A beginner’s guide to solution small-angle X-ray scattering (SAXS)

Beginner’s Guide

Melissa Gräwert and Dmitri Svergun (European Molecular Biology Laboratory, Hamburg Unit, Germany)

Solution small-angle X-ray scattering (SAXS) is a powerful technique, which is complementary to different bioanalytical and structural methods. With straightforward data collection procedures and minimal restrictions in sample environment, information on size, shape and conformational flexibility of biological macromolecules and complexes in near native solutions can be rapidly obtained. Here, we highlight the recent developments that have advanced SAXS to a versatile tool with the capability to enrich almost any biochemical study.

Introduction

Solution small-angle X-ray scattering (SAXS) occupies a unique niche in the landscape of methods available for biological research. As implied by its name, the measurements are performed in solution making SAXS a powerful tool to analyse macromolecules in their quasi-native environment. The sample can easily be probed as a function of a number of variables such as temperature, ionic strength, pH, time etc. From the scattering experiments, overall molecular size and weight, as well as degree of compactness and flexibility, are readily determined. In addition to this analytical information, the scattering profiles also comprise structural information. For many cases it is possible to directly reconstruct the overall shape of the macromolecule—without any prior knowledge about the system, albeit at low resolution (ca. 10–20 angstroms; Å). SAXS is even more powerful when combined with other higher resolution structural methods such as nuclear magnetic resonance spectroscopy, cryogenic electron microscopy and protein crystallography. Atomic substructures or incomplete models can be validated, expanded or altered giving meaningful explanations of the structural behaviour of a system in solution.

The ability to collect and interpret SAXS data was demonstrated by French scientist Andre Guinier as early as 1938. From there, it took more than half a century for it to achieve its current status: a high-throughput technique addressing a vast spread of biological questions. SAXS is employed by experts as well as novices for either pure structural analysis or as part of integrative studies in combination with complementary methods. Most recently, pharmaceutical companies have started adding SAXS analysis to their workflows: the characterization of various biopharmaceuticals makes it profitable to invest in SAXS analyses.

Major milestones that fuelled the development of solution SAXS

There are three main aspects that have led to the increase in popularity of solution SAXS.

1) Developments in hardware: improved X-ray sources (both synchrotrons and laboratory cameras) combined with the integration of robotic systems have significantly reduced not only the required sample volume but have also tremendously accelerated the experiments. Nowadays, even tricky samples, which are difficult to obtain in large amounts, can be examined within minutes (Figures 1 and 2).

2) Enhanced methods for data analysis, particularly automation of SAXS analysis, has played the most significant role. In the 1990s, a number of software...
tools became available to facilitate the interpretation of scattering patterns from biological systems. Three important advances represent the groundwork for all further progress (Figure 3): i) the possibility to accurately calculate theoretical scattering curves from (high- as well as low-resolution) models and compare these with the experimental data, ii) the implementation of algorithms to generate the distance probability function, p(r), from the measured intensities and iii) the ability to automatically generate simple shape models \textit{ab initio} (i.e. from scratch) within minutes of data collection.

The more recent development of sophisticated methods that have supported the important role of solution SAXS in integrative structural biology are summarized later in the article (Figures 4 and 5). 3) Last, a very important driving force for mainstreaming biological SAXS has been the community effort to establish standards for the validation of SAXS data. A task force was formed in 2012 to define essential criteria for data reporting and interpretation. Publishing guidelines were agreed and commitments to adhere to these were made by the wider biological SAXS community. Furthermore, it is now possible to share and cross-check data through dedicated databases. These measures not only lead to validation and authenticity checks of the data, but also make the method more approachable for non-experts. With these guidelines the novice user can be more assured in the process of data analysis.

In this Beginner’s Guide to SAXS, we will first take a closer look at the experimental set-up and sample requirements to perform a SAXS experiment. Then, advanced data interpretation approaches are described and illustrated by selected examples. We conclude with a summary and outlook of the future of SAXS.

**Experimental set-up**

**Basic principles of the scattering experiment:**

The biological X-ray scattering experiment itself is conceptually simple. The dissolved macromolecules are illuminated with X-rays of a defined energy (i.e. at a specific X-ray wavelength, \(\lambda\), typically about 1 Å; Figure 1A). The intensities of the scattered X-rays, \(I\), are recorded on an appropriate area detector. As the sample is isotropic, i.e. the molecules are tumbling around independently of each other, the intensity distribution can be radially averaged. This results in a 1D scattering curve, usually plotted as \(I(s)\) vs \(s\); where \(s = 4\pi\sin(\Theta)/\lambda\) and \(2\Theta\) is the angle between the incident and scattered beam. A very important step is to record a second measurement of the identical solution lacking the sample of interest. This profile contains the scattering contributions from the solvent, the sample holder as well as the machine set-up and is used for background subtraction (Figure 1B). Overall, the electron density of proteins (~0.43 electrons/Å³) are only slightly higher than the average electron density of water (0.33 electrons/Å³). It is this discrepancy that gives rise to the (rather weak) scattering signal. The small difference in electron density decreases even further with the addition of components to the buffer. Many first-time SAXS users are surprised just how sensitive the method is to small discrepancies between correctly and incorrectly matched buffers.

Another important prerequisite for a successful SAXS experiment is to ensure sample purity. Scattering signals of the sample as well as impurities are simply summed up; in particular, the presence of aggregates displays a challenge. These might be present in small numbers (low concentrations) but due to their large size and the fact that the scattering is proportional to the square of the particle volume (V²) they can lead to severe alterations of the scattering curves, often making these uninterpretable.

As an experimentalist, the two main challenges lie in providing pure samples and ensuring that the buffer does indeed match the composition of the sample.

![Figure 1. SAXS data collection. A) Schematic drawing of an experimental set-up. B) Exemplary presentation of data reduction. C) Comparison of two different SAXS beamlines highlighting the technological advances.](image-url)
solvent. Typical sample volumes needed are 40–200 µl at concentrations of 5–10 mg/ml (comparable to other structural studies).

**New developments in hardware**

The higher demand for biological solution SANS has fuelled the development of beamlines suitable for, and dedicated to, such experiments at most synchrotron facilities (see https://lightsources.org/). At modern third-generation synchrotrons, sample volumes and exposure times are reduced due to small but powerful beam sizes (Figure 1C). Here, a new challenge is posed: perturbation of the sample due to radiation damage. The photolysis of water leads to the production of free hydroxyl (OH•) and hydroperoxyl (HO2•) radicals as well as solvated electrons. Moreover, highly reactive free radicals are formed by X-ray absorption of proteins containing higher atomic mass elements such as metallo-proteins. Ideally, the aim in data collection is to outrun radiation damage and collect data before the sample is destroyed/aggregated.

To reduce the damage in synchrotron set-ups the sample is pushed through the capillary during exposure. Furthermore, small molecules that function as radical scavengers (for example 3–5% glycerol) can be added to the sample (and buffer). To assess the radiation damage, the data is collected as short individual frames, which are compared with each other. Frames that show alterations after initial exposure are then discarded and not used for further processing. Unfortunately, it is still not possible to predict whether a sample might be prone to radiation damage prior to a SANS experiment and thus, this must be treated on a case-by-case basis, and on-site ‘sample pampering’ might be required.

The progress in X-ray sources has been accompanied by advances in X-ray detectors. Solution SANS profited strongly from single-photon-counting pixel detectors with negligible readout dead-times. These are important prerequisites for time-resolved SANS measurements (see the Future Outlook section).

Despite the many available synchrotron beamlines offering biological solution SANS, there can still be difficulties in obtaining beam time. This limitation, coupled with the fresh interest of pharmaceutical companies to add SANS to their workflows, has renewed the overall interest in laboratory X-ray sources. The latter are, in many cases, able to yield data of sufficient quality to gain information about the quaternary static structure and some protein characterization facilities are now offering the possibility of in-house SANS analysis.

Lab-based sources and solution SANS beamlines have strongly profited from the developments in sample delivery robotics. These ensure precise and efficient sample loading and cell cleaning, reducing the overall measurement time and ensuring consistency in sample filling during long beamshifts (especially at night).

One of the most recent developments of biological SANS beamlines has been the implementation of SEC-SANS. Here, SEC stands for size-exclusion chromatography and refers to the use of specific columns attached to a HPLC set-up. With this, different components comprised within a sample can be separated according to their size. By diverting the outlet of such a column directly to the SANS sample holder, scattering profiles can be collected from the separated components. The first SEC-SANS experiment was reported in 2004. Five years later, the bioSAXS beamline at synchrotron Soleil, SWING, was the first station to offer such a set-up on a regular basis. Now, SEC-SANS is regarded as a standard operation mode by all the major synchrotron facilities. At the P12 beamline in Hamburg, additional detectors are implemented for biophysical sample characterization in parallel (Figure 2). The automation of static SANS through robotic delivery systems and the increase in data quality with SEC-SANS have expanded the application of SANS to sample types
that were previously difficult to measure. Thus, new insights into challenging biological probes can be obtained; including large and transient complexes as well as flexible macromolecules etc. Some of these examples are briefly discussed in the Future Outlook section (Figure 5).

Recent advances in data interpretation

Recent advances in hardware have led to a critical increase in data volume (Figure 1). The interpretation of all these data sets has been facilitated with the implementation of automated pipelines for online data handling. Analysis programs are run in the background during data collection, ensuring that users promptly receive a first summary on data quality, overall parameters and even preliminary shapes. This allows users to evaluate the quality of the data ‘on the spot’ and consequently react to events such as radiation damage, unwanted sample aggregates or mismatched buffers. In a similar manner, the automated processing of SEC-SAXS data is of tremendous value: formerly it would take hours to process one SEC run; now, the evaluation occurs within minutes.

The unique information comprised in a scattering profile and the overall ease of data collection have allowed SAXS to emerge as an important asset in integrative structural biology. Here, the complementary nature of SAXS to many other bioanalytical and structural methods is essential. The latest advancements of modelling methods include:

Rigid body modelling and docking algorithms

As outlined in Figure 4, a number of programs are available to combine existing high-resolution substructures to fully understand the structural properties of a more complex particle in solution. For example, (A) guided by the scattering curve, missing fragments can be added to incomplete atomic structures, (B) complex formation may be studied by allowing subunits to rotate and re-orient themselves to each other or (C) flexible linkers can be built to connect various domains. Many of these programs allow setting various constraints such as symmetry or imposing specific contact sites. Local
refinement starting from a (reasonably good) model is also possible, either manually, by moving and turning substructures in molecular graphic programs or automatically, with the implementation of an algorithm for normal mode analysis (D).

Studying flexibility with an ensemble approach

Initially implemented in the program EOM (Figure 4E), ensemble modelling describes the SAXS profile from a flexible macromolecule with the co-existence of different conformations in solution. These are selected from a large pool of structures that are either randomly generated (based on expected length of the molecule) or based on available substructures to which missing sections are added. Valuable information is gained when comparing the characteristics of the selected confirmations with those of the overall pool (such as radius of gyration, $R_g$, and maximum dimension, $D_{max}$).

Examples: structural studies

The possibilities of using SAXS in biological studies are vast and there are many examples in the literature. Here, we have selected two studies of different medically relevant proteins to summarize various aspects of SAXS data analysis (Figure 5).

Insulin formulations

SAXS plays an important role in drug development, in particular, for formulation studies where SAXS can be applied to study the oligomerization state of drugs at high concentrations (>100 mg/ml) or in the presence of different ions. A typical question to address with SAXS is the nature of oligomeric mixtures. The scattering curves shown in Figure 5A depict the behaviour of an insulin formulation upon dilution. Previously, the simplified assumption had been that stable hexamers dissociate into active monomers. However, with this SAXS analysis it could be shown that the equilibrium is not that simple and higher reversible dodecamer species are present within the formulation. This is seen with the improved fit (red curve) to the scattering data when dodecamers are considered.

Characterization of higher order species

In this example, two different dimeric species of monoclonal antibodies (mAb) are compared (Figure 5B). These are obtained during different processing steps. Thereby, Dimer 2 (yellow), is more compact and structurally more rigid than Dimer 1 (blue). This is observed in (i) comparison of the scattering curves in log plot: Dimer 2 shows a slower descent and has a more pronounced bump typical of multi-domain proteins. (ii) The distance probability function of Dimer 2 shows the typical distribution of a globular protein, but the maximum particle size $D_{max}$ of Dimer 1 significantly exceeds that of Dimer 2. (iii) The reploting of the curve as a so-called dimensional Kratky plot (($sR_g)^2I(s)/I(0)$ vs $sR_g$) offers information on the degree of flexibility. Here, Dimer 2 shows a bell-shaped curve with a peak position at $s = \sqrt{3}$, typical of globular particles, whereas the Kratky plot of Dimer 1 exhibits a plateau. This is expected for extended and possibly flexible proteins. Finally (iv), the averaged ab initio shape reconstruction of Dimer 1 is extended whereas that of Dimer 2 is more compact. As evident from the rotated view, both unbiased models displayed approximately the same thickness and are flat. The obtained SAXS results can be readily related to the functional properties of the two dimeric species.
Future outlook: new sources, new opportunities

In summary, SAXS has emerged as a powerful tool to probe biological macromolecules under a large variety of conditions in solution. With the ensemble approach, SAXS is one of the very few structural tools capable of the quantitative characterization of metastable systems such as multi-domain proteins with flexible linkers, unfolded and intrinsically disordered proteins; these represent a rich and yet unexplored reservoir of new drugs due to their involvement in many disease-related signalling pathways including cancer, cardiovascular disease, amyloidosis, neurodegenerative diseases and diabetes.

The increasing amenability of membrane proteins (MPs) to SAXS also supports further drug development, as a high percentage of MPs are the target of therapeutic agents. The detergents required for solubilization of the MPs previously posed a major problem in SAXS analysis as the presence of free detergent micelles obscures the signal from the protein. However, with SEC-SAXS, scattering data can be collected from loaded micelles alone as they are separated from the free micelles.

From a methodological point of view, considerable effort is being focused on time-resolved SAXS. At modern third-generation synchrotrons, intense X-ray pulses as short as 100 ps can already be delivered with specified set-ups. With sophisticated triggering systems, submillisecond time resolution allows one to follow biological processes in action. Even more impressive specifications are expected with the development of the new (fourth-) generation synchrotrons.

The present progress in cryogenic microscopy questions the future of other high-resolution structural methods. For SAXS, which can be used directly in native solutions, the examples described here highlight its present and future importance both as a stand-alone and as complementary technique. SAXS is unique in many ways including the rapid analysis in solution, observation of changes with time or upon the addition of ligands as well as the characterization of macromolecular interactions and structural flexibility.

A) Oligomeric mixture of Insulin formulation

B) Comparison of higher ordered structures formed during monoclonal antibody production

Figure 5. Examples of SAXS analysis of therapeutically relevant molecules. A) Oligomeric mixture of an insulin formulation. B) Comparison of higher ordered structures formed during mAb production (see reference Plath et al. 2016 in Further reading).
Dr Dmitri Svergun is Group Leader and Senior Scientist at the European Molecular Biology Laboratory Unit in Hamburg, Germany. His group runs and further enhances the dedicated high brilliance synchrotron beamline – P12 at DESY’s third-generation storage ring, PETRA III. He has developed many novel approaches for SANS data analysis and interpretation which are implemented in the world’s most used SANS program package, ATSAS. Furthermore, he is strongly recognized by the field for his leadership through community-focused projects and education. Before moving to Germany, Dmitri was affiliated with the Institute of Crystallography in Moscow. Email: svergun@embl-hamburg.de

Dr Melissa Gräwert is a structural biologist working in Dmitri Svergun’s group. After obtaining her PhD in biophysics, she did a postdoc at the Technical University of Munich (Germany) for which she received a prestigious industrial stipend from the Peter und Traudl Engelhorn Foundation. In 2012, she joined Dmitri’s group as a postdoctoral fellow before becoming a Senior Technical Officer in 2018. Her expertise includes the biophysical and structural characterization of proteins and complexes. Currently, her main research focus is on the implementation and constant development of SAXS as an emerging technique for the characterization of biological therapeutics such as monoclonal antibodies.

Email: graewert@embl-hamburg.de

Further reading

- Comprehensive monograph of biological small-angle X-ray scattering

- Concise review with references to recent developments in biological small-angle X-ray scattering

- Detailed protocol on sample preparation and challenges of biological small-angle X-ray scattering

- Comprehensive overview of various software programs comprised in the ATSAS suite (developed at EMBL Hamburg)

- Scientific article describing the data presented in Figure 5
  - Plath, F., Ringler P., Graff-Meyer A., et al. (2016) Characterization of mAb dimers reveals predominant dimer forms common in therapeutic mAbs. mAbs 8, 928 –940

Further reviews