

# Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

## Title

Structural analysis of flexible proteins in solution by Small Angle X-ray Scattering combined with crystallography

## Permalink

<https://escholarship.org/uc/item/27r7m28b>

## Authors

Tsutakawa, Susan E.

Hura, Greg L.

Frankel, Ken A.

et al.

## Publication Date

2006-05-25

# **Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography**

Susan E. Tsutakawa<sup>1</sup>, Greg L. Hura<sup>2</sup>, Ken A. Frankel<sup>3</sup>, Priscilla K. Cooper<sup>1</sup>, and John A. Tainer<sup>1,4</sup>

<sup>1</sup>Life Science Division, <sup>2</sup>Physical Biosciences Division and <sup>3</sup>Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

<sup>4</sup>The Scripps Research Institute, La Jolla, CA

To whom correspondence should be addressed:

John A. Tainer, Ph.D.

Professor, Department of Molecular Biology  
and The Skaggs Institute for Chemical Biology

The Scripps Research Institute  
10550 North Torrey Pines Road

Mailstop: MB4

La Jolla, CA 92037

phone: (858) 784-8119

fax: (858) 784-2289

e-mail: [jat@scripps.edu](mailto:jat@scripps.edu)

## ABSTRACT

In the last few years, SAXS of biological materials has been rapidly evolving and promises to move structural analysis to a new level. Recent innovations in SAXS data analysis allow *ab initio* shape predictions of proteins in solution. Furthermore, experimental scattering data can be compared to calculated scattering curves from the growing data base of solved structures and also identify aggregation and unfolded proteins. Combining SAXS results with atomic resolution structures enables detailed characterizations in solution of mass, radius, conformations, assembly, and shape changes associated with protein folding and functions. SAXS can efficiently reveal the spatial organization of protein domains, including domains missing from or disordered in known crystal structures, and establish cofactor or substrate-induced conformational changes. For flexible domains or unstructured regions that are not amenable for study by many other structural techniques, SAXS provides a unique technology. Here, we present SAXS shape predictions for PCNA that accurately predict a trimeric ring assembly and for a full-length DNA repair glycosylase with a large unstructured region. These new results in combination with illustrative published data show how SAXS combined with high resolution crystal structures efficiently establishes architectures, assemblies, conformations, and unstructured regions for proteins and protein complexes in solution.

## INTRODUCTION

Proteins are dynamic flexible molecules in solution, and conformational changes are often as important for protein function as the catalytic active site. Flexibility enables proteins (a) to be regulated allosterically either by post-translational modification, by small molecules or by other proteins; (b) to have multiple functions that can be coordinated; (c) to interact with multiple structurally distinct partners through the same region that can mold to different shapes; (d) to have reversible interactions through entropic disorder-to-order transitions, (e) to have increased processivity or specificity by molding to the substrate or to the interacting partner (Dyson and Wright, 2002).

Flexibility, as a characteristic of proteins, may take on many forms. Flexibility can be conferred through flexible linkers connecting rigid domains or through two domains that can move relative to each other through a pivot point. It also can include large stretches of sequence that can assume multiple conformations that may or may not have secondary structure on their own (Dyson and Wright, 2002). In this paper, we define these latter regions as unstructured. These regions become structured upon binding to substrate, post-translational modification or interaction with partners. This disorder-to-order transition either provides a structural platform for the catalytic function or promotes secondary interfaces for interactions with other proteins and propagation of a signal (Dyson and Wright, 2002).

Eukaryotic proteins are more likely than bacterial proteins to be predicted to have unstructured regions. Genomic analysis by PONDR, a disorder prediction program, predicts that 2-21% of bacterial proteins within a particular genome have long unstructured regions (>50 aa), compared to 25-41% of eukaryotic proteins (Dunker et al., 2000; Dunker et al., 2005). This increased frequency of eukaryotic proteins with disordered regions is likely correlated with the more complex architecture. It has been observed that eukaryotic complexity is not encoded so much in additional proteins but in longer gene sequences (Lander et al., 2001; Venter, 2001). Extensions added during evolution typically provide platforms for complex protein-protein interactions and post-translational modification that help coordinate enzymatic or signalling pathways. Given the regulatory nature of these regions, they are likely to be intrinsically unstructured or flexible.

Characterization of flexible proteins in solution can provide the structural framework for understanding protein function. However, high resolution techniques are not well adapted for studying flexible proteins, particularly large protein complexes, in solution. Protein crystallography provides high resolution atomic structures, and if one is lucky, conformational changes can be observed through subunits not symmetrically related in the asymmetric unit, regions of increased B values, and distinct crystals with different cofactors or partners. However, limiting domain movement by crystal packing is still a major concern for most protein crystals, and regions that are flexible are often either not visible in the electron density or have been removed to enable crystallization. Nuclear magnetic resonance (NMR) can provide atomic resolution information of proteins in solution, but it is still limited by the size of the proteins that can be routinely studied. Electron microscopy (EM), in particular cryo-EM, can provide valuable low resolution information on proteins in solution through single particle analysis. Cryo-EM can be especially powerful in combination with computational fitting of component crystal structures so that detailed information relevant to biology and function can be obtained, such as in the recent cryo-EM and crystal structures of pathogenic pili and their pilin subunits (Craig et al., 2006). However, cryo-EM requires very large protein or protein complexes (>300 kD) or unusual symmetry for effective analyses.

Small angle X-ray scattering (SAXS) is more broadly able to provide low resolution structural information for flexible proteins in solution to complement crystallographic analyses. Although SAXS is not a high resolution method, detailed knowledge about bio-macromolecules, increased computing power and recent developments in algorithms have enabled extraction of much more detailed information from SAXS data analysis (Chacon et al., 1998; Petoukhov and Svergun, 2003; Petoukhov and Svergun, 2005; Takahashi et al., 2003; Zipper and Durchschlag, 2003). SAXS is also most powerful as a hybrid method since combination with other structural information can provide

significant biological insight into protein conformations in solution. The relative ease and speed of SAXS sample preparation and data collection means that SAXS can readily be used in conjunction with other methods. Recent SAXS programs which allow the incorporation of high resolution structures as a starting point upon which further SAXS analysis may be conducted, build upon the success of protein crystallography. SAXS analysis can reveal the structures of flexible proteins, protein complexes, induced conformational changes, and unstructured regions that are not amenable to high resolution protein crystallographic methods. Furthermore SAXS can characterize the assembly, architecture, and conformational states for reversible molecular complexes that form in response to environmental stress and cell state. These dynamic or reversible complexes are critical for key decisions points in cell biology including pathway controls and coordination. Here, we discuss applications and new results of SAXS analysis to protein structure analysis in the context of informative published SAXS results including several from our own research or from associated researchers at the SIBYLS beamline located in the Advanced Light Source at Lawrence Berkeley National Laboratory.

## SAXS BASICS

When matter is irradiated by an X-ray beam, one of the primary interactions is the elastic scattering of X-rays by electrons. X-rays thus scattered by spatially correlated electron pairs constructively and destructively interfere in accordance with Bragg's law. The longer the electron pair distance, the smaller the angle at which interference effects are observable. This is readily apparent by contrasting the scattering profiles of large proteins relative to small proteins. Large proteins will scatter more strongly at smaller angles than small proteins and much more strongly than buffer. In the last case electrons are exclusively correlated on small length scales. Traditionally SAXS was defined by scattering at angles where the small angle cosine and sine approximations apply. However, practically SAXS analysis may use scattering at angles up to the primary scattering peak of water at a momentum transfer of  $\approx 2\text{\AA}^{-1}$  which thereafter overwhelms all other signals. (Momentum transfer  $s = (4\pi/\lambda)\sin(\theta/2)$  where  $\lambda$  is the wavelength of x-rays and  $\theta$  is the scattering angle relative to the incident beam.) For proteins randomly oriented in solution, all configurations of the macromolecule are sampled during the experiment and the scattering profile is azimuthally symmetric around the primary beam. Because the signal difference between buffer and protein is small, the scattering profile of the buffer solution must be subtracted from that of the protein. This also removes background due to parasitic scattering of the primary beam; a “halo” of x-rays surrounding the primary beam, and ultimately the limiting factor on the smallest angles attainable at a given instrument. The scattering profile is dependent on the size and shape of the macromolecule and is a unique “signature” of the protein. In their study of the protein data bank, Sokolova et coworkers have shown that no two unique structures have identical calculated scattering profiles (Sokolova et al., 2003a; Sokolova et al., 2003b).

The x-ray scattering from proteins is a function of their electron density. The scattering profile from a buffer-subtracted macromolecular solution is the Fourier transform of the spatially averaged auto-correlation function of the electron density (Eqn, 1) (Feigin and Svergun, 1987) where  $I$  is the intensity,  $\rho(\mathbf{r})$  is the electron

$$\text{Eqn. 1} \quad I(s) = 4\pi \int_0^{D_{\max}} \langle \rho(\mathbf{r}) * \rho(-\mathbf{r}) \rangle \frac{\sin(sr)}{sr} r^2 dr$$

$$\text{Eqn. 2} \quad p(r) = r^2 \langle \rho(\mathbf{r}) * \rho(-\mathbf{r}) \rangle$$

density,  $D_{\max}$  is the maximum dimension of the protein and  $\mathbf{r}$  is the vector describing the spatial position. The spatially averaged auto-correlation function of the electron density (also known as the Patterson function) multiplied by  $r^2$  is called the pair distribution function or  $p(r)$  and is shown in Equation 2. For particles with homogeneous electron density the  $p(r)$  plot is an electron pair histogram describing the number of electrons within a macromolecule that are a distance  $r$  away from one another. For particles with heterogeneous electron density such as DNA – protein complexes the

$p(r)$  is a histogram of distances, within the particle, weighted by the product of the electron densities at the two ends of the connecting vector. From this plot, the  $D_{\max}$ , the radius of gyration ( $R_g$ ) as well as gross features of the shape (globular, elongated, bilobal) can be determined (Svergun and Koch, 2003).

Helpful for determining the stoichiometry of complex components in the solution, the forward scattering is proportional to the number of electrons squared and thus also proportional to the molar mass squared. Practically, experimenters may determine the mass of a particle by comparing the forward scattering of a protein whose mass is known to the forward scattering of the protein in question. The concentration of protein in the solutions must also be known. This process allows the experimenters to extract the instrument dependent proportionality constant for an absolute mass determination. SAXS is less exact if there is heterogeneity in the complexes formed, but this limitation may be overcome through in-line isolation of complexes by gel filtration chromatography.

## PRACTICAL ASPECTS OF SAXS DATA COLLECTION.

Multiple beamlines are set up for SAXS of biological samples around the world, and more SAXS beamlines are in the planning stages. We manage the SIBYLS beamline at the Advanced Light Source, which is a unique resource with two endstations, one for protein crystallography and one for small angle X-ray scattering (Trame et al., 2003). The purpose of the design was to facilitate synergy between these two structural techniques, and the rapid exchange between endstations ( $< 1$  hour) enables users to do both analyses on the same trip. Another characteristic of the SIBYLS beamline for SAXS analysis is that the beam is fairly broad at the sample position (typically  $4 \times 1.5$  mm) but is focused at the detector to a  $300 \times 300$  micrometer spot. The benefit for SAXS analysis is that the sample area is maximally exposed to the beam but the beam is narrowly focused at the detector and there can be a correspondingly small beam stop. The monochromator is actually two double crystal monochromators, a set of Si(111) single crystals that provides high spectral resolution ( $\Delta\lambda/\lambda$  is less than  $10^{-4}$ ) and a set of multilayer crystals that provides lower spectral resolution ( $\Delta\lambda/\lambda$  is less than  $10^{-2}$ ) but substantially higher flux. The automated monochromator allows rapid wavelength tuning without a change in the target position, and the wavelength range is between 8 and 14 keV for the Si(111) crystal monochromator and 8 and 12 keV for the multilayer. For SAXS studies, the ability to change wavelength enables optimization of data collection for protein size. For most proteins between 30-150 kD, we routinely use approximately 1.11 Å wavelength. There are also four slits arranged from the monochromator to the sample station to collimate the beam and minimize parasitic scattering that would increase the background scatter.

The equipment required for measuring SAXS data is very similar to standard protein crystallography (PX). The predominant difference is that the beam stop and detector are often a meter or more away from the sample in order to spread the SAXS signal over a larger area for detection and for separation of the signal from the primary beam. At the SIBYLS beamline, the sample cell (developed by Hiro Tsuruta) is 1 mm in thickness with mica windows and holds 12  $\mu$ l of solution. The two dimensional detector is a MAR165 CCD detector, and the sample is separated from the detector with a 1.5 meter flight tube under vacuum to minimize air scatter. Although scattering quality is highly protein dependent, we have obtained accurate shape predictions with protein concentrations as little as 2 mg/ml (see below). Dilution curves are used to make sure that proteins remain monodisperse at high protein concentrations during data collection. The buffer also significantly scatters the X-ray, and a buffer blank is used to subtract scatter from the buffer solution as well as eliminate any systematic errors in the data. Although minimization of electron dense buffer components is advised, we generally find that protein monodispersity is a more important criterion. Thus, we have collected data on samples in high salt buffer (1 M NaCl) and in phosphate-based buffers. We have also been able to collect data on samples in detergents but have found not all detergents to be amenable to SAXS studies due to high intrinsic scattering of the detergents. More information about the beamline can be found on our web site: <http://www.bl1231.als.lbl.gov>.

One of the most powerful aspects of SAXS is that data collection is rapid. We generally take two exposures, a short exposure of 5-10 secs to obtain accurate data close to the beam stop and a long exposure of 50-100 secs for more accurate high resolution data. Most samples take between ten minutes and one hour, depending on the number of protein concentrations that are run. Since most of the time is spent in sample handling, addition of sample handling robots will greatly accelerate the process.

## APPLICATIONS OF SAXS TO STRUCTURAL ANALYSIS

*Ab initio* Shape Prediction. Given that no two proteins with unique structures in the PDB have the same calculated SAXS profile (Sokolova et al., 2003a), the combination of SAXS data and protein-based knowledge may in fact determine unique three dimensional structures at low resolution. However current SAXS *ab initio* shape prediction are more accurately low resolution (~10 Å) envelopes whose calculated scattering curves are consistent with the experimental scattering curve. Very basic protein properties are used to significantly reduce the number of acceptable three dimensional envelope that agree with the experimental scattering curve. Comparison of the *ab initio* shape predictions from experimental SAXS data of proteins in solution and their corresponding crystal structures demonstrate that *ab initio* programs can be surprisingly accurate.

Equation 1 describes how the scattering profile may be calculated from structural models of bio-macromolecules. Several *ab initio* algorithms have been written to create a low resolution envelope whose calculated scattering curve matches the experimental data (Chacon et al., 1998; Svergun, 1999; Svergun et al., 2001; Walther et al., 2000). For example, one of the most used algorithms, GASBOR, compares calculated scattering profiles from thousands of configurations of a chain of dummy residues to the measured scattering curve (Svergun et al., 2001). Simulated annealing is used with additional penalties for non-protein-like density. Typically, GASBOR is run multiple (>10) times and the results are averaged for a final shape. A comparison of individual runs may be evaluated for convergence on a final model. Often all runs converge on a similar structure and the final accuracy and precision of the experiment is determined by the finer agreement between individual runs. GASBOR may also be started from a crystal structure that includes the hydration shell or symmetry may be enforced. The accuracy of the final model is expected to vary from protein to protein depending on its shape, size and if symmetry is enforced.

We have collected and analyzed a wide variety of proteins at the SIBYLS beamline. Fig. 1 shows several examples of shape predictions determined from experimental SAXS data collected at the SIBYLS beamline using GASBOR on proteins with known structure. In these examples no information from the crystal structure was used as input to the GASBOR program. The excellent agreement between the SAXS models and the respective crystal structures demonstrates that SAXS analysis can generate accurate predictions of the overall protein envelope. These proteins differ in size, shape, and symmetry. OGG1 is a monomeric 39 kD protein that acts as a glycosylase in DNA repair. The shape prediction for OGG1 roughly follows the surface of the OGG1 crystal structure and is notably asymmetric. Glucose isomerase is a tetrameric protein of 173 kD and is involved in xylose metabolism in bacteria. This shape prediction was calculated from scattering of a 2 mg/ml protein solution and with P4 symmetry enforced. Proliferating cell nuclear antigen (PCNA) is a trimeric 86 kD protein that is involved in a number of DNA transactions such as replication and DNA repair. Information about the symmetry of the protein was not used as input to the program. In particular, the shape prediction for PCNA was remarkable in that it accurately predicted the trimeric ring structure and the hole that is formed. The hole in the crystal structure is approximately 35-36 Å across. In other protein analyses, we have found holes to be generally problematic for GASBOR. The differences between the SAXS shape predictions and the crystal structures reflect a number of parameters, including the SAXS capability of detecting flexible regions, relative increase in flexibility of proteins in solution as opposed to packed in a crystal, and the low resolution nature of the SAXS data. The latter parameter may be partially decreased by obtaining SAXS data out to higher angles (Davies et al., 2005), but there is a theoretical limit to the resolution of the SAXS data.

Understandably, there are caveats to *ab initio* shape predictions and to SAXS analysis in general. One of the areas needed for development in SAXS analysis is the confidence level in the shape reconstruction. It is well appreciated by everyone in the field that multiple different shape predictions can agree to similar extents with an experimental scattering curve. Convergence of the majority, if not all, of the shape predictions to a unique shape is indicative of a correct solution. However, the number of runs and the quality measures of the accuracy in shape predictions have varied widely in different publications involving SAXS data. A more quantitative measure of the solution uniqueness is needed. GASBOR is particularly useful when a great deal is already known about the structure, such as the multimerization state and symmetry. However, results are less reliable when Dmax, an input parameter in GASBOR, is difficult to assign. Both long flexible regions and the corruption of the monodispersed signal due to small amounts of aggregation result in difficult Dmax assignment. Part of the overall problem is intrinsic to the nature of flexible proteins in solution. Heterogeneity in conformation cannot be currently identified or accounted for unless structures are already known or postulated. *Ab initio* shape predictions based on mixed populations give averaged conformations and can therefore be misleading (Heller, 2005). Thus, an area for development is the identification and assignment of multiple conformations that can be input into SAXS programs such as OLIGOMER (Konarev et al., 2003). The number of amino acids is also a required input for GASBOR and may be difficult to determine for large oligomeric complexes where the stoichiometry is not known. As with any single technique, conclusions based on SAXS data analysis need to be carefully considered and should be tested and validated with additional experiments.

Assembly of individually determined structures into a complex. Successful high resolution structure analysis often requires the divide-and-conquer approach, where individual domains or subunits but not the whole protein or complex can be crystallized or studied by NMR. SAXS data can provide a framework for piecing together individual high resolution structures. On a simple but still informative level, SAXS can be used to observe complex formation based on the Rg. For example, a Rg of 50.9 Å calculated by SAXS analysis of an Mre11/Rad50 ATPase complex indicated that the complex formed a heterotetramer with two Mre11 subunits and two Rad50 ATPase domains (Hopfner et al., 2001). Addition of ATP to the heterotetramer reduced the Rg to 46.3 Å, suggesting that ATP conformationally induced a tighter complex formation. However, a more direct and still powerful approach is to compare the scattering curves calculated from a rigid body refinement model to the experimental curve.

An example of the utility of comparing scattering curves comes from auto-inhibited Ca<sup>2+</sup>/calmodulin-dependent protein kinase II (CAMKII). The crystal structures of the individual domains were known, and there were EM structures of the whole protein (Hoelz et al., 2003; Kolodziej et al., 2000; Morris and Torok, 2001; Rosenberg et al., 2005). CAMKII is a dodecameric enzyme with two hexameric rings formed by a central association domain complex, and the kinase domains are fixed exterior to the central rings. However, the EM structures were in disagreement as to how the main association domains of the protein were juxtaposed to the kinase domain. Rosenberg and coworkers analyzed the full-length protein by SAXS and used semi-manual rigid body refinement of the kinase and association domain structures to determine that the inactive kinase domains lie flat in a ring exterior to the central assembly domains (Rosenberg et al., 2005). *Ab initio* shape predictions based on the experimental scattering curve agreed with this model. If instead the kinase domains lay above the plane of the association domains, the calculated scattering curve disagreed strongly with the experimental one. The SAXS scattering curve of an activated CAMKII was significantly different, with an apparent increase in Dmax of 100 Å. The authors speculated that upon activation, the kinase domains are released from their position to positions above and below the plane of the association domains. A second SAXS example is determining the oligomerization state in solution of the Holliday junction motor RuvB protein (Putnam et al., 2001). SAXS data provided an overall radius of gyration (Rg) for the particle of 40 Å with a Dmax of 120 Å, consistent with the crystal subunit structure and a generated hexamer assembly model. The scattering curve calculated from the hexamer model was similar to the experimental scattering curve, and an *ab initio* spherical



harmonic reconstruction fit the experimental curve with a small number of parameters and matched the hexamer model. These results showed how SAXS data can exclude packing arrangements that give different radii of gyration and maximum interparticle distances, such as more globular arrangements or asymmetric hexamers that are more elliptical and would have increased maximum intraparticle distances. The SAXS generated hexamer model appears likely to be correct as it showed good agreement with experimentally determined SAXS data, EM reconstructions, other hexameric AAA-class ATPases, and the crystallographic packing interfaces.

At SIBYLS, we have increasingly found that combining crystal structures of subunits with SAXS characterizations of the complex to be extremely powerful and efficient in the characterizations of reversible complexes. Such dynamic or reversible complexes act in many critical aspects of cell biology, and are typified by assemblies, such as DNA sliding clamps that encircle DNA and provide binding sites for many DNA replication and repair enzymes, including polymerase, flap endonuclease, and DNA ligase. Recently, we used SAXS in combination with crystallography to help elucidate the mechanisms by which sliding clamps like PCNA coordinate multi-step DNA transactions, which have been mysterious. Besides the new PCNA structure noted above (Fig. 1), we determined structures of *Sulfolobus solfataricus* DNA ligase and heterotrimeric PCNA separately by x-ray diffraction, and characterized their complex by SAXS experiments (Pascal et al., 2006). In the structure of this heterotrimeric DNA sliding clamp, the three distinct subunits assemble into a protein ring resembling the homotrimeric PCNA of humans, but with three unique protein-binding sites that facilitate the assembly of discrete multi-protein complexes. In the absence of a nicked DNA, the *Sulfolobus solfataricus* DNA ligase has an open, extended conformation by both crystallographic and SAXS studies. In the reversible complex amenable only to SAXS studies, the DNA ligase retains an open, extended conformation and binds to a single PCNA3 subunit (Fig. 2). Yet, a closed, ring-shaped conformation of ligase catalyzes a DNA end-joining reaction stimulated by PCNA. This open-to-closed switch in the conformation of DNA ligase is accommodated by a malleable interface with PCNA that serves as an efficient platform for DNA ligation. SAXS combined with crystallography thus established a flexible model for DNA ligase function that allows an extended ligase to bind the PCNA platform for DNA replication and repair and then to close into a ring around DNA for end joining. In essence SAXS was an essential component to show that ligase in solution can act like a molecular watch band that can extend or wrap around DNA as a reversible functional complex.

Observation of loops or domains whose structure has not been determined. In crystallographic studies, flexible loops or disordered regions are often either not visible in the electron density or have already been eliminated in the cloning stage to produce a sample that can be crystallized. SAXS can provide an effective complementary technique that can build low resolution models for these regions onto the known structure. Svergun and his coworkers have developed a suite of programs, including CREDO, based on their *ab initio* programs that can build dummy atoms or C $\alpha$  chains onto the crystal structure to show the relative location of these “missing” regions (Petoukhov et al., 2002). The known structure is fixed and then dummy residues are placed randomly within the input Dmax region. During the simulated annealing runs, the fit to the experimental scattering curve is optimized, and a model for the relative positioning of the unknown region is generated.

In their study of the *E. coli* primase, Corn and coworkers used this approach to identify the relative position of two domains which putatively connected by a long flexible linker (Corn et al., 2005). The structure of the polymerase domain was determined by solving the crystal structure, and the regulatory domain had strong homology to a known structure. SAXS data were collected on the full length protein. The polymerase domain was used as input to the program CREDO, which built density on the polymerase structure to best fit the scattering profile (Fig. 3). The homology model of the regulatory domain fit the additional density built by CREDO. The regulatory domain was positioned tightly against the polymerase domain, though unexpectedly far away from the proposed path of the primer/template (indicated by an arrow). By increasing the MgCl<sub>2</sub> concentration they observed increases in Dmax (Fig. 3). These studies further complemented fluorescence resonance energy transfer (FRET) experiments. Both results suggested that the regulatory domain is bound to

the polymerase domain through polar contacts and salt bridges which may be disrupted for function. Subsequent FRET studies by the authors indicated that the regulatory domain acts *in trans*, and they proposed that release of the two domains is an essential step toward regulation of the polymerase.

Ligand-induced conformational changes. Proteins often function through mechanistic domain motions that are controlled by binding small molecules such as nucleotide triphosphates, and SAXS provides a uniquely rapid and effective tool for observing gross domain movements. SAXS studies can be done with the same protein preparation under solution conditions that are identical to optimized biochemical assay conditions. Variables can include nucleotides, salt, buffer type, and pH. Comparison of the scattering curves can identify changes in flexibility and, with *ab initio* analysis, determination of domain shifts relative to each other. Time-resolved studies add another dimension to characterization of the conformational changes. Furthermore, automation of SAXS sample loading would open the way for medium throughput drug studies to identify compounds that induce a conformational change or changes in protein flexibility (Fischetti et al., 2004).

A recent study on an AAA ATPase, p97, illustrates the power of SAXS to observe conformational changes in large proteins (Davies et al., 2005). AAA ATPases are a remarkable family of proteins whose functions are controlled by binding and hydrolysis of ATP. Different nucleotide states confer large conformational changes that traverse the entire protein. p97 is a large protein; a single subunit is approximately 806 residues, and p97 forms a homohexamer. In the nucleotide-free form, the protein showed an expanded structure with an increase of Dmax of approximately 25 Å. Comparison of structures with bound AMP-PNP, ADP, or ADP-AIF<sub>x</sub> suggests that the specific nucleotide state of one set of ATPase domains alters the flexibility and relative position of two other domains and sets up a path for transfer of energy from the ATPase domain to interacting proteins through these large conformational changes. The authors note that there were significant differences from the EM studies, which they attribute to the higher resolution of the SAXS data. The data, collected at the Advanced Photon Source, extended to an s of 0.73 Å<sup>-1</sup> and provided the higher angle data critical for the detailed features observed in the SAXS shape predictions. An exciting development in the field is the use of molecular dynamics to introduce flexibility in crystallographic structures to match SAXS scattering curves (Karplus and Kuriyan, 2005; Wu et al., 2005); this technique could provide atomic resolution hypotheses on how proteins transition from one conformation to another. Although p97 is a relatively large protein for molecular dynamics, improvements in computing power as well as developments in the algorithms will increase the size possible for this type of study.

Possibility for visualization of flexible regions in proteins. Although flexible or unstructured domains are especially problematic for crystallographic studies, these regions can be uniquely characterized by SAXS. On one level, flexible and unstructured regions can be identified by the shape of the scattering curve on a Kratky plot (Doniach et al., 1995; Semisotnov et al., 1996). The Kratky plot graphs  $I(s) \cdot s^2$  vs s, where I is the intensity and s is the momentum transfer. Kratky plots for natively folded proteins are parabolic in appearance and  $I(s) \cdot s^2$  rises and falls in a relatively sharp peak with increasing s. For unfolded proteins, the  $I(s) \cdot s^2$  approaches a constant maximum value. For proteins with both structured and unstructured regions, the Kratky plot shows almost an additive nature with an initial parabolic nature and a tail at high s that resembles unfolded protein (Moncoq et al., 2004). As an aside for protein crystallographers, we have found that identification of flexible regions in constructs by SAXS analysis has aided in design and selection of crystallization targets.

NEIL1 is a protein where one quarter of the protein is predicted to be unstructured. It is a DNA glycosylase that removes oxidized pyrimidines; the catalytic domain is located in the N-terminus (Bandaru et al., 2002; Hazra et al., 2002). Sequence analysis by PONDR identified the C-terminal residues 282-390 as disordered (Bandaru et al., 2004). A deletion construct (2-342) with part of the C-terminus removed was successfully crystallized (Doublié et al., 2004). In the crystallographic model, only residues from the N-terminus to residue 290 were observable in the electron density, indicating that PONDR successfully predicted the disordered region to within 10 residues. We have

done SAXS analysis on full-length NEIL1 (Fig. 4). The Kratky plot is typical of a protein with both folded and unstructured regions. *Ab initio* shape prediction from the GASBOR program revealed an oblong structure. The crystal structure for NEIL1 fit well in the wider region, with the C-terminus facing the more narrow region. One loop which forms a crystallographic contact with another NEIL1 molecule in the crystal packing extends from the shape envelope. It is likely that in the full-length protein, this loop makes contact with the C-terminal region. This preliminary study indicates the ability of SAXS to visualize unstructured regions, although more work needs to be done to verify this conclusion.

## PERSPECTIVES

High resolution structure has transformed biology, raising our understanding of proteins from one-dimensional amino acid sequences to three-dimensional structures: Identification of the functional importance of a particular amino acid has been greatly facilitated by knowing its location within the protein and its chemical environment. Similarly, SAXS promises to raise this understanding to the next dimension since detailed characterizations of the protein conformational states, architectures, and reversible assemblies can provide important clues to how proteins carry out their functions. With advanced computing and the growing data base of solved structures, SAXS can be the optimal means to answer questions at many levels of resolution and to bridge the gaps between high resolution crystal structures and the large and dynamic complexes functioning in cells.

The SAXS field has been evolving rapidly in recent years and is increasingly becoming a critical technique for structural biology. First, SAXS provides useful information from almost every sample. Second, the minimal nature of the sample requirements compared to other structural techniques, with no requirement for labeling or for crystals and a large molecular weight range, is enabling SAXS analysis to become an increasingly powerful and important tool for probing structure in solution. Third, already many protein crystallography laboratories, which use the SAXS endstation at the SIBYLS beamline, consider SAXS analysis a standard characterization of their samples to characterize states of folding, aggregation, assembly, and conformation. Access to SAXS endstations at synchrotrons is available through general user proposals, and beamline staff can aid collaborative projects. As more scientists take advantage of SAXS analysis, we will gain a greater fundamental understanding of how proteins function and are controlled through dynamic features including conformational flexibility, reversible assemblies, and dynamic complexes. The emerging grand challenge of characterizing molecular machines that control cell biology, including responses to endogenous and exogenous stress, will require the type of powerful and efficient analysis in solution that SAXS can provide.

## ACKNOWLEDGEMENTS.

We thank John Pascal, Tom Ellenberger, Sankar Mitra, Tapas Hazra, Jacob Corn, and James Berger for collaborating on the structural studies and for helpful discussions. We thank James Holton, Jane Tanamachi, Christine Trame, and Hiro Tsuruta for their help during design and commissioning of the beamline. We thank Jill Fuss and Quen Cheng for constructive discussions and comments on the manuscript and Vladimir Uversky for helpful discussions on unstructured proteins. The SIBLYS beamline at the Advanced Light Source and these studies on the integration of crystallographic and SAXS analyses and for the characterization of reversible complexes and modified proteins were supported by the National Institutes of Health (NIH) Structural Cell Biology of DNA Repair Machines P01 grant CA92584 and by the Office of Science, Office of Biological and Environmental Research, U.S. Department of Energy, under Contract Number DE-AC02-05CH11231.

## FIGURE LEGENDS

**Fig. 1.** *Ab initio* shape prediction by SAXS analysis

SAXS analysis of OGG1, glucose isomerase, and PCNA collected at the SIBYLS beamline, ALS. (A) Buffer-subtracted scattering profiles and (B) electron pair distribution plots of OGG1 (blue), glucose isomerase (purple), and PCNA (green). Human OGG1 (5 mg/ml) was analyzed in 25 mM Tris pH 7, 150 mM NaCl, 1mM TCEP, 10% glycerol, 0.5 mM EDTA. Work was done in collaboration with Tapas Hazra and Sankar Mitra (University of Texas, Medical Branch, Galveston). *Streptomyces rubiginosus* glucose isomerase (Hampton Research) at 2 mg/ml was analyzed in 20 mM Tris pH 7.5, 1mM TCEP. Human PCNA (4 mg/ml) was analyzed in 25 mM Tris pH 7.6, 150 mM NaCl, 0.1 mM EDTA and 5 mM DTT. Work was done in collaboration with John Pascal and Tom Ellenberger (Harvard University Medical School). (C) Averaged and filtered *ab initio* shape prediction calculated by GASBOR and aligned using the DAMAVER program suite (Volkov and Svergun, 2003). The DAMAVER program suite was used to overlay the shape predictions onto the crystals structures (ribbon) of, respectively, OGG1 (1KO9.pdb), glucose isomerase (8XIA.pdb), and PCNA (1AXC,) (Bjoras et al., 2002; Carrell et al., 1989; Gulbis et al., 1996)

**Fig. 2.** *Ab initio* shape prediction of complex between DNA ligase and PCNA.

SAXS analysis of *Sulfolobus solfataricus* ssLig, PCNA1, PCNA2, and PCNA3. The protein concentration was 4 and 8 mg/ml, and the sample buffer was 25 mM Tris pH 7.5, 50 mM NaCl, and 1 mM TCEP. Averaged and filtered *ab initio* shape prediction was calculated by GASBOR and aligned using the DAMAVER program suite (Volkov and Svergun, 2003).

**Fig. 3.** Use of SAXS to identify relative positioning of two domains in *E. coli* primase, with the structure of only one domain used as input. Work was published in collaboration with Jacob Corn and James Berger (Corn et al., 2005). Electron pair distribution function for *E. coli* primase under varying salt conditions. Low salt buffer (20 mM HEPES, pH 7.5, 75 mM potassium glutamate) is colored in black, buffer supplemented with 150 mM MgCl<sub>2</sub> in red and buffer supplemented with 300 mM MgCl<sub>2</sub>. Consensus shape prediction (transparent spheres) for the regulatory domain of *E. coli* primase overlaid with the homology model (ribbon). The envelope of the polymerase domain is shown in solid spheres. The arrow indicates the postulated path of the primer/template. Shape prediction was calculated by CREDO and averaged using the DAMAVER program suite.

**Fig. 4.** Use of SAXS to predict conformation of an unstructured region in NEIL1 glycosylase. (A) Buffer-subtracted scattering profile of NEIL1 (5 mg/ml) dialyzed in PBS and 1 mM TCEP. Inset: Kratky plot of NEIL1. The non-parabolic nature of the high *s* region indicates that there is unstructured regions present in the protein (B) Averaged and filtered *ab initio* shape prediction calculated by GASBOR and aligned using the DAMAVER program suite. The ribbon structure of NEIL1 N-terminus (1TDH.pdb) was manually fit into the envelope (Doublié et al., 2004). The arrow indicates the C-terminal residue of the crystal structure. Work was done in collaboration with Tapas Hazra and Sankar Mitra.

## BIBLIOGRAPHY

- Bandaru, V., S. Sunkara, S.S. Wallace, and J.P. Bond, 2002. A novel human DNA glycosylase that removes oxidative DNA damage and is homologous to Escherichia coli endonuclease VIII. *DNA Repair* 1: 517-529.
- Bandaru, V., W. Cooper, S.S. Wallace, and S. Doublié, 2004. Overproduction, crystallization and preliminary crystallographic analysis of a novel human DNA-repair enzyme that damage recognizes oxidative DNA damage. *Acta Crystallographica Section D-Biological Crystallography* 60: 1142-1144.
- Bjoras, M., E. Seeberg, L. Luna, L.H. Pearl, and T.E. Barrett, 2002. Reciprocal "flipping" underlies substrate recognition and catalytic activation by the human 8-oxo-guanine DNA glycosylase. *J Mol Biol* 317: 171-7.
- Carrell, H.L., J.P. Glusker, V. Burger, F. Manfre, D. Tritsch, and J.F. Biellmann, 1989. X-ray analysis of D-xylose isomerase at 1.9 Å: native enzyme in complex with substrate and with a mechanism-designed inactivator. *Proc Natl Acad Sci U S A* 86: 4440-4.
- Chacon, P., F. Moran, J.F. Diaz, E. Pantos, and J.M. Andreu, 1998. Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophysical Journal* 74: 2760-2775.
- Corn, J.E., P.J. Pease, G.L. Hura, and J.M. Berger, 2005. Crosstalk between primase subunits can act to regulate primer synthesis in trans. *Mol Cell* 20: 391-401.
- Davies, J.M., H. Tsuruta, A.P. May, and W.I. Weis, 2005. Conformational changes of p97 during nucleotide hydrolysis determined by small-angle X-ray scattering. *Structure* 13: 183-195.
- Doniach, S., J. Bascle, T. Garel, and H. Orland, 1995. Partially Folded States of Proteins - Characterization by X-Ray-Scattering. *Journal of Molecular Biology* 254: 960-967.
- Doublié, S., V. Bandaru, J.P. Bond, and S.S. Wallace, 2004. The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proc Natl Acad Sci U S A* 101: 10284-10289.
- Dunker, A.K., Z. Obradovic, P. Romero, E.C. Garner, and C.J. Brown, 2000. Intrinsic Protein Disorder in Complete Genomes. *Genome Informatics* 11: 161-171.
- Dunker, A.K., M.S. Cortese, P. Romero, L.M. Iakoucheva, and V.N. Uversky, 2005. Flexible nets - The roles of intrinsic disorder in protein interaction networks. *Febs Journal* 272: 5129-5148.
- Dyson, H.J., and P.E. Wright, 2002. Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology* 12: 54-60.
- Feigin, L.A., and D.I. Svergun, 1987. *Structure Analysis by Small-Angle X-ray and Neutron Scattering* Plenum Press, New York.
- Fischetti, R.F., D.J. Rodi, D.B. Gore, and L. Makowski, 2004. Wide-angle X-ray solution scattering as a probe of ligand-induced conformational changes in proteins. *Chem Biol* 11: 1431-43.
- Gulbis, J.M., Z. Kelman, J. Hurwitz, M. O'Donnell, and J. Kuriyan, 1996. Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell* 87: 297-306.
- Hazra, T.K., T. Izumi, I. Boldogh, B. Imhoff, Y.W. Kow, P. Jaruga, M. Dizdaroglu, and S. Mitra, 2002. Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc Natl Acad Sci U S A* 99: 3523-3528.
- Heller, W.T., 2005. Influence of multiple well defined conformations on small-angle scattering of proteins in solution. *Acta Crystallographica Section D-Biological Crystallography* 61: 33-44.
- Hoelz, A., A.C. Nairn, and J. Kuriyan, 2003. Crystal structure of a tetradecameric assembly of the association domain of Ca<sup>2+</sup>/calmodulin-dependent kinase II. *Mol Cell* 11: 1241-51.
- Hopfner, K.P., A. Karcher, L. Craig, T.T. Woo, J.P. Carney, and J.A. Tainer, 2001. Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* 105: 473-85.
- Karplus, M., and J. Kuriyan, 2005. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* 102: 6679-85.

- Kolodziej, S.J., A. Hudmon, M.N. Waxham, and J.K. Stoops, 2000. Three-dimensional reconstructions of calcium/calmodulin-dependent (CaM) kinase II $\alpha$  and truncated CaM kinase II $\alpha$  reveal a unique organization for its structural core and functional domains. *J Biol Chem* 275: 14354-9.
- Konarev, P.V., V.V. Volkov, A.V. Sokolova, M.H.J. Koch, and D.I. Svergun, 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* 36: 1277-1282.
- Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissole, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, *et al.*, 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- Moncoq, K., I. Broutin, C.T. Craescu, P. Vachette, A. Ducruix, and D. Durand, 2004. SAXS study of the PIR domain from the Grb14 molecular adaptor: A natively unfolded protein with a transient structure primer? *Biophysical Journal* 87: 4056-4064.
- Morris, E.P., and K. Torok, 2001. Oligomeric structure of  $\alpha$ -calmodulin-dependent protein kinase II. *J Mol Biol* 308: 1-8.
- Pascal, J.M., O.V. Tsodikov, G.L. Hura, W. Song, E.A. Cotner, S. Classen, A.E. Tomkinson, J.A. Tainer, and T. Ellenberger, 2006. A flexible interface between DNA ligase and a heterotrimeric sliding clamp supports conformational switching and efficient ligation of DNA. *Mol Cell* in press.
- Petoukhov, M.V., and D.I. Svergun, 2003. New methods for domain structure determination of proteins from solution scattering data. *Journal of Applied Crystallography* 36: 540-544.
- Petoukhov, M.V., and D.I. Svergun, 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical Journal* 89: 1237-1250.
- Petoukhov, M.V., N.A. Eady, K.A. Brown, and D.I. Svergun, 2002. Addition of missing loops and domains to protein models by x-ray solution scattering. *Biophys J* 83: 3113-25.
- Putnam, C.D., S.B. Clancy, H. Tsuruta, S. Gonzalez, J.G. Wetmur, and J.A. Tainer, 2001. Structure and mechanism of the RuvB Holliday junction branch migration motor. *Journal of Molecular Biology* 311: 297-310.
- Rosenberg, O.S., S. Deindl, R.J. Sung, A.C. Nairn, and J. Kuriyan, 2005. Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. *Cell* 123: 849-60.
- Semisotnov, G.V., H. Kihara, N.V. Kotova, K. Kimura, Y. Amemiya, K. Wakabayashi, I.N. Serdyuk, A.A. Timchenko, K. Chiba, K. Nikaido, T. Ikura, and K. Kuwajima, 1996. Protein globularization during folding. A study by synchrotron small-angle X-ray scattering. *Journal of Molecular Biology* 262: 559-574.
- Sokolova, A.V., V.V. Volkov, and D.I. Svergun, 2003a. Prototype of a database for rapid protein classification based on solution scattering data. *Journal of Applied Crystallography* 36: 865-868.
- Sokolova, A.V., V.V. Volkov, and D.I. Svergun, 2003b. Database for rapid protein classification based on small-angle X-ray scattering data. *Crystallography Reports* 48: 959-965.

- Svergun, D.I., 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal* 76: 2879-2886.
- Svergun, D.I., and M.H.J. Koch, 2003. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics* 66: 1735-1782.
- Svergun, D.I., M.V. Petoukhov, and M.H.J. Koch, 2001. Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal* 80: 2946-2953.
- Takahashi, Y., Y. Nishikawa, and T. Fujisawa, 2003. Evaluation of three algorithms for ab initio determination of three-dimensional shape from one-dimensional solution scattering profiles. *Journal of Applied Crystallography* 36: 549-552.
- Trame, C., A.A. MacDowell, R.S. Celestre, H.A. Padmore, D. Cambie, E.E. Domning, R.M. Duarte, N. Kelez, D.W. Plate, J.M. Holton, K. Frankel, S. Tsutakawa, H. Tsuruta, J.A. Tainer, and P.K. Cooper. 2003. SIBYLS - A SAXS and protein crystallography beamline at the ALS, pp. 502-505, *In* T. Warwick, (ed.) *Synchrotron Radiation Instrumentation*, Eighth International Conference. American Institute of Physics, San Francisco.
- Venter, J.C., 2001. The sequence of the human genome (vol 292, pg 1304, 2001). *Science* 292: 1838-1838.
- Volkov, V.V., and D.I. Svergun, 2003. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* 36: 860-864.
- Walther, D., F.E. Cohen, and S. Doniach, 2000. Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules. *Journal of Applied Crystallography* 33: 350-363.
- Wu, Y.H., X. Tian, M.Y. Lu, M.Z. Chen, Q.H. Wang, and J.P. Ma, 2005. Folding of small helical proteins assisted by small-angle X-ray scattering profiles. *Structure* 13: 1587-1597.
- Zipper, P., and H. Durchschlag, 2003. Modeling of protein solution structures. *Journal of Applied Crystallography* 36: 509-514.

Figure 1

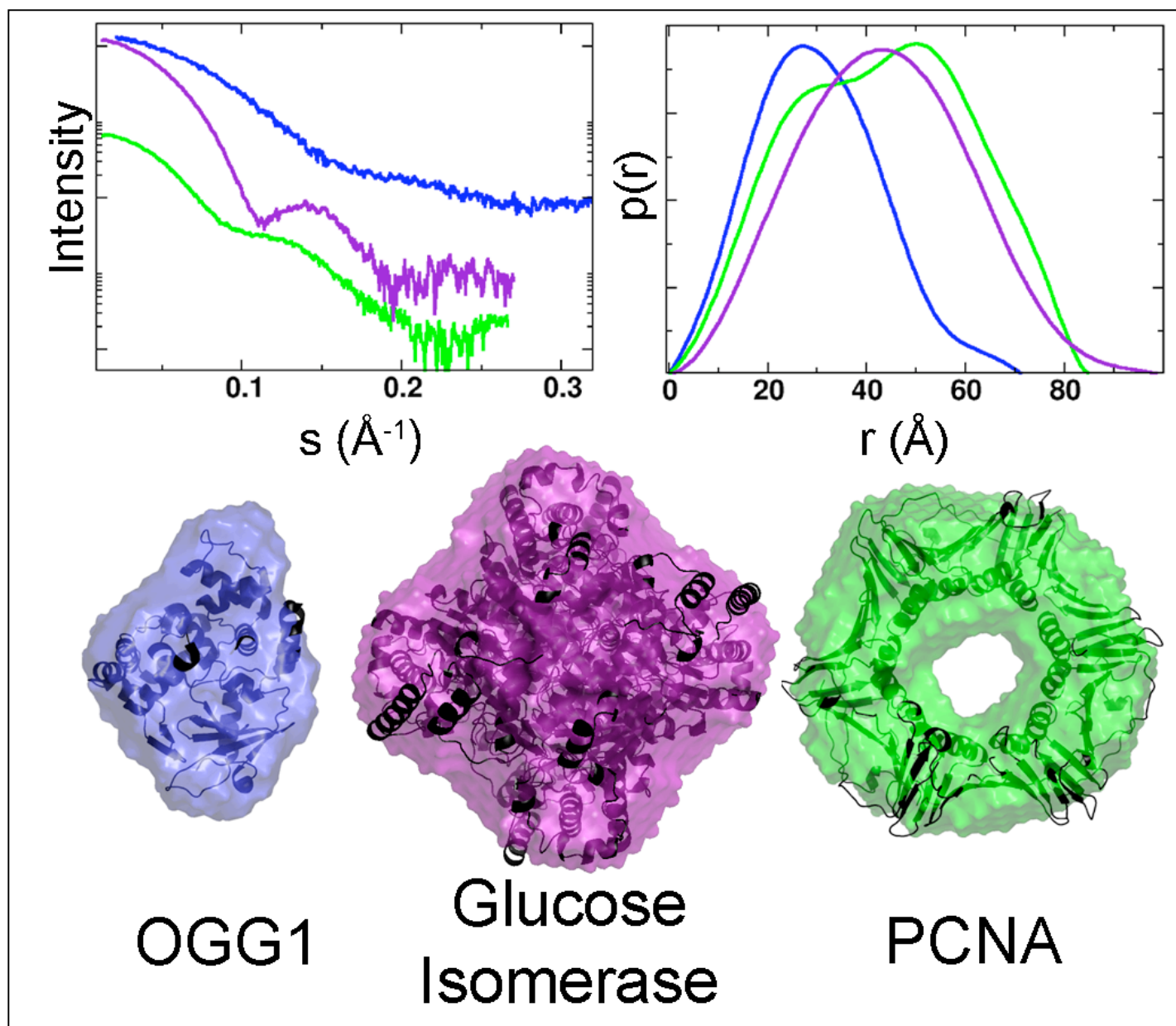




Figure 2

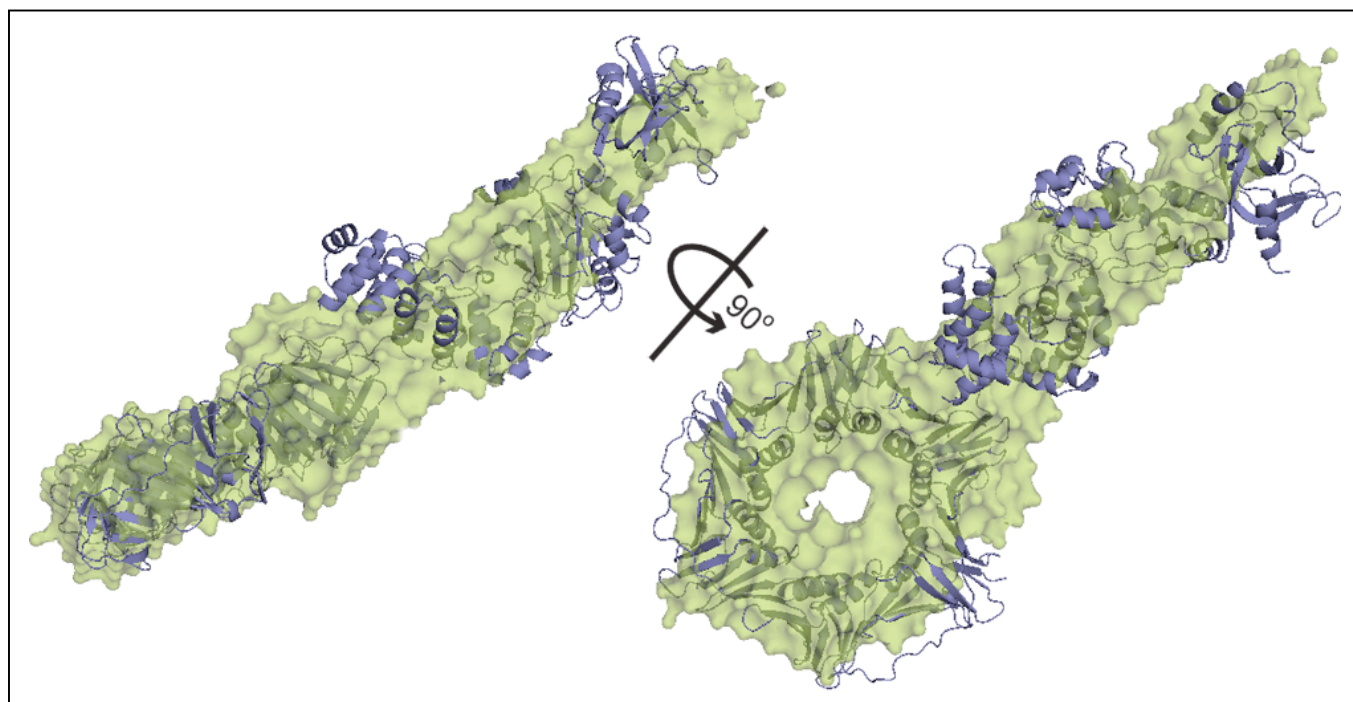


Figure 3

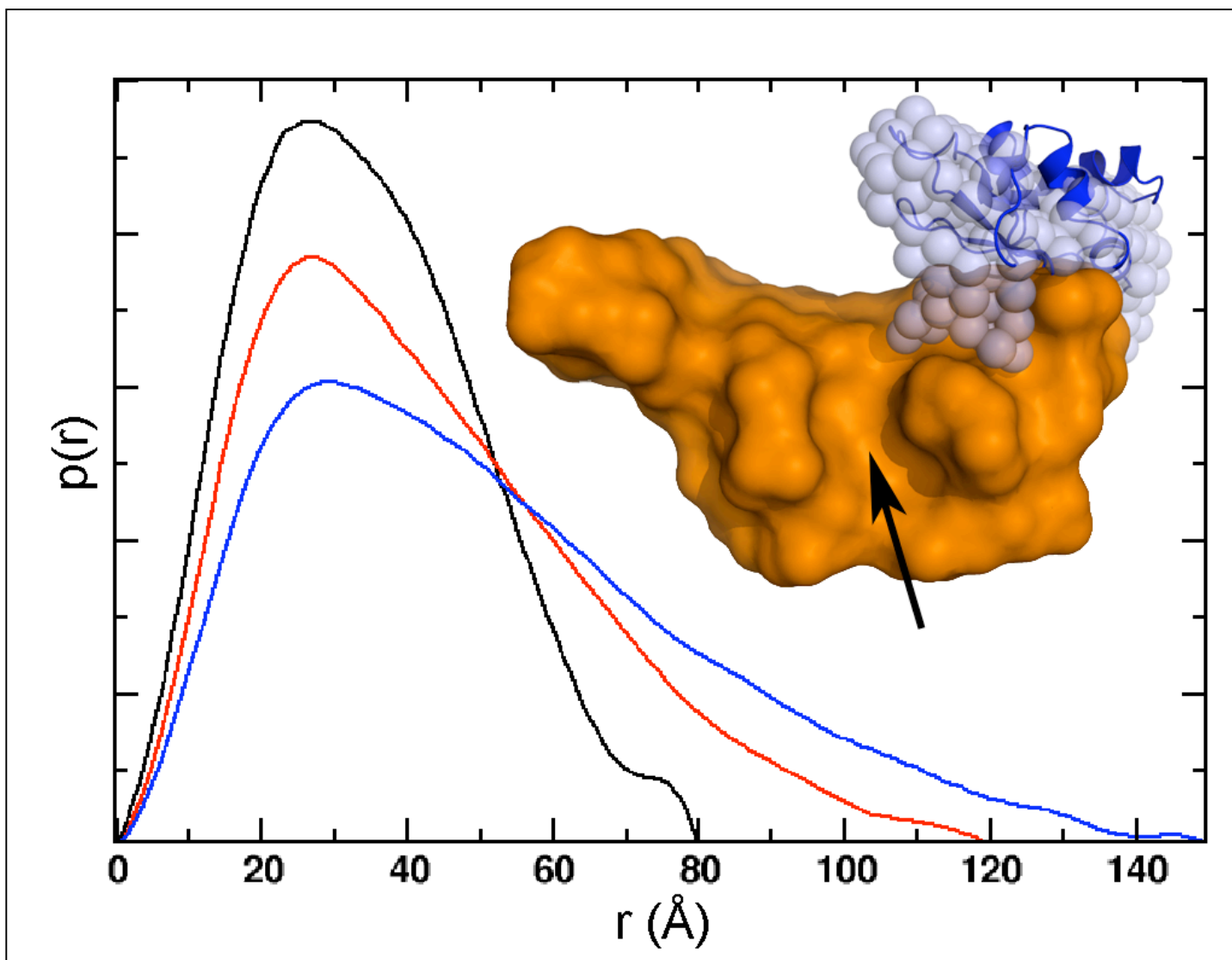


Figure 4

