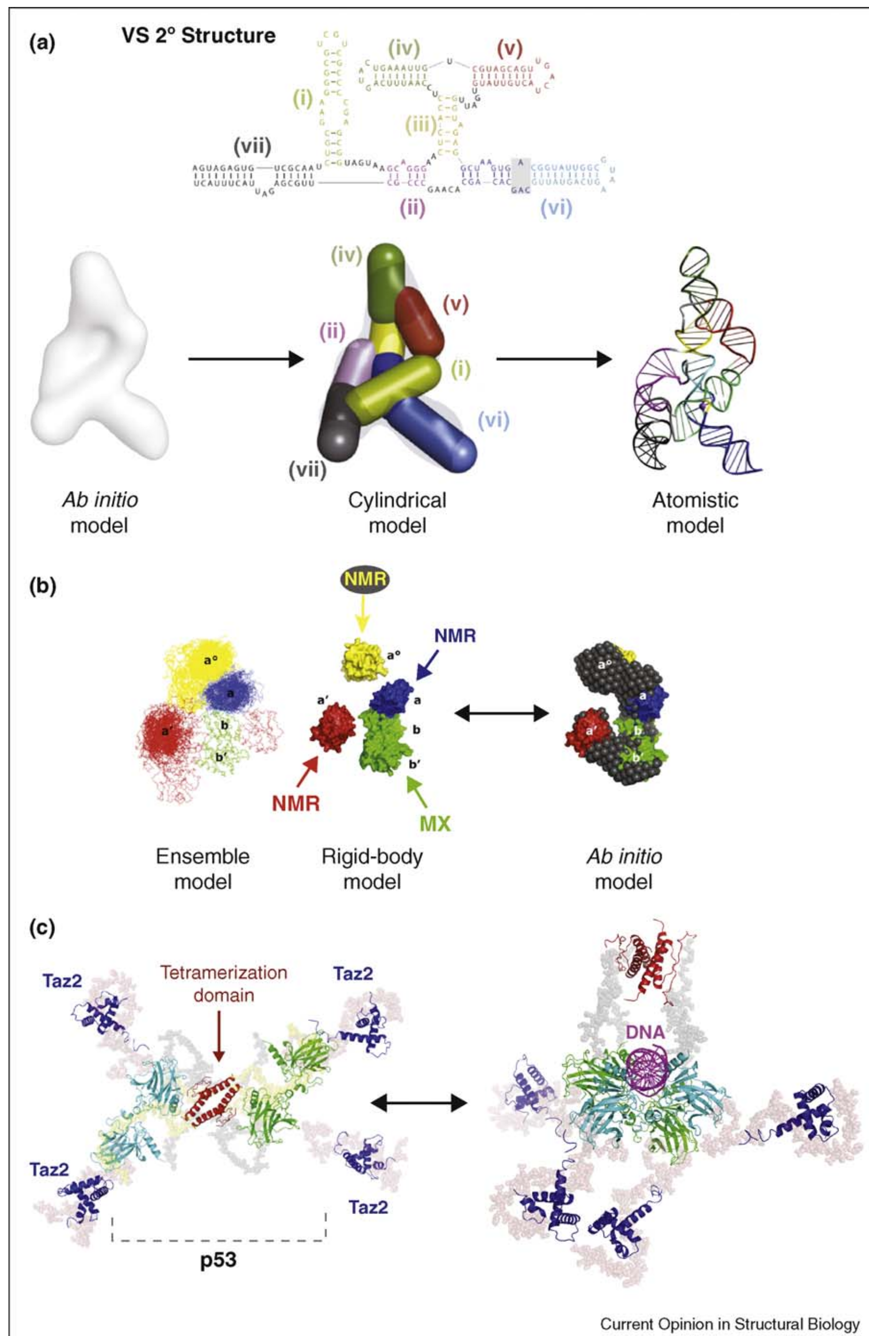
SAXS provides powerful restraints for modeling large macromolecular assemblies [20]. Particularly for folded RNA whose building block generally consists of 22 Å helical measure, a SAXS-based *ab initio* model can serve as a suitable structural framework for model building. The VS ribozyme is the largest known nucleolytic ribozyme for which there is no determined crystal structure. Decades of research have dissected the ribozyme into its component catalytic pieces, identified catalytically important residues and contacts and led to partial models of the abstract ribozyme. Lipfert *et al.* [21••] collected SAXS data on four component pieces of the ribozyme to develop a cylindrical model such that they could unambiguously assign the various helices from the VS secondary structure to the *ab initio* model [21••]. Subsequently, the assigned cylindrical models were converted to residues specific RNA helices and single-stranded nucleotides were added to link together the assigned helical regions (Figure 3a). The entire model was energy minimized using the known tertiary interactions as distance restraints with XPLOR-NIH and refined through several rounds against the SAXS profile producing a plausible structural model describing how the VS ribozyme positions the substrate strand for cleavage.

The explicit shape information within a SAXS profile is particularly useful for assembling a complete picture of a multidomain protein when only the high-resolution MX or NMR structures are known of the individual domains. SAXS data collected on the full-length ERp72, a 5-domain (a°–a–b–b′–a′) protein disulfide isomerase was used in a rigid-body refinement incorporating a°, a and a′ NMR structures with a single b–b′ MX structure [22]. In a Monte Carlo-type sampling of the starting conformations, the rigid-body modeling converged to a C-shaped structure with all three catalytic domains on the same side. Similar successes were reported for the extracellular adherence protein of *S. aureus* [23], picornaviral loop-to-loop replication complex [24], the PDZ domain of SAP97 [25] and the large heteromeric α,β-importin-cap-binding complex [26].

Incorporating SAXS data into NMR structure determination is an active area of development [27••,28•,29•]. Residual dipolar coupling (RDC) measurements from NMR provide orientational information between helical regions of proteins or nucleic acid structures and when combined with the shape information from SAXS, considerable improvements in modeling multidomain proteins and complexes can be achieved. Bax and colleagues [30•] used RDCs to deduce the correct point group for dimeric TolR thus creating a C_2-restrained dimer space. Subsequent fits to the experimental SAXS data identified the correct TolR dimer assembly. For RNA, a sparse set of RDCs combined with SAXS can be used to refine a starting homology model such as the case with $tRNA^{val}$ [27••].

In an impressive multistep deconstruction of a multiprotein DNA complex, Blackledge and colleagues [31••] were able to determine the solution structural model of the entire tumor suppressor p53 protein tetramer in complex with DNA and the Taz2 domain. p53 is an intrinsically disordered protein interspersed with two core structural domains. Unlike the previous application of RDCs, here RDCs measured along the polypeptide backbone captured the conformational ensemble of the intrinsically disordered regions. The experimental RDCs were subsequently used to further refine a preliminary SAXS-based p53–DNA complex [32]. Additional SAXS and NMR measurements of free p53, and in complex with DNA or Taz2 or both demonstrated preferred changes in the flexibility of the 175 kDa p53 tetramer and proposed a unique biological perspective of full-length p53 (Figure 3c) where the core domains of tetrameric p53 wrap around the DNA with the intrinsically disordered regions acting as outstretched arms presenting the N-terminal domain of p53 for the recruitment of additional transcription factors. SAXS captures the conformational

Figure 3

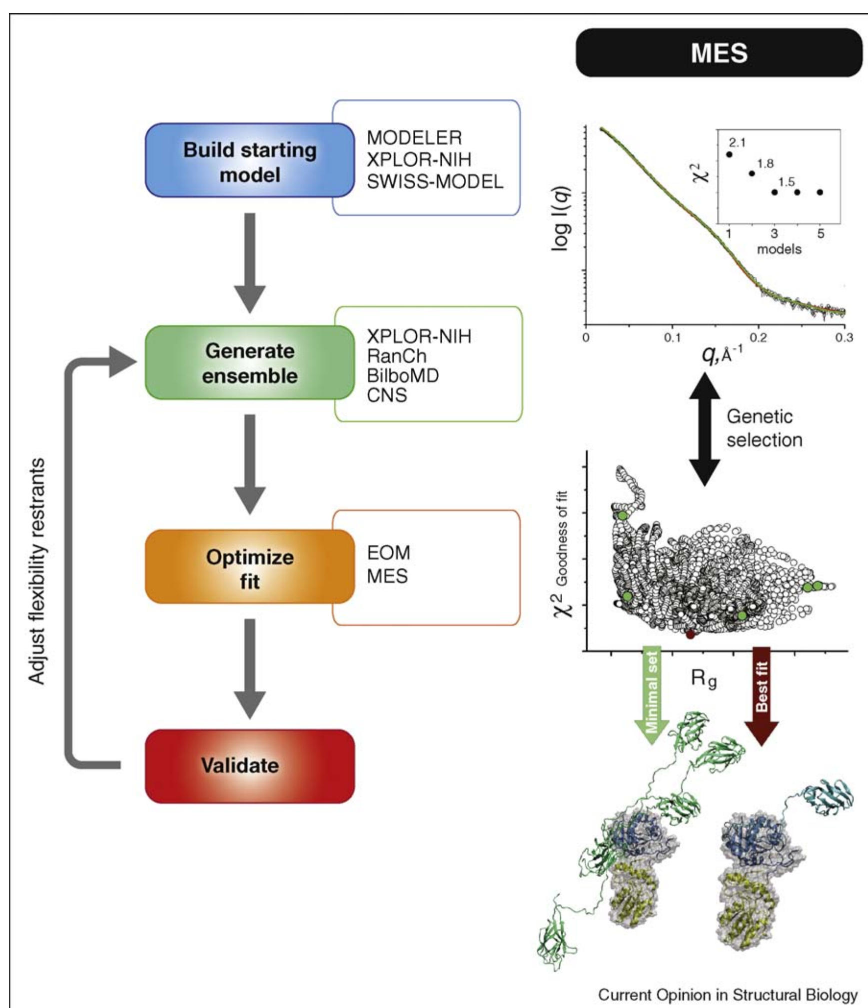


SAXS extends modeling of most large multidomain macromolecular machines and suggests that flexibility and unstructured regions are generally critical for function. **(a)** Construction of the VS ribozyme solution structure [21••] started with the *ab initio* model which was used to place cylindrical elements that corresponded to the helical elements described by its secondary structure. The resulting cylindrical model was then converted to a residue specific model and subject to cycles of energy minimization refinement to produce the final atomistic solution model of the entire VS ribozyme. **(b)** The entire ERp72 solution structure [22] was achieved by refining the SAXS data of ERp72 against a starting model composed of its individual domains that were previously determined by a combination of NMR and MX. 380 separate rigid-body modeling runs converged to show a C-shape ensemble consistent with a separate *ab initio* reconstruction. **(c)** The full-length p53–Taz2–DNA complex (right) [31••] was modeled from SAXS data of the all protein complex (left) [32]. The core (cyan and green) and tetramerization (red) domains were determined by MX and Taz2 was determined by NMR. The three high-resolution structures were incorporated into a rigid-body analysis to produce the final cross-shaped structure showing fully extended arms held together by the tetramerization domain. Flexible regions were modeled with dummy residues (gray beads). Additional SAXS data collected in the presence of DNA (magenta) was then subject to additional rigid-body modeling using an MX-derived structure of the core domain bound to DNA. The binding of Taz2 to p53 and the associated p53–DNA complex facilitated visualization of the N-terminal regions of p53, an observation not seen in a previous EM study.

ensemble and can be used as the sole restraint for modeling flexible proteins [33••]. Intrinsically disordered proteins [1,2] are a significant fraction of the eukaryotic genome and a comprehensive structural understanding of these proteins and complexes will require an integrated structural approach as described above.

For DNA machinery, a SAXS milestone was the solution structure of the multiprotein DSBR complex of DNA bound to DNA–PK, which consists of Ku70/80 (Ku) and DNA-dependent protein kinase catalytic subunit (DNA-PKcs) [33••]. DNA–PK initiates DSB repair by nonhomologous end joining (NHEJ). SAXS combined with live cell imaging and biochemical approaches helped characterize DNA–PK complex assembly and disassembly. In the Ku80 C-terminal region, a flexible arm extends from the DNA-binding core to recruit and retain DNA-PKcs at DSBs. Notably, SAXS allowed definition of the assembled DNA-PKcs dimer and of the resulting site-specific autophosphorylation that opens DNA-PKcs to promote its release from DNA ends. These SAXS results show how protein and DNA interactions regulate large rearrangements to control DNA–PK biological functions as a macromolecular machine orchestrating NHEJ

Figure 4



SAXS results suggest that conformational variation in solution is a general functional feature of macromolecular assemblies. SAXS results provide restraints for ensemble modeling that fully accounts for the scattering species in solution. The primary structure of each component is analyzed for flexible regions and regions that demarcate domain boundaries. Structures for the domains can be determined by experimental or computational (MODELER, XPLOR-NIH, SWISS-MODEL) methods and compiled into a single starting model. The initial model is used to generate the ensemble of 2000–10 000 conformations. Experimental SAXS data are then used as a constraint to select the set of members from the ensemble that best describes the data (EOM or MES). Examination of the selected set should provide testable hypotheses subject to experimental methods such as limited proteolysis or chemical probing. Additional ensembles can be generated by changing the starting conditions. In the case of MES, the example shows that the minimal number of members required to explain the data is 3 and is largely described by a structure with an extended domain in multiple conformations. The graph of χ^2 versus R_g shows the entire ensemble with the each χ^2 goodness-of-fit calculated for each member against the experimental data.

initiation. Such control of DNA product release is emerging as a general feature of DNA repair. Structures of endonuclease IV, which acts in backbone cleavage for DNA base excision repair, reveal a strained DNA substrate complex that reverts after catalysis to an ideal geometry to hold rather than release cleaved DNA product [34]. Endonuclease V substrate and product complexes with damaged DNA show DNA is wedged opened and held by a minor groove–damage sensor motif [35]. Similarly stable protein binding without DNA damage removal can sculpt damaged DNA to switch from base to nucleotide excision repair pathways [36]. These protein complexes and DNA distortions appear to have evolved to promote direct handoffs of repair pathway intermediates, and SAXS provides an optimal means for defining such complexes and conformations.

The use of SAXS in model building has been greatly facilitated by its explicit incorporation in the Integrative Modeling Platform (IMP) [37••]. IMP is a computational tool that can integrate a variety of biophysical experimental observations into spatial restraints for structural modeling [38]. For SAXS, an IMP refinement requires an initial protein structure derived from either a crystal structure or comparative homology modeling that is used in an optimization routine. IMP targets the experimental $P(r)$-distribution in a χ^2 minimization where the model with user defined rigid bodies is continuously adjusted. This is similar to X-ray crystallographic refinement where the model is built and refined against the experimental data in real space with stereochemical restraints derived from the chemical structure.

The conformational ensemble

The SAXS profile is a direct interrogation of the thermodynamic ensemble. Typical synchrotron-based SAXS measurements will measure 10 000 billion molecules simultaneously and for a protein or complex that may contain an intrinsically flexible linker between multiple domains, a single model would be grossly insufficient to explain the data. Provided an initial starting model, two promising approaches for modeling the ensemble are pushing SAXS into an exciting new direction, the Ensemble Optimization Method (EOM) [39••] and Minimal Ensemble Search (MES) [40••]. The starting model, often a multidomain *Franken-protein* composed of pieces from homology models and high-resolution NMR/MX structures (Figure 4), is constructed and the missing linkers are either modeled via a random sampling from a library of coil conformations (EOM) or created during high temperature MD simulation (MES). In both algorithms, a large sampling of the conformational space is sought (>10 000 conformations) and subject to a genetic algorithm to efficiently search for the set of models that best fits the data in reciprocal space. MES has an adaptive ensemble search algorithm that also minimizes the ensemble size during the optimization (Figure 4). These techniques appear enormously successful based upon analyses in several systems: identifying correct subunit positions for DNA pol δ [41•], demonstrating the flexibility in full-length protein kinase R [17••], establishing the configurational space of Lys-63 linked tetraubiquitin [42], and detailing the 4° structure of flavorubredoxin [43]. While these approaches sacrifice atomic details, the broader scope of knowledge gained from the SAXS modeling is often appropriate to develop hypotheses and guide biochemical experiments.

Conclusions and outlook

As structural biology advances to include larger and more flexible systems, SAXS is emerging as an all-purpose tool in the structural biologists toolkit. SAXS can characterize sample quality, provide the first structural insights into a novel complex, identify assembly states in solution, and detect functional conformational changes in the presence of substrates or ligands [3,5]. Synchrotron-based facilities now extend SAXS into the high-throughput regime where 15 μL samples at 1 mg/mL and 0.5-s exposures are practical [5••]. Nonetheless, sample quality is a paramount decider for the utility of any SAXS data and is particularly important for nucleic acids where mis-folded material may obscure the interpretation for biologically important structures [44•]. For MX, crystallization provides quality control and even improves the sample because the crystallization process itself is indeed an additional purification. For SAXS and NMR, poor quality samples that is aggregated or polydisperse, still allow data collection. We expect improvements to occur with high-throughput strategies where the exploration of a wide variety of different conditions will lead to quality SAXS data [5••].

Computational tools for analyzing and modeling SAXS data are an exciting area of development. Deconstructing a protein into its constituent stable domains for MX or NMR studies has become routine in structural biology. The combination of MX, NMR, and SAXS is a powerful strategy that defines the overall macromolecular architectures of assemblies in solution thereby providing biology with a comprehensive structure. Large multidomain macromolecular machines with flexible and unstructured regions [1] are now tractable to direct structural investigations. Likewise, a high-resolution $P(r)$-distribution may prove to be an ideal target function for protein folding algorithms thereby producing empirically validated structural models. Nonetheless, the many powerful applications of SAXS will fundamentally require an active community that continues to advance computational methods through publicly accessible databases and tools.

SAXS promises to address fundamental questions, for example induced folding versus conformational selection in events such as DNA repair, where the DNA is either captured [45] or converted into a distorted structure

[34–36]. SAXS allows modeling of large multidomain macromolecular machines and furthermore suggests that flexibility and unstructured regions are generally critical for biological functions. SAXS is showing that flexibility in *apo*-receptors is reduced upon ligand binding for many RNA and protein receptors. Moreover, SAXS suggests conformational variation is a general functional feature of macromolecules in solution; thus, strengthening the notion that accurate macromolecular analyses will require comprehensive approaches.

Acknowledgements

We sincerely thank Alan Fersht and Martin Blackledge for providing material for Figure 3. We thank Jan Lipfert, SIBYLS staff members G Hura and M Hammel for discussions and RT Batey for SAM-I. The authors acknowledge salary and other support for SAXS technologies at the SIBYLS beamline (BL12.3.1) of the Advanced Light Source at Lawrence Berkeley National Laboratory by United States Department of Energy (DOE) program Integrated Diffraction Analysis Technologies (IDAT) under contract DE-AC02-05CH11231, the Molecular Assemblies: Genes and Genomes Integrated Efficiently (MAGGIE) project (DOE; DE-FG0207ER64326), and National Cancer Institute Structural Cell Biology of DNA Repair Machines grant CA92584.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Fink AL: **Natively unfolded proteins**. *Curr Opin Struct Biol* 2005, **15**:35-41.

2. Wright PE, Dyson HJ: **Linking folding and binding**. *Curr Opin Struct Biol* 2009, **19**:31-38.

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

4. Tsuruta H, Irving TC: **Experimental approaches for solution X-ray scattering and fiber diffraction**. *Curr Opin Struct Biol* 2008, **18**:601-608.

5. •• Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL 2nd, Tsutakawa SE, Jenney FE Jr, Classen S, Frankel KA, Hopkins RC *et al.*: **Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)**. *Nat Methods* 2009, **6**:606-612.

This article presents an automated system and data analysis pipeline for SAXS data collection. The results on 50 samples indicate the relatively small amounts of material needed, the robust features of SAXS for efficient throughput, and the high percentage of samples for which structural data can be obtained.

6. Glatter O: **New method for evaluation of small-angle scattering data**. *J Appl Crystallogr* 1977, **10**:415-421.

7. Moore PB: **Small-angle scattering — information-content and error analysis**. *J Appl Crystallogr* 1980, **13**:168-175.

8. Svergun DI: **Determination of the regularization parameter in indirect-transform methods using perceptual criteria**. *J Appl Crystallogr* 1992, **25**:495-503.

9. Hansen S: **Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering**. *J Appl Crystallogr* 2000, **33**:1415-1421.

10. Hong XG, Hao Q: **High resolution pair-distance distribution function *P*(*r*) of protein solutions**. *Appl Phys Lett* 2009, **94**:.

11. Glatter O, Kratky O: *Small Angle X-ray Scattering*. London/New York: Academic Press; 1982.

12. •• Williams RS, Moncalian G, Williams JS, Yamada Y, Limbo O, Shin DS, Groocock LM, Cahill D, Hitomi C, Guenther G *et al.*: **Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair**. *Cell* 2008, **135**:97-109.

SAXS was used to determine the correct crystallographic dimer in solution and to further test the residues involved in the Mre11 dimerization interface. SAXS of the Mre11 mutants demonstrated a mixed monomer–dimer population suggesting disruption of the interface can be achieved by the incorporation of charged residues in the hydrophobic interface.

13. Chrencik JE, Brooun A, Zhang H, Mathews II, Hura GL, Foster SA, Perry JJ, Streiff M, Ramage P, Widmer H *et al.*: **Structural basis of guanine nucleotide exchange mediated by the T-cell essential Vav1**. *J Mol Biol* 2008, **380**:828-843.

14. Garst AD, Heroux A, Rambo RP, Batey RT: **Crystal structure of the lysine riboswitch regulatory mRNA element**. *J Biol Chem* 2008, **283**:22347-22351.

15. Nishimura N, Hitomi K, Arvai AS, Rambo RP, Hitomi C, Cutler SR, Schroeder JI, Getzoff ED: **Structural mechanism of abscisic acid binding and signaling by dimeric PYR1**. *Science* 2009. 1181829.

16. • Mills RD, Trewhella J, Qiu TW, Welte T, Ryan TM, Hanley T, Knott RB, Lithgow T, Mulhern TD: **Domain organization of the monomeric form of the Tom70 mitochondrial import receptor**. *J Mol Biol* 2009, **388**:1043-1058.

The X-ray crystal structure of Tom70 crystallized as a dimer where each monomer was composed of discrete N-terminal and C-terminal domains. The linker region between the two domains was missing from the electron density map, which consequently made the assignment of the correct domain topology ambiguous. Comparison of the SAXS profile with the two plausible models resolved the ambiguity and a further *ab initio* shape reconstruction provided additional support for the correct choice.

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

SAXS combined with advanced computational approaches were employed to characterize the SAXS to define the shapes and conformations of mPNK complexes including their functional flexibility. On the basis of several thousand conformations of mPNK and their calculated SAXS profiles, the SAXS validated best-fitting model had the FHA linker in an extended conformation. To test the position of DNA predicted from SAXS envelopes, the authors performed an exhaustive search of the protein–DNA interfaces by an EMAP docking approach, an algorithm that searches for the minimal ensemble (MES) of the conformations from the pool of all generated conformations in MD simulations. The combined envelopes for DNA complexes and flexible attachment of the FHA domain to the catalytic segment, elucidated by SAXS, suggests how flexibility enables the interactions of mPNK with diverse DNA substrates and protein partners required for effective orchestration of DNA-end repair.

18. Fan L, Fuss JO, Cheng QJ, Arvai AS, Hammel M, Roberts VA, Cooper PK, Tainer JA: **XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations**. *Cell* 2008, **133**:789-800.

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

Nbs1 links the Mre11 nuclease at a DNA double strand break to the CTIP protein that opens DNA surrounding the break. MX, SAXS, and proteolysis results show that the Nbs1 folded core is conserved from pombe to human Nbs1 and that disease-associated Nbs1 mutations fall into the folded and extended N-terminal FHA-BRCT1-BRCT2 domains defined by SAXS and MX. As all parts of a structure contribute to the experimental X-ray scattering, SAXS was used to provide an objective, quantitative measure of structural similarity for pombe and human Nbs1 and to provide molecular insights into the consequences of patient-derived NBS mutations. SAXS shows that the fused, extended, FHA-BRCT1-BRCT2 domains are flexibly linked to C-terminal Mre11-binding and ATM-binding motifs. With the SAXS results, Nbs1 emerges as an extended, flexible binding nexus coordinating interactions of the Mre11-Rad50 core, the checkpoint kinase ATM/Tel1, and the DNA-end-processing factor CTIP.

20. Petoukhov MV, Svergun DI: **Analysis of X-ray and neutron scattering from biomacromolecular solutions**. *Curr Opin Struct Biol* 2007, **17**:562-571.

21. Lipfert J, Ouellet J, Norman DG, Doniach S, Lilley DM: **The complete VS ribozyme in solution studied by small-angle X-ray scattering**. *Structure* 2008, **16**:1357-1367.
••
Building structural models into electron density usually occurs by interpreting an electron density map in terms of the amino or nucleic acid building blocks. Here, the authors interpret a density map in terms of larger helical building blocks and build the entire ribozyme using an *ab initio* SAXS model.

22. Kozlov G, Maattanen P, Schrag JD, Hura GL, Gabrielli L, Cygler M, Thomas DY, Gehring K: **Structure of the noncatalytic domains and global fold of the protein disulfide isomerase ERp72**. *Structure* 2009, **17**:651-659.

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

24. Claridge JK, Headey SJ, Chow JY, Schwalbe M, Edwards PJ, Jeffries CM, Venugopal H, Trewhella J, Pascal SM: **A picornaviral loop-to-loop replication complex**. *J Struct Biol* 2009, **166**:251-262.

25. Goult BT, Rapley JD, Dart C, Kitmitto A, Grossmann JG, Leyland ML, Lian LY: **Small-angle X-ray scattering and NMR studies of the conformation of the PDZ region of SAP97 and its interactions with Kir2.1**. *Biochemistry* 2007, **46**:14117-14128.

26. Dias SM, Wilson KF, Rojas KS, Ambrosio AL, Cerione RA: **The molecular basis for the regulation of the cap-binding complex by the importins**. *Nat Struct Mol Biol* 2009, **16**:930-937.

27. Grishaev A, Ying J, Canny MD, Pardi A, Bax A: **Solution structure of tRNAVal from refinement of homology model against residual dipolar coupling and SAXS data**. *J Biomol NMR* 2008, **42**:99-109.
••
Starting with a $tRNA^{phe}$ based homology model, experimental SAXS data and residual dipolar coupling measurements of $tRNA^{val}$ are used to refine the starting homology model generating a final solution model of $tRNA^{val}$. The authors present a general strategy for future combined investigation utilizing both SAXS and RDCs.

28. Wang J, Zuo X, Yu P, Byeon IJ, Jung J, Wang X, Dyba M, Seifert S, Schwieters CD, Qin J *et al.*: **Determination of multicomponent protein structures in solution using global orientation and shape restraints**. *J Am Chem Soc* 2009, **131**:10507-10515.
•
Correct placement and orientation of individual protein components in a large complex is a challenging problem in folding and EM maps. The authors combine RDCs and SAXS data to develop a straightforward approach for assembling multicomponent structures.

29. Wang J, Zuo X, Yu P, Xu H, Starich MR, Tiede DM, Shapiro BA, Schwieters CD, Wang YX: **A method for helical RNA global structure determination in solution using small-angle X-ray scattering and NMR measurements**. *J Mol Biol* 2009, **393**:717-734.
•
Determining the global architecture of a folded RNA in terms of its helical building blocks can be readily achieved in a combined approach integrating both SAXS and NMR based residual dipolar coupling measurements. The structural periodicity of the RNA helical backbone provides unique RDC measurements for orienting helices. However, an ambiguity arises when measurements are made in one alignment medium. Here, SAXS resolves the ambiguity and provides additional independent data for refining a family of structures.

30. Parsons LM, Grishaev A, Bax A: **The periplasmic domain of TolR from *Haemophilus influenzae* forms a dimer with a large hydrophobic groove: NMR solution structure and comparison to SAXS data**. *Biochemistry* 2008, **47**:3131-3142.
•
NMR measurements encompassing both NOEs and RDCs were used to determine the solution structure of the periplasmic domain of TolR. Additional SAXS studies were used to confirm the correct dimer model derived by the NMR studies, providing an independent assessment of the model in solution.

31. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, Fersht AR: **Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain**. *Proc Natl Acad Sci U S A* 2008, **105**:5762-5767.
••
The authors use a combined approach of NMR, X-ray crystallography, and SAXS to develop a structural model of p53 in complex with DNA. They reconcile a previous EM structure of the p53–DNA complex and show a remarkable agreement with the SAXS-based model derived from rigid-body modeling. More importantly, the authors demonstrate by SAXS that the tetramerization of p53 does not occur explicitly through C-terminal or N-terminal domains. Systematic truncations of p53 were explored to define the requirements for quaternary structure formation.

32. Tidow H, Melero R, Mylonas E, Freund SM, Grossmann JG, Carazo JM, Svergun DI, Valle M, Fersht AR: **Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex**. *Proc Natl Acad Sci U S A* 2007, **104**:12324-12329.

33. Hammel M, Yu Y, Mahaney BL, Cai B, Ye R, Phipps BM, Rambo RP, Hura GL, Pelikan M, So S *et al.*: **Ku and DNA-dependent protein kinase (DNA–PK) dynamic conformations and assembly regulate DNA binding and the initial nonhomologous end joining complex**. *J Biol Chem* 2010, **285**:1414-1423.
••
The assembly states of DNA–PK were studied by SAXS in the presence of DNA. Two types of structural conversions were observed demonstrating monomeric and dimeric forms of DNA-PKcs. The results show a unique structural comparison of the *ab initio* SAXS models with previously determined EM reconstructions.

34. Garcin ED, Hosfield DJ, Desai SA, Haas BJ, Bjoras M, Cunningham RP, Tainer JA: **DNA apurinic–apyrimidinic site binding and excision by endonuclease IV**. *Nat Struct Mol Biol* 2008, **15**:515-522.

35. Dalhus B, Arvai AS, Rosnes I, Olsen OE, Backe PH, Alseth I, Gao H, Cao W, Tainer JA, Bjoras M: **Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair**. *Nat Struct Mol Biol* 2009, **16**:138-143.

36. Tubbs JL, Latypov V, Kanugula S, Butt A, Melikishvili M, Kraehenbuehl R, Fleck O, Marriott A, Watson AJ, Verbeek B *et al.*: **Flipping of alkylated DNA damage bridges base and nucleotide excision repair**. *Nature* 2009, **459**:808-813.

37. Forster F, Webb B, Krukenberg KA, Tsuruta H, Agard DA, Sali A: **Integration of small-angle X-ray scattering data into structural modeling of proteins and their assemblies**. *J Mol Biol* 2008, **382**:1089-1106.
••
Computational package designed to use SAXS explicitly as a spatial restraint for model building.

38. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT *et al.*: **Determining the architectures of macromolecular assemblies**. *Nature* 2007, **450**:683-694.

39. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI: **Structural characterization of flexible proteins using small-angle X-ray scattering**. *J Am Chem Soc* 2007, **129**:5656-5664.
••
An important advance for incorporating multiple models in the interpretation of SAXS data is the Ensemble Optimization Method (EOM). An ensemble of atomistic protein models is constructed with missing linkers computationally modeled. Efficient sampling of the conformational space employs a genetic algorithm to efficiently search for the set of models that best fits the data in reciprocal space.

40. Pelikan M, Hura GL, Hammel M: **Structure and flexibility within proteins as identified through small angle X-ray scattering**. *Gen Physiol Biophys* 2009, **28**:174-189.
••
A robust means to examine the conformational variation manifest in SAXS results. The Minimal Ensemble Search (MES) employs high temperature MD simulation, samples conformational space by a genetic algorithm, and applies an ensemble search algorithm to minimize the ensemble size during the optimization.

41. Jain R, Hammel M, Johnson RE, Prakash L, Prakash S, Aggarwal AK: **Structural insights into yeast DNA polymerase delta by small angle X-ray scattering (SAXS)**. *J Mol Biol* 2009, **394**:377-382.
•
DNA polδ is a multisubunit polymerase consisting of a catalytic core Pol3. SAXS data collected on the ternary Pol3–Pol31–Pol32 was used as an empirical constraint for structural modeling using a high-resolution structure of Pol3 and homology models derived from human orthologs. The resulting model is consistent with prior EM studies and demonstrates a model consistent with biochemistry.

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

43. Petoukhov MV, Vicente JB, Crowley PB, Carrondo MA, Teixeira M, Svergun DI: **Quaternary structure of flavorubredoxin as revealed by synchrotron radiation small-angle X-ray scattering**. *Structure* 2008, **16**:1428-1436.

44. Rambo RP, Tainer JA: **Improving small angle X-ray scattering**
• **data for structural analyses of the RNA world**. *RNA*, http://www.rnajournal.org/cgi/doi/10.1261/rna.1946310

A comprehensive examination of folding artifacts and their influences during SAXS data collection and model building. The authors offer a simple but robust approach for making reliable SAXS samples.

45. Parker JB, Bianchet MA, Krosky DJ, Friedman JI, Amzel LM, Stivers JT: **Enzymatic capture of an extrahelical thymine in the search for uracil in DNA**. *Nature* 2007, **449**:433-437.

46. Franke D, Svergun DI: **DAMMIF, a program for rapid ab-initio**
• **shape determination in small-angle scattering**. *J Appl Crystallogr* 2009, **42**:4.

A significantly enhanced version of DAMMIN that incorporates a dynamic search volume for efficient sampling of the search space for generating *ab initio* bead models.

47. Chacon P, Diaz JF, Moran F, Andreu JM: **Reconstruction of protein form with X-ray solution scattering and a genetic algorithm**. *J Mol Biol* 2000, **299**:1289-1302.

48. Svergun DI, Petoukhov MV, Koch MH: **Determination of domain structure of proteins from X-ray solution scattering**. *Biophys J* 2001, **80**:2946-2953.

49. Montange RK, Batey RT: **Structure of the *S*-adenosylmethionine riboswitch regulatory mRNA element**. *Nature* 2006, **441**:1172-1175.