
High Resolution Distance Distributions Determined by X-Ray and Neutron Scattering

10

Henry Y.H. Tang, John A. Tainer, and Greg L. Hura

Abstract

Measuring distances within or between macromolecules is necessary to understand the chemistry that biological systems uniquely enable. In performing their chemistry, biological macromolecules undergo structural changes over distances ranging from atomic to micrometer scales. X-ray and neutron scattering provide three key assets for tackling this challenge. First, they may be conducted on solutions where the macromolecules are free to sample the conformations that enable their chemistry. Second, there are few limitations on chemical environment for experiments. Third, the techniques can inform upon a wide range of distances at once. Thus scattering, particularly recorded at small angles (SAS), has been applied to a large variety of phenomenon. A challenge in interpreting scattering data is that the desired three dimensional distance information is averaged onto one dimension. Furthermore, the scales and variety of phenomenon interrogated have led to an assortment of functions that describe distances and changes thereof. Here we review scattering studies that characterize biological phenomenon at distances ranging from atomic to 50 nm. We also distinguish the distance distribution functions that are commonly used to describe results from these systems. With available X-ray and neutron scattering facilities, bringing the action that occurs at

H.Y.H. Tang
Molecular Biophysics and Integrated Bioimaging,
Lawrence Berkeley National Laboratory, 1 Cyclotron Rd,
Berkeley, CA 94720, USA

J.A. Tainer
Molecular Biophysics and Integrated Bioimaging,
Lawrence Berkeley National Laboratory, 1 Cyclotron Rd,
Berkeley, CA 94720, USA

Department of Molecular and Cellular Oncology, The
University of Texas M. D. Anderson Cancer Center,
Houston, TX, USA

G.L. Hura (✉)
Molecular Biophysics and Integrated Bioimaging,
Lawrence Berkeley National Laboratory, 1 Cyclotron Rd,
Berkeley, CA 94720, USA

Department of Chemistry and Biochemistry, University
of California, Santa Cruz, CA, USA
e-mail: glhura@lbl.gov

the atomic to the micrometer scale is now reasonably accessible. Notably, the combined distance and dynamic information recorded by SAS is frequently key to connecting structure to biological activity and to improve macromolecular design strategies and outcomes. We anticipate widespread utilization particularly in macromolecular engineering and time-resolved studies where many contrasting experiments are necessary for resolving chemical mechanisms through structural changes.

Keywords

SAXS • Pair distribution • Pair correlation • Protein structure • Resolution • Molecular ruler

10.1 Introduction

X-ray and neutron scattering from solutions provide information at atomic to intra-organelle distances for which microscope based visible light techniques cannot be applied. In addition scattering can be applied at any biologically relevant solution condition excellently complimenting techniques that provide greater detail but only under restrictive conditions. Significant advances in the collection and analysis of high quality data have been made over the last ten years leading to several important results. Recent examples include studies on polyketide synthase (Edwards et al. 2014), elastin (Baldock et al. 2011), photosynthesis (Stingaciu et al. 2016), chromatin (Andresen et al. 2013; Falk et al. 2015) and microbial chromatin-like systems (Hammel et al. 2016). Recent advances in genomic sequencing and purification have led to a large increase in the number of targets that would greatly benefit from increased structural understanding. Demand for structural information is outpacing the capabilities of current main stream experimental techniques. Scattering reliably provides results on most samples under nearly any solution condition and is capable of high-throughput and multi-condition analyses (Hura et al. 2009, 2013a), so the impact of these techniques can potentially be very profound.

A primary asset of X-ray and neutron scattering is the ability to elucidate relatively detailed or high resolution structural features. This fact is in contradiction to what is often stated about SAXS

and SANS (collectively SAS), as a web search will quickly attest. The perception that the techniques are low resolution largely stems from the use of SAS to estimate 3D shape from macromolecules. 3D shape is a point of convergence between structural methods of macromolecular crystallography (MX), electron microscopy (EM) and nuclear magnetic resonance (NMR) techniques, so it is quite natural to compare results at this level. The resolution of SAS shapes is lower relative to other structural techniques but not because high resolution features do not contribute. SAS results do not preserve directional information and the resulting ambiguity reduces the precision of shape determination. The loss of directional information is the price paid for the ability to work effectively and efficiently in solution. Despite the lack of directional information, shapes from SAS experiments are calculated by supplementing data with assumptions about compactness and density connectivity that have proven to be valid and are valuable because of key advantages of working in solution (Franke and Svergun 2009). However, this may not be the most important asset of SAS.

A better illustration of the resolution capabilities of scattering data comes from examining differences between scattering profiles collected from a macromolecule undergoing a small structural change. To demonstrate the accuracy and multiple scales at which scattering techniques can measure, we will start by developing the formalism that is applied for most

macromolecules through use of the pair distribution function or $P(r)$. Comparisons of the $P(r)$, attained through Fourier transformation of the primary data, provides insight into the resolution possible. Many examples of the value of this function are described in the literature on particles that are 1–10 nm in maximum dimension. Below we highlight the precision possible but rather than dwell on this length scale, for which there are several excellent tools (Semenyuk and Svergun 1991; Nielsen et al. 2009; Bergmann et al. 2000), we will explore either side of this scale. We will review scattering as a tool to measure the distances between water molecules in bulk water at atomic length scales (0.1 nm). We will also describe the use of metallic labels to characterize small changes over large distances on the order of 50 nm.

We aim to highlight how these broad length scales (0.1–50 nm) are bridged by scattering. We believe such a treatise is important since instruments with sufficient flux and large detectors are increasingly available, providing access to these regimes in a single experiment on time scales of 1 s or less. We endeavor to enable investigators to begin extending their X-ray or neutron scattering analysis to understand multiple important phenomena. In addition, we seek to clarify the differences between distance distributions extracted from each approach which can be confusing because they describe related quantities. The measurement of structural changes a macromolecule undergoes as part of its mechanism has wide spread

application and scattering provides this capability with sub-nanometer resolution.

10.2 General Formalism of Extracting Distances from Scattering and Diffraction

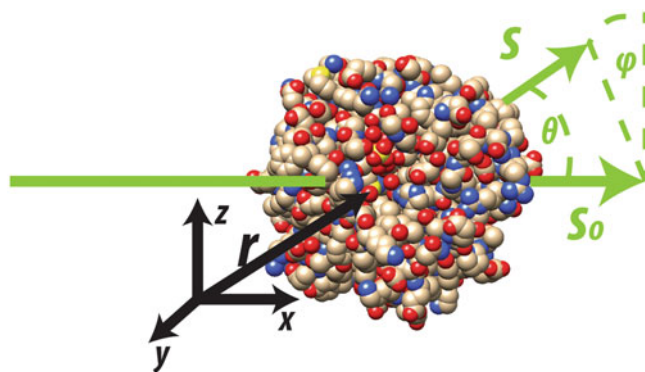
Reconstructing a macromolecular structure from its coherent scattering starts with two fundamental equations. Equations 10.1 and 10.2 describe the relationship between measured scattering intensities $I(\mathbf{q})$ and the structure of the scattering material $\rho(\mathbf{r})$. Bold face variables indicate vectoral quantities. The quantity $\rho(\mathbf{r})$ will either describe the electron density for X-rays or nuclear scattering length density for neutrons as a function of position from some origin r (Fig. 10.1).

$$f(\mathbf{q}) = \int_V \rho(\mathbf{r}) e^{i(\mathbf{r} \cdot \mathbf{q})} d\mathbf{r} \quad (10.1)$$

$$I(\mathbf{q}) = \sum_j \sum_k f_j(\mathbf{q}) f_k^*(\mathbf{q}) e^{i(\mathbf{r}_j - \mathbf{r}_k) \cdot \mathbf{q}} \quad (10.2)$$

In Eq. 10.1, $\mathbf{q} = 2\pi(\mathbf{s}_0 - \mathbf{s})/\lambda$ is the vectoral change in momentum in momentum between a photon or neutron prior to its interaction with material and after its interaction with the material. The quantity \mathbf{q} is therefore often referred to as momentum transfer. We are interested in the coherent scattering where momentum changes are solely directional,

Fig. 10.1 Coordinates for describing the scattering from an object. The coordinate system used to describe the scattering density is in *black*. The incident beam and outgoing scattering are described by the *green* vectors and coordinates



with the change in direction, $\mathbf{s}_o - \mathbf{s}$, between the incident and the outgoing angle. The vectors \mathbf{s}_o and \mathbf{s} are unit vectors describing the direction of the incident and outgoing scattering. The $f(\mathbf{q})$ function is often referred to as the “form” function and involves an integral over the volume V of the scattering particle (the portion of the sample illuminated by the incident X-ray beam).

In two common biological approaches that have been powerfully used together (Putnam et al. 2007), macromolecular crystallography and solution bioSAS from monodisperse samples, the scattering material is organized in two extreme ways. These are extreme in that with crystallography all particles are, ideally, perfectly ordered with respect to one another while in solution SAS, ideally, none of the particles are ordered in relation to one another. The organization of the material in these two different ways also distinguishes diffraction from scattering.

In most biological cases the total scattering from a macromolecule is not much stronger than the scattering of water regardless of whether neutrons or X-rays are used. Water of hydration is therefore an important contribution. However, for simplicity we neglect hydration now and will deal with some aspects later. For more complete detailed descriptions, we urge readers to consult several excellent resources (Schneidman-Duhovny et al. 2016; Koch et al. 2003; Poitevin et al. 2011).

In crystallographic diffraction, where all molecules are oriented identically, the vectoral distance of one atom relative to another is retained. In SAS, upon averaging all orientations of the macromolecules relative to the incident beam, the vectoral information is reduced to scalars. This split between the two techniques can be seen from the $I(\mathbf{q})$ recorded from either type of sample. For crystallographic diffraction, \mathbf{q} must be measured in two angular dimensions since the intensity of diffraction spots containing essential information vary in two dimensions.

In diffraction, a copy of an object located at \mathbf{r}_j can be found at \mathbf{r}_k oriented identically relative to the probing beam. The relation to the diffraction intensity can be made more useful when it takes

into account the repeating regular translation of one scattering unit relative to another by describing $\mathbf{r}_{jk} = \mathbf{r}_j - \mathbf{r}_k$ in terms of the three crystal lattice dimensions. By replacing the general \mathbf{r}_{jk} with lattice units, the equations describing crystallographic diffraction simplify. As we proceed and consider scattering from solutions, the same conditions do not hold and a different formalism is applied.

For bioSAS from dilute solutions of biomolecules, the scattering objects of interest are far enough away from one another that they may be treated as uncorrelated. There is no structure between individual particles. In this case the only term to consider in Eq. 10.2 is the self-scattering, shown in Eq. 10.3, where $(\mathbf{r}_j - \mathbf{r}_k) = 0$. With solution scattering, the components are in every orientation relative to the probing beam. Thus an extra mathematical operation is required to describe SAS data. The averaging over all orientations is indicated by the triangular brackets and is accomplished by integration over the traditional spherical coordinate system with axes r , θ and φ . This averaging results in loss of vectoral information but retention of scalar distance information. Therefore, only a one dimensional convolution of $\rho(\mathbf{r})$ can be determined, the pair distribution function or $P(r)$. The relationship between the intensity and $P(r)$ is shown in Eq. 10.3.

$$\begin{aligned} I(q) &= \langle f(\mathbf{q})f^*(\mathbf{q}) \rangle \\ &= \int_0^{D_{max}} 4\pi r^2 P(r) \frac{\sin qr}{qr} dr \quad (10.3) \end{aligned}$$

Since

$$\langle e^{i\mathbf{r}\cdot\mathbf{q}} \rangle = \frac{\int_0^\pi \sin \varphi d\varphi \int_0^{2\pi} e^{iqr\cos\theta} d\theta}{\int_0^\pi \sin \varphi d\varphi \int_0^{2\pi} d\theta} = \frac{\sin(qr)}{qr}$$

The momentum transfer $|\mathbf{q}| = q = (4\pi \sin(\theta/2))/\lambda$, now varies in one angular dimension, θ . Thus all information could be collected by a one dimensional strip detector. This is typically not done because by collecting on an area detector and averaging, the signal to noise is of greater quality.

Despite this mapping of three dimensional information on a one dimensional space, high resolution capability stems from the use of wavelengths (λ) on the order of 1 Å. When scattering is done with X-rays and neutrons, it is intrinsically capable of measuring scalar distances at the same resolution as crystallography because the wavelengths used are the same.

An often unappreciated advantage in solution scattering is that the scattering objects are not in contact as they are in crystallography. This lack of contact limits the integration of the distance r up to the maximum dimension of the particle D_{\max} . In crystallography the boundary between scattering units can be difficult to decipher making the analogue to the $P(r)$ function, called the Patterson function, less directly interpretable. Conversely the $P(r)$ function is a histogram of pair distances of scattering density. Its properties are pictorially illustrated in Fig. 10.2 on a toy system. Note that changes in the position of one atom influences significant portions of the distribution.

For macromolecules in real systems, the $P(r)$ can be very sensitive to conformational changes or modifications. Figure 10.3 illustrates the accuracy achievable on a protein where no

atom changes more than 5 Å. This level of accuracy is achievable because (1) the protein is either in one state or another and (2) a substantial proportion of the molecule moves. Had the protein sampled a range of conformations in its apo state relative to a single conformation in the ligand-bound state, the $P(r)$ functions for the apo and ligand-bound state would be more difficult to discern. Furthermore, if only a small portion of the molecule is changing, the $P(r)$ function would be less sensitive to the change. For example, while the addition of a single amino acid to a terminus of lysozyme may be detectable in the $P(r)$ function, the same cannot be said for one of the ribosomal proteins within a ribosome. While several factors influence the resolution of changes that can be determined, the $P(r)$ function has successfully and uniquely detected subtle changes.

Until recently, existing methods for extracting the $P(r)$ function from SAS data have been suitable. However, as data at higher angles are now routinely collected, particularly with X-rays, available tools are showing limitations. New methods are sure to arise for even greater definition and sensitivity of macromolecular structure. Of particular importance, as higher angles are

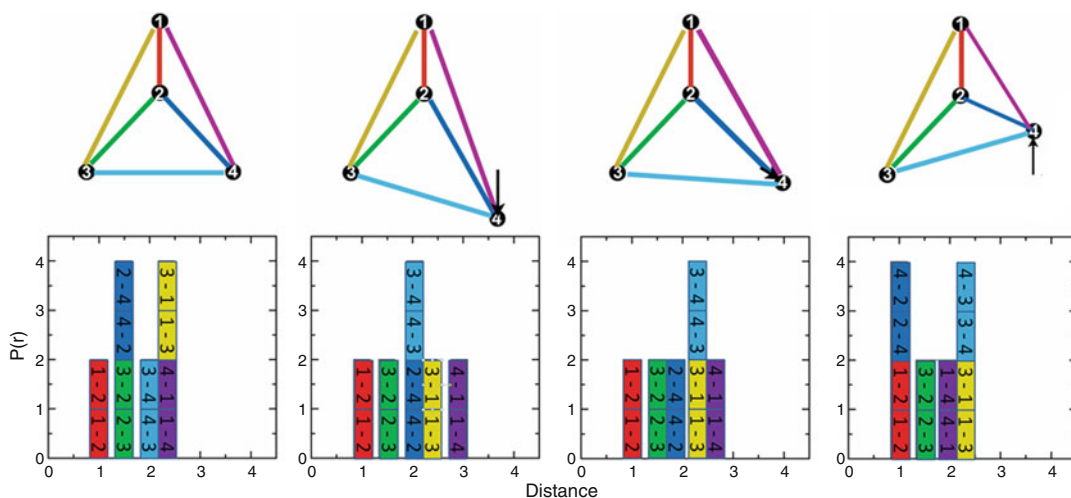


Fig. 10.2 The pair distribution function, $P(r)$, is a histogram of distances. A toy model of a scattering is composed of four labeled components. The $P(r)$ function is

shown below. As the spatial distribution changes within the system the $P(r)$ function can be dramatically sensitive to these changes

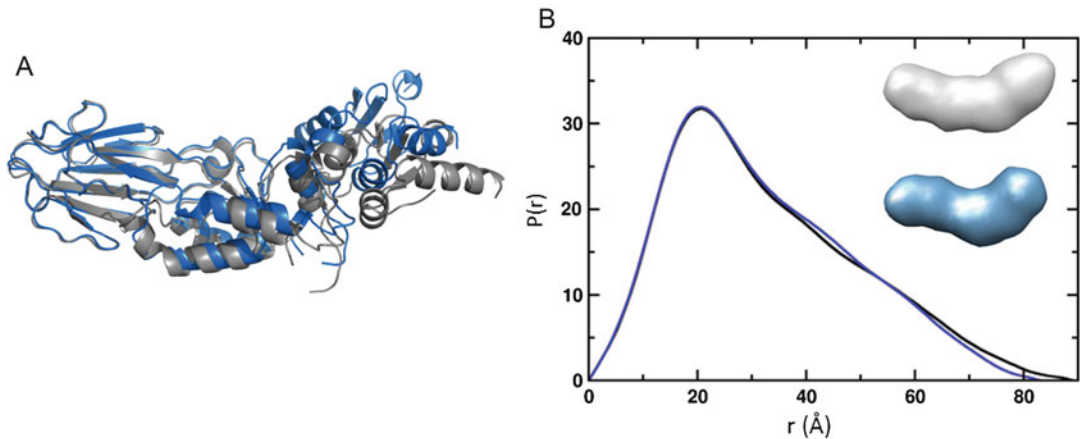


Fig. 10.3 The $P(r)$ function provides high resolution information. (a) The protein NBS1 undergoes a conformational change from an extended (*black*) in an apo state or contracted (*blue*) when binding a small peptide (Williams et al. 2009). In this conformational change no atom moves more than 5 Å. (b) The difference can be detected in the

$P(r)$ function. The decrease in maximum dimension from 87 to 83 Å among other changes are visible. However these differences would be difficult to discern from SAS generated shapes on the two samples (inset)

measured, the influence of water which we consider below.

10.3 Measuring the Distance Between Water Molecules in the Liquid Phase

Liquids have structure as readily indicated by strong and non-monotonic features from their scattering. Nearly all scattering from biological samples will contain a significant signal from the structure of liquid water, as shown in Fig. 10.4a, b. Learning about the liquid water signal is of value if for no other reason than to de-convolute it from the signal of interest. Further investigation may be warranted though, as the structural interactions of water underlying this signal are also of important consequence for many processes in biology.

The molecules in liquid water are constantly translating and rotating due to thermal fluctuations. Thus the distance between any single water molecule and its instantaneous neighbor will vary with time. However, since all water molecules are identical, the influence of thermal

energy is actually out-sized; in exchange for breaking bonds holding two waters together, another near equivalent bond with a different water is made. A water molecule will have a steady number of neighbors, shown in Fig. 10.4c. Water is famously polar giving a preferred directionality to interactions with neighbors. Consequently, certain spaces around a given water molecule are preferred relative to others. The strength of polarity competes with van der Waals interactions producing an optimal distance between water molecules. The influence of a water molecule goes beyond those in direct contact and extends to correlated second and third shells until thermal fluctuations dominate. These structural characteristics have a profound impact on the chemistry and biology of water. Furthermore, since water is a small molecule with a total of ten electrons it can be simulated in detail to provide insights into the quantum mechanics of inter-molecular bonds. Scattering measurements on water have played a fundamental role in these areas.

Interpretation of scattering data from molecular liquids focuses on the relationship between molecules rather than on the structure of the

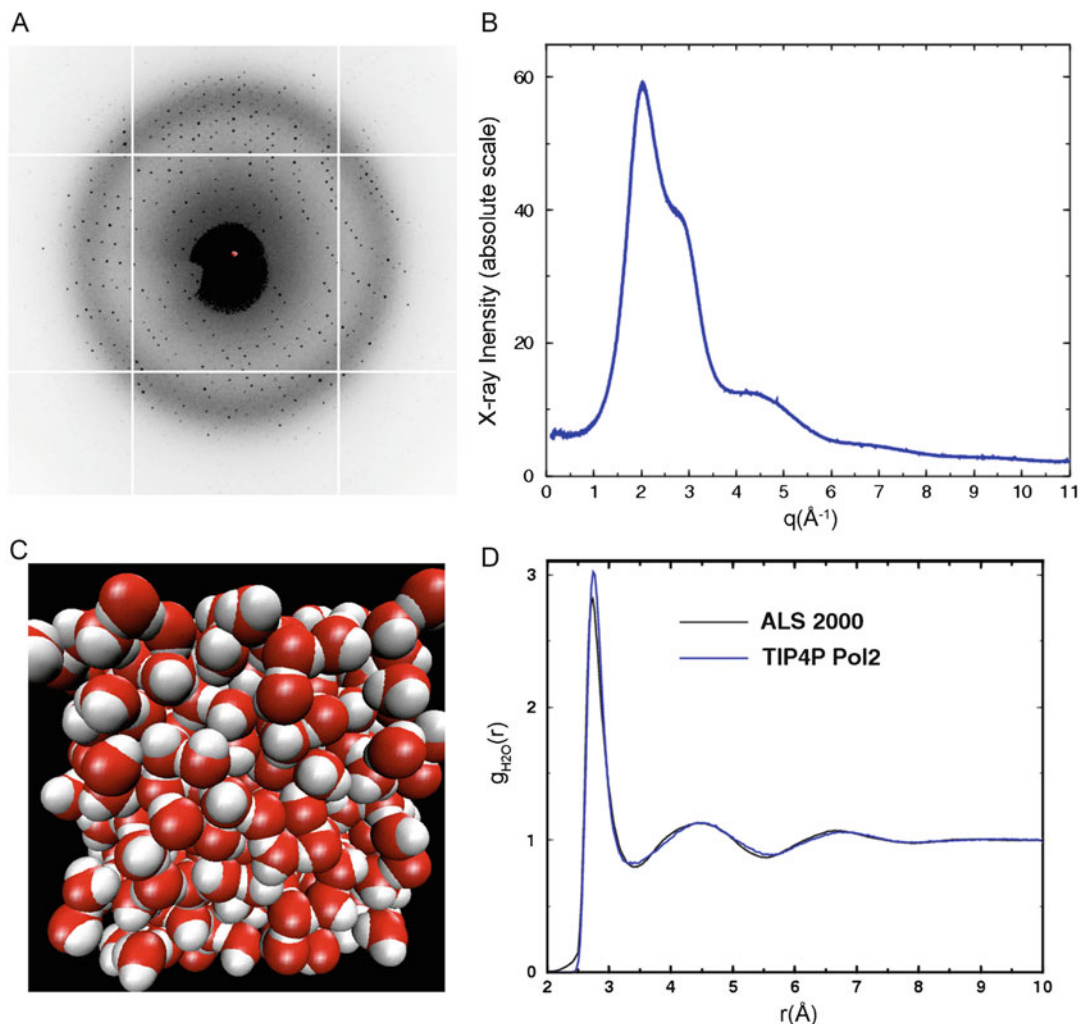


Fig. 10.4 Scattering from water structure. (a) Underlying protein crystallographic data is a ring due to water. Because the protein is organized on a repeating lattice in the crystal, the diffraction spots must be characterized in two dimensions on the detector. In contrast the structures related to water are oriented in all directions relative to beam producing a symmetric ring that may be represented in one dimension (b) by integrating around central incident beam. (c) Water structure stems from several

features including hydrogen bonding and van der Waals interaction. (d) The van der Waals interactions prevent water molecules from overlapping which can be seen in the radial distribution function $g_{\text{H}_2\text{O}}(r)$ at small r up to 2.5 \AA . The main correlation is with next nearest neighbors of which there are five. This can be determined by integrating the area under the peak of the first correlation shell which is centered at 2.8 \AA

molecule itself. For this reason, the quantity of interest from Eq. 10.2 is the opposite of what it was for both crystallography and scattering from macromolecules. The $\rho_{\text{H}_2\text{O}}(\mathbf{r})$, that describes a water molecule, in Eq. 10.1 is known and has been tabulated for both X-rays (Morin 1982;

Hura et al. 2000) and neutrons. Using the known $\rho_{\text{H}_2\text{O}}(\mathbf{r})$, a form factor for water has been calculated, $f_{\text{H}_2\text{O}}(\mathbf{r})$. The scattering from water is thus composed of two parts, the scattering within a water molecule (intra-molecular scattering, a known quantity), and the scattering between

different water molecules (inter-molecular scattering) where $\mathbf{r}_j - \mathbf{r}_k \neq 0$ as defined in the second part of Eq. 10.4. Intra-molecular scattering, $I_{\text{intra}}(q) = \langle f(q)f^*(q) \rangle$, can be calculated, however this calculated quantity can only be utilized when the measurement has been calibrated on an absolute scale. Absolute measurements can be challenging since few detectors are accurately calibrated. However, using calibrants or extremely high angles one can isolate the inter-molecular scattering $I_{\text{inter}}(q)$.

As in crystallography where a formalism is introduced that anticipates molecules will be on a lattice with extracted indices, for non-crystalline materials a construct has been created that anticipates the kind of inter-molecular structure we expect in liquids. This formalism is called the radial pair correlation function or $g(\mathbf{r}_j - \mathbf{r}_k)$. The $g(\mathbf{r}_j - \mathbf{r}_k)$ can be considered a probability weighting function that describes the probability that a neighboring water molecule has a specific orientation and distance relative to any other given water molecule. The quantity $\rho_o g(\mathbf{r}_j - \mathbf{r}_k) d\mathbf{V}$ is the expected number of molecules that will be found at a distance \mathbf{r}_k in a volume element $d\mathbf{V}$ from a molecule at \mathbf{r}_j . The constant ρ_o is the average molecular density that is macroscopically measurable, calculable from $\sim 1 \text{ g/cm}^2$ for water. By introducing this term, we can exchange one of the sums over all molecular pairs in the second part of Eq. 10.4 for an integral over the probability weighting function as in Eq. 10.5. This assumes that the population distribution of configurations has reached an equilibrium within the liquid.

The correlated structures in water will be in every orientation relative to the X-ray beam involving a spherical integration, with results similar to those described in the preceding section. All directional information is lost and the exponential term is further reduced to a ‘‘sinc’’ ($\sin(x)/x$) function of scalars and $g(r)$. The result of the spherical integration is shown in Eq. 10.5 (neglecting a term that is only of significance with strongly absorbing material). The function

$g(r)$ has several key properties. The value of $g(r)$ describes the relative number of scatterers outside the bulk density having a center-to-center distance of r from a molecule, as shown in Fig. 10.4d. At small distances within the van der Waals diameter, $g(r)$ will have a value of zero since this presents a no overlap zone. At large r , $g(r)$ will be 1 as correlation has been lost and numbers have reached bulk average density, ρ_o . The vectorial $g(\mathbf{r}_j - \mathbf{r}_k)$ can be calculated from a molecular dynamics calculation as can be its scalar form $g(r)$.

$$I(q) = 2\langle f(q)f^*(q) \rangle + \left\langle \sum_j \sum_{k \neq j} f(\mathbf{q})f^*(\mathbf{q}) e^{i(\mathbf{r}_j - \mathbf{r}_k) \cdot \mathbf{q}} \right\rangle \quad (10.4)$$

$$I(q) = 2\langle f(q)f^*(q) \rangle + \left\langle \sum_k f(\mathbf{q})f^*(\mathbf{q}) \int_V \rho_o g(\mathbf{r}_j - \mathbf{r}_k) e^{i(\mathbf{r}_j - \mathbf{r}_k) \cdot \mathbf{q}} d\mathbf{V}_j \right\rangle$$

$$I(q) = 2\langle f(q)f^*(q) \rangle \left(1 + 4\pi\rho_o \int_0^\infty r^2 (g(r) - 1) \frac{\sin(qr)}{qr} dr \right) \quad (10.5)$$

Water was almost certainly one of the first targets of X-rays and neutrons as it is both easy to attain and of critical importance. However, since water is of such great importance, there is a tremendous demand for precision. The $g(r)$ function for water remains intensively studied and even debated (Brookes and Head-Gordon 2015; Clark et al. 2010; Gallo et al. 2016; Amann-Winkel et al. 2016). Each structural detail has wide spread implications. Though not presented here, further refinements have been made by starting at the atomic ($\rho_o(\mathbf{r})$ and $\rho_H(\mathbf{r})$) rather than the molecular level. Working from this basis allows an exploitation of a unique property of neutron scattering. Since deuterium scatters neutrons more strongly than hydrogen one can make use of a contrast change between deuterated water and hydrogenated water for extraction of the pair correlation function of hydrogens $g_{\text{HH}}(r)$. As X-rays are scattered by electrons, scattering experiments report mainly

on the oxygens $g_{OO}(r)$. This is particularly true since oxygen electronegativity will draw electrons from hydrogen (Head-Gordon and Hura 2002).

A challenge for converting scattering data to real spatial information is that the two are related through a Fourier transform. For early investigators in particular but also of concern today are three features of scattering data that stymie Fourier transformation. (1) Noise in data adds unphysical Fourier terms, (2) sparsely sampled data from point detectors or other experimental factors may mean missing Fourier terms and (3) data are always truncated both at high and low angles with disastrous effects for Fourier transformation. When direct transformation is attempted on truncated data, an infinite set of non-physical Fourier terms are required. Modern detectors and bright sources have improved signal, increased the sampling and the angular range collected, greatly reducing challenges faced by early experimentalists. Thus an understanding of the structure of water as determined by scattering has been emerging.

General features that are agreed upon from analysis of $g(r)$ is that a water molecule strongly influences its nearest neighbors. The $g(r)$ as determined from X-ray scattering measurements from water is shown in Fig. 10.4d and has on average 5 nearest neighbors, that sit 2.8 Å from the center of any given water molecule. The presence of a water molecule perturbs structure as far as 10 Å away. This first coordinated shell is followed by a drop in density below bulk levels. Two more peaks are discernable in addition to the first, showing the minimum distance a water molecule maintains its influence. These length scales are larger than cavities within macromolecules affecting many important phenomena such as metabolite and drug binding. Thus accounting for the structural influence of water remains a major challenge for *in silico* based drug screening among other fields.

10.4 Measuring Changes in Distance Distributions Over Long Length Scales Using Labels

We now move to measuring changes in large scalar distances within a macromolecular assembly. Such measurements in solution can provide key biological insights. Outside of a crystalline state, there is almost always a population distribution of distances that cannot be adequately represented with a single structure. Biological systems may sample important states infrequently and thus these states occur in only a small subset of a population. The challenge for solution based techniques is to define this distribution and changes in this distribution as a function of some perturbation. A large variety of techniques have evolved to quantify these dynamic distributions, each with assets that are worth contrasting.

Like many other techniques, scattering can make use of labels to aid in following specific parts of macromolecules, increasing signal and accuracy. There are several types of labels and several ways of measuring the distance between them. This section is focused on the use of heavy metal labels in the context of X-ray scattering (SAXS) (schematically illustrated in Fig. 10.5a). Conceptually, the experiment is quite simple, however there is still a substantial investment into synthesis relative to other label based techniques for which synthesis has become more routine.

SANS based approaches have been used to measure distance distributions for many years. Selective deuteration of parts of a macromolecule adds extra scattering cross-section to that part. Collecting data from such samples in a mixture of H₂O and D₂O can make the non-deuterated portion invisible. This approach was applied to identify the relative distance between ribosomal components before atomic resolution structures were available. Several other interesting systems have been elucidated

this way as discussed elsewhere in this book. Increased access to SANS instruments plus their increased brightness are certain to profoundly increase this type of application.

10.5 FRET and EPR for Measurement of Scalar Distances

Before we focus on SAXS we draw attention to important alternative non-scattering based techniques. When molecular rulers are required, FRET and EPR are often utilized with excellent effect. FRET and EPR labels are commercially supported, reducing sample preparation challenges. They have been powerfully applied in scenarios for which scattering techniques are difficult. For example, FRET can be applied *in vivo* and as a single molecule technique. EPR can be conducted at low concentrations and with membrane proteins. For these reasons FRET and EPR should be strongly considered for at least complementary information to scattering.

FRET and EPR also have specific challenges. Both techniques have an optimal range for distance measurements, beyond which they are no longer reliable. For FRET, this range is usually from 1.5 to 6 nm depending on the dynamics of the biomolecule and the size of the label (Lam et al. 2012). For EPR, the optimal range can extend from 1.5 to 2.5 nm for continuous-wave EPR, and up to 8 nm for pulsed wave EPR (Schiemann and Prisner 2007).

In addition to limitations in distance measurements, both FRET and EPR have specific experimental challenges. For FRET to accurately measure distance, both labels need to have the freedom to sample all rotational orientations. Limiting the rotation of one label relative to the other can increase measurement error significantly, from 10% at 5 nm to 50% at 1.5 nm. For EPR, the sensitivity of the spin label to the environment can be both a blessing and a curse. To increase the signal-to-noise for EPR, measurements are often taken at 50–80 K, sometimes for 10–12 h.

For systems and questions where these experimental limitations do not pose a challenge, both

FRET and EPR will provide significant insights. Scattering remains inherently complementary. Scattering measurements can be conducted with and without FRET or EPR labels to determine the influence of labels.

10.6 Scattering from Metal Labels for SAXS Measurements

X-rays scatter from electrons, thus heavy elements scatter more strongly than lighter elements. Most biological SAXS measurements are difference experiments where the solvent is subtracted from the solution containing sample. When this is done, the relative signal of each contributor is scaled not by the square of the electron density but by the square of the electron density difference between the scattering object and solvent. The average electron density is 0.332, and 0.44 electrons/Å³ for water and protein respectively. For gold, the electron density is approximately 4.6 electrons/Å³. Using these values, in a difference experiment the scattering intensity at zero angle of a gold particle is 1650-fold larger than that of a protein of equivalent size (Fig. 10.5b). Differences of this order of magnitude are usually worthwhile to pursue despite challenges in preparation. For example, concentrations can be reduced to the nanomolar range or time scales can be reduced to milliseconds.

Gold labeling of biological systems has relied heavily on gold-sulfur bonding though other strategies are possible. Cysteines are a natural target in proteins while thiolated bases can be incorporated into DNA. As gold nanoclusters present a large surface, care must be taken to ensure the desired labeling ratio is eventually purified. Once the gold-sulfur bond is made, further reactions must be reduced. Thus the portion of the gold surface that is not involved in the desired bonding is protected by other chemical groups like thiolated polyethylene glycols. For control of bonding, the redox state of solutions during preparation must be carefully controlled as the gold-sulfur bond competes with potential disulfide bonding.

The larger the scattering from the label the simpler the subsequent analysis will be for the extraction of distances between two labels. However, any label can be disruptive to the biological system under investigation and a variety of heavy atoms and heavy atom clusters have been found useful including mercury (Vainshtein et al. 1980), lead (Grishaev et al. 2012), rubidium (Horkay et al. 2006), and terbium (Miakelye et al. 1983). Many of these studies also manipulated the scattering contrast by either using anomalous scattering properties of the metal atom or changing the solvent scattering to match that of the biological component. Depending on the strength of the signal and size of the label, data collection and analysis increase in complexity (Zettl et al. 2016; Mathew-Fenn et al. 2008a, b).

10.7 Extracting Length Information from the Scattering Curve

Contributions to the total scattering from labeled systems can be conveniently grouped. There are five types of contributions: the intra-label, intra-biomolecule, inter-label, inter-biomolecule, and finally scattering due to correlations between biomolecules and labels. Analysis of these terms vary in complexity. In the simplest case, the labels scatter so strongly that the scattering contribution of the biological macromolecule are negligible (Hura et al. 2013b). In this case only two terms are significant: the intra-label and inter-label terms.

With labels that scatter on the same scale as the biological macromolecule, all terms must be considered. In an effort to measure basic properties of DNA, small gold labels of 1.5 nm diameter have been used (Mathew-Fenn et al. 2008a, b). The intra- and inter-biomolecule scattering can be measured on the system without labels. The more difficult component is the scattering cross term due to correlations between the biomolecule and the label. Several strategies may be applied including the measurement of the system labeled at each point independently or modifying the scattering power of the label

using resonant X-ray energies (Zettl et al. 2016) or varying label size or composition as reported in several studies.

In a system using equivalent labels at all labeling points and that scatter overwhelmingly, the analysis is similar to that used in the previous section for water. A distinction can be made between the experimental observable desired from labels relative to that from water. In the case of water, the absolute number of coordinating waters is an important experimental result. The number of coordinated labels is almost always known and if uncertain can be tested by varying the labeling strategy. For example one can test the agreement between the scattering of the label alone and the macromolecule labelled at one point. If these results are in poor agreement, the macromolecules may be multimerizing, adding additional label correlations that must be accounted for. Defining a distribution of labels relative to the bulk density is not necessary and so rather than work with $g(r)$ which is a comparison to bulk density, a relative distribution is desired. Thus, starting from Eq. 10.4 we define a weighing distribution $P(D)$ where D is the distance between label j and k . The labels all have the same $f(q)$ so we can utilize Eq. 10.6.

$$I(q) = 2\langle f^2(q) \rangle + 2\langle f^2(q) \rangle \int_0^\infty P(D) \frac{\sin qD}{qD} dD \quad (10.6)$$

Rearranging terms to focus on the inter-label distance distribution and taking into account both concentration factors and instrumental parameters with two constants (k_1 and k_2) we arrive at the correlation scattering function (CSF), which is a Fourier transform of $P(D)$.

$$CSF = \frac{I(q)}{k_1 \langle f^2(q) \rangle} - k_2 = \int_0^\infty P(D) \frac{\sin qD}{qD} dD$$

Experimentally, k_1 and k_2 can be determined. However, they can also be treated as fitted parameters. The CSF should oscillate about 0 and at wide q the inter-particle contribution should be negligible as the inter-label distance must be larger than the label size. Drifts from an

oscillation about 0 indicate either some level of aggregation in the labels.

We have applied this scenario to monitor protein mediated DNA repair (Hura et al. 2013b). Labeling both ends of damaged DNA with nominally 5 nm diameter gold labels, we followed the end-to-end distribution as proteins and metabolites in the repair pathway were added. We contrasted both short (31 base pair) and long (up to 71 base pair) DNA. The shorter DNA substrate accommodates a single protein

footprint analogous to what can be done with FRET. The longer DNA substrate accommodates multiple proteins allowing the observation of cooperative effects common in DNA repair processes. Example results from this study are shown in Fig. 10.5c–e where we measure distances between labels that are 30 nm apart. These results can be extended to longer DNA strands or length scales as most modern SAXS instruments can sufficiently capture small angle data.

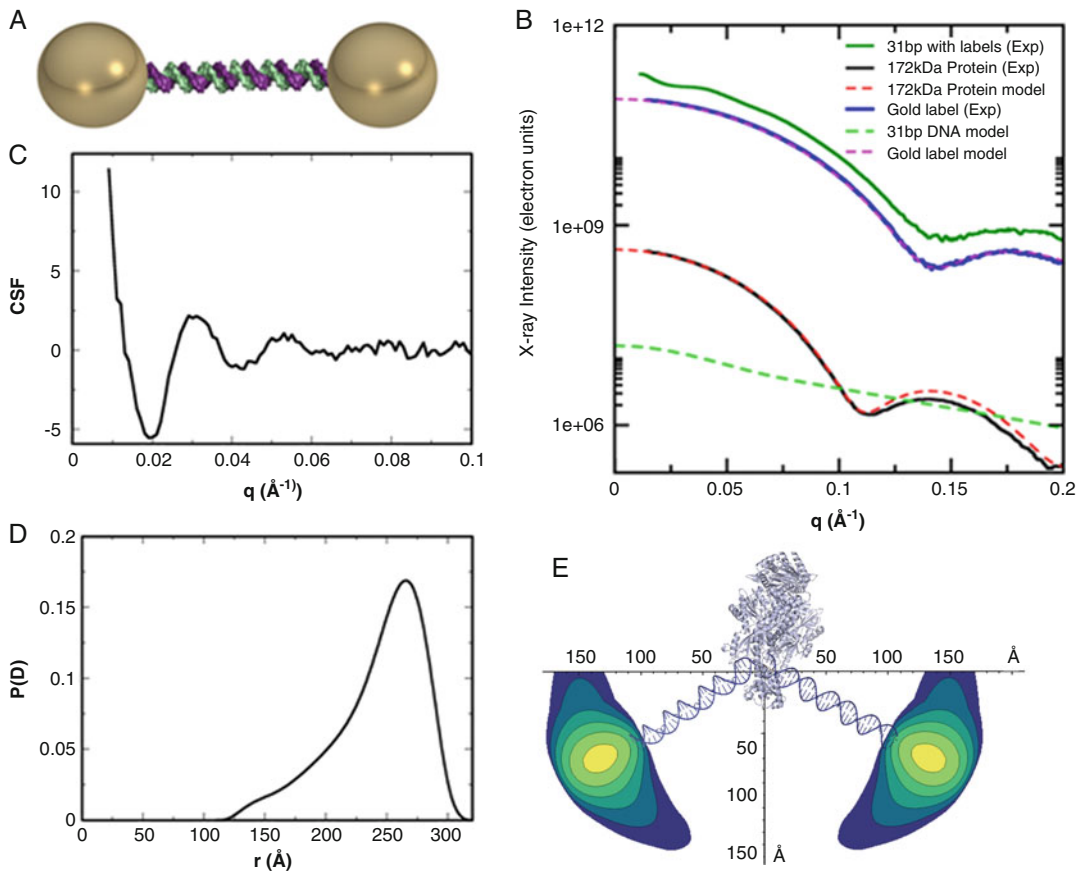


Fig. 10.5 SAXS measures of long-distance distributions with gold labels on DNA. (a) Schematic representation of gold end-labeled DNA. (b) Experimentally measured scattering power of 5 nm gold is 5400 times higher than DNA and 500-fold greater than a globular protein of two times its diameter. (c) The CSF is derived from dividing two SAXS profiles, the labeled system through by the label alone. (d) The $P(D)$ distribution characterizing the

distances between labels as DNA is manipulated by a DNA processing enzyme. (e) Contour maps of the gold labels can be drawn based on the distribution shown in (d) combined with crystallographic information that the protein system bends DNA. The bending of the DNA is dynamic, sampling very dramatic bending angles that could not be deduced crystallographically

10.8 Conclusion

SAS as employed to study biological macromolecules in solution is a very flexible and powerful technology with widespread application. For most samples, a minimal amount of preparation is required to provide a comprehensive characterization of macromolecular structure. High flux sources and new detectors are capable of characterizing wide temporal and spatial scales – all with one sample and data collection. Here we worked through some details of analysis for probing distances between molecules in a liquid which can be pushed down to 0.1 nm resolution. Large area detectors provide access to the necessary angles that can be used to characterize the details of hydration layers around proteins. More work is required on tool development to utilize this information that has become routinely available.

For specialized cases where the organization or movement of subassemblies within a larger assembly is of central importance, samples may be modified so that these pieces have additional contrast. For X-rays, specific points may be labeled with metallic nanoclusters or for neutrons, regions may be deuterated. Since required sample quantities are already quite small and continue to decrease, the same sample preparation may be used to study the labeled system in a variety of contexts, providing unique insights into function. Here, we considered some of the detailed analysis required to extract information from labels separated by distances of 50 nm or greater.

While there have been big leaps in recent progress for EM and with the free electron laser for X-rays, advances in SAS data collection and analysis have been more wide spread and continuous. Access and utilization has grown, creating an ever larger community that contributes to analysis tools and interpretation. We anticipate that due to its widespread applicability and throughput, SAS will increasingly be looked to for complementary and unique structural information on a rapidly expanding set of targets from genomic and macromolecular engineering fields.

10.9 Few Assorted Experimental/Computational Tips

SAS can monitor high resolution structural changes in solution when conducted as a relative measurement to other SAS data or an atomic resolution model.

Distance distributions are not always calculated from SAS in the same way and therefore do not always quantify the same scattering density distribution. Make sure you understand the assumptions that are part of a particular distance distribution.

In quantifying distances between scattering density, utilize any information that may be of value, whether it be bulk density or known semi-periodicity particular to the sample, to create the most intuitive distance distribution for your system.

Distance distributions are never direct Fourier transforms of SAS data as SAS data is finite and are often derived from fitting data. Always examine the quality of the fit of the Fourier transform.

References

- Amann-Winkel K, Bellissent-Funel MC, Bove LE, Loerting T, Nilsson A, Paciaroni A, Schlesinger D, Skinner L (2016) X-ray and neutron scattering of water. *Chem Rev* 116:7570–7589
- Andresen K, Jimenez-Useche I, Howell SC, Yuan CL, Qiu XY (2013) Solution scattering and FRET studies on nucleosomes reveal DNA unwrapping effects of H3 and H4 tail removal. *PLoS One* 8:e78587
- Baldock C, Oberhauser AF, Ma L, Lammie D, Siegler V, Mithieux SM, Tu Y, Chow JY, Suleman F, Malfois M, Rogers S, Guo L, Irving TC, Wess TJ, Weiss AS (2011) Shape of tropoelastin, the highly extensible protein that controls human tissue elasticity. *P Natl Acad Sci USA* 108:4322–4327
- Bergmann A, Fritz G, Glatter O (2000) Solving the generalized indirect Fourier transformation (GIFT) by Boltzmann simplex simulated annealing (BSSA). *J Appl Crystallogr* 33:1212–1216
- Brookes DH, Head-Gordon T (2015) Family of oxygen-oxygen radial distribution functions for water. *J Phys Chem Lett* 6:2938–2943
- Clark GNI, Hura GL, Teixeira J, Soper AK, Head-Gordon T (2010) Small-angle scattering and the structure of ambient liquid water. *Proc Natl Acad Sci U S A* 107:14003–14007

- Edwards AL, Matsui T, Weiss TM, Khosla C (2014) Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *J Mol Biol* 426:2229–2245
- Falk SJ, Guo LY, Sekulic N, Smoak EM, Mani T, Logsdon GA, Gupta K, Jansen LE, Van Duyne GD, Vinogradov SA, Lampson MA, Black BE (2015) Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science* 348:699–703
- Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr* 42:342–346
- Gallo P, Arnann-Winkel K, Angell CA, Anisimov MA, Caupin F, Chakravarty C, Lascaris E, Loerting T, Panagiotopoulos AZ, Russo J, Sellberg JA, Stanley HE, Tanaka H, Vega C, Xu LM, Pettersson LGM (2016) Water: a tale of two liquids. *Chem Rev* 116:7463–7500
- Grishaev A, Anthis NJ, Clore GM (2012) Contrast-matched small-angle x-ray scattering from a heavy-atom-labeled protein in structure determination: application to a lead-substituted calmodulin-peptide complex. *J Am Chem Soc* 134:14686–14689
- Hammel M, Amlanjyoti D, Reyes FE, Chen JH, Parpana R, Tang HY, Larabell CA, Tainer JA, Adhya S (2016) HU multimerization shift controls nucleoid compaction. *Sci Adv* 2:e1600650
- Head-Gordon T, Hura G (2002) Water structure from scattering experiments and simulation. *Chem Rev* 102:2651–2669
- Horkay F, Hecht AM, Rochas C, Bassler PJ, Geissler E (2006) Anomalous small angle x-ray scattering determination of ion distribution around a polyelectrolyte biopolymer in salt solution. *J Chem Phys* 125
- Hura G, Sorenson JM, Glaeser RM, Head-Gordon T (2000) A high-quality x-ray scattering experiment on liquid water at ambient conditions. *J Chem Phys* 113:9140–9148
- Hura GL, Menon AL, Hammel M, Rambo RP, Poole FL II, Tsutakawa SE, Jenney FE Jr, Classen S, Frankel KA, Hopkins RC, Yang S-J, Scott JW, Dillard BD, Adams MWW, Tainer JA (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods* 6:606–U683
- Hura GL, Budworth H, Dyer KN, Rambo RP, Hammel M, McMurray CT, Tainer JA (2013a) Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat Methods* 10:453–454
- Hura GL, Tsai CL, Claridge SA, Mendillo ML, Smith JM, Williams GJ, Mastroianni AJ, Alivisatos AP, Putnam CD, Kolodner RD, Tainer JA (2013b) DNA conformations in mismatch repair probed in solution by X-ray scattering from gold nanocrystals. *Proc Natl Acad Sci U S A* 110:17308–17313
- Koch MHJ, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* 36:147–227
- Lam AJ, St-Pierre F, Gong YY, Marshall JD, Cranfill PJ, Baird MA, McKeown MR, Wiedenmann J, Davidson MW, Schnitzer MJ, Tsien RY, Lin MZ (2012) Improving FRET dynamic range with bright green and red fluorescent proteins. *Nature Methods* 9:1005. –+
- Mathew-Fenn RS, Das R, Silverman JA, Walker PA, Harbury PAB (2008a) A molecular ruler for measuring quantitative distance distributions. *PLoS One* 3
- Mathew-Fenn RS, Das R, Harbury PAB (2008b) Remeasuring the double helix. *Science* 322:446–449
- Miakelye RC, Doniach S, Hodgson KO (1983) Anomalous x-ray-scattering from terbium-labeled parvalbumin in solution. *Biophys J* 41:287–292
- Morin LRM (1982) Molecular-form factors and photon coherent scattering cross-sections of water. *J Phys Chem Ref Data* 11:1091–1098
- Nielsen SS, Toft KN, Snakenborg D, Jeppesen MG, Jacobsen JK, Vestergaard B, Kutter JP, Arleth L (2009) BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *J Appl Crystallogr* 42:959–964
- Poitevin F, Orland H, Doniach S, Koehl P, Delarue M (2011) AquaSAXS: a web server for computation and fitting of SAXS profiles with non-uniformly hydrated atomic models. *Nucleic Acids Res* 39:W184–W189
- Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 40:191–285
- Schiemann O, Prisner TF (2007) Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Q Rev Biophys* 40:1–53
- Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A (2016) FoXS, FoXSDock and multiFoXS: single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res* 44:W424–W429
- Semenyuk AV, Svergun DI (1991) Gnom – a program package for small-angle scattering data-processing. *J Appl Crystallogr* 24:537–540
- Stingaciu LR, O’Neill H, Liberton M, Urban VS, Pakrasi HB, Ohl M (2016) Revealing the dynamics of thylakoid membranes in living cyanobacterial cells. *Sci Rep* 6:19627
- Vainshtein BK, Feigin LA, Lvov YM, Gvozdev RI, Marakushev SA, Likhtenshtein GI (1980) Determination of the distance between heavy-atom markers in hemoglobin and histidine-decarboxylase in solution by small-angle x-ray-scattering. *FEBS Lett* 116:107–110
- Williams RS, Dodson GE, Limbo O, Yamada Y, Williams JS, Guenther G, Classen S, Glover JNM, Iwasaki H, Russell P, Tainer JA (2009) Nbs1 flexibly tethers Ctp1

and Mre11-Rad50 to coordinate DNA double-strand break processing and repair. *Cell* 139:87–99

Zettl T, Mathew RS, Seifer S, Doniach S, Harbury PAB, Lipfert J (2016) Absolute intramolecular distance

measurements with angstrom-resolution using anomalous small-angle x-ray scattering. *Nano Lett* 16:5353–5357