

# The power of SANS, combined with deuteration and contrast variation, for structural studies of functional and dynamic biomacromolecular systems in solution

Emilie Mahieu<sup>1</sup>, Ziad Ibrahim<sup>1,2,#</sup>, Martine Moulin<sup>2</sup>, Michael Härtle<sup>2</sup>, Bruno Franzetti<sup>1</sup>, Anne Martel<sup>2</sup>, and Frank Gabel<sup>1,\*</sup>

<sup>1</sup>Univ. Grenoble Alpes, CEA, CNRS, IBS, 38044 Grenoble, France

<sup>2</sup>Institut Laue-Langevin, 38044 Grenoble, France

<sup>#</sup>*Present address*: Leicester Institute of Structural and Chemical Biology, University of Leicester, United Kingdom

**Abstract.** Small-angle neutron scattering (SANS), combined with macromolecular deuteration and solvent contrast variation (H<sub>2</sub>O/D<sub>2</sub>O exchange) allows focussing selectively on the signal of specific proteins in multi-protein complexes or mixtures of isolated proteins. We illustrate this unique capacity by the example of a functional protein-degradation system in solution, the PAN-20S proteasome complex in the presence of a protein substrate, ssrA-tagged GFP. By comparing experimental SANS data with synthetic SAXS (small-angle X-ray scattering) data, predicted for the same system under identical conditions, we show that SANS, when combined with macromolecular deuteration and solvent contrast variation, can specifically focus on the conformation of the PAN unfoldase, even in the presence of very large GFP aggregates. Likewise, structural information of native GFP states can be visualized in detail, even in the presence of the much larger PAN-20S unfoldase-protease oligomers, which would dominate the overall scattering signal when using X-rays instead of neutrons.

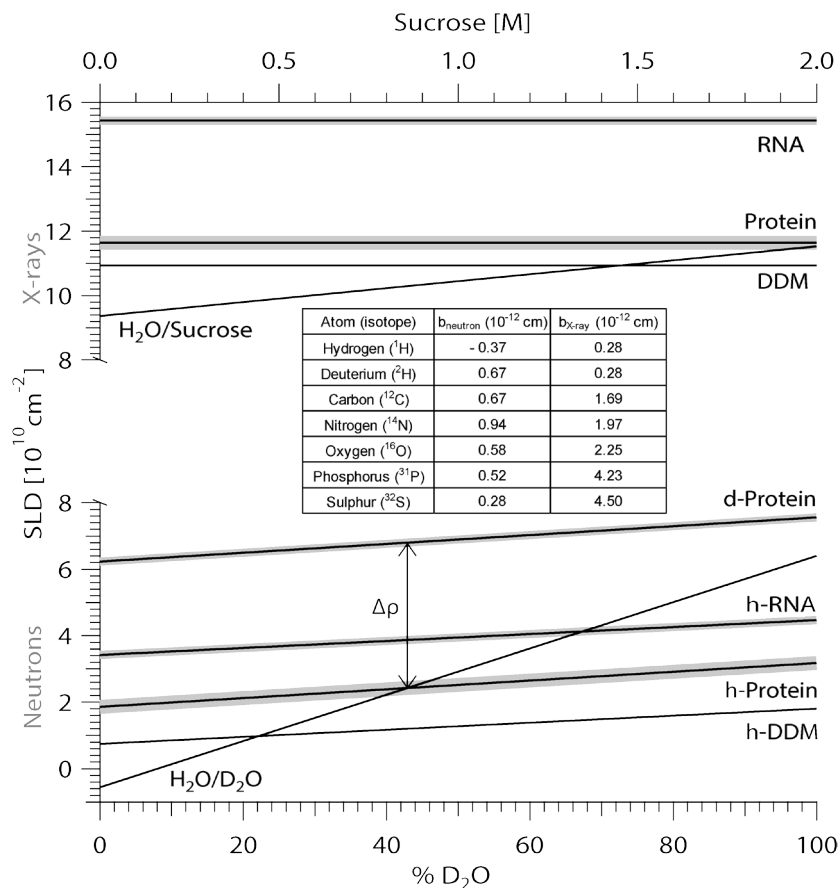
## 1 Introduction

Small-angle X-ray (SAXS) and neutron (SANS) scattering provide structural information on biomacromolecules in solution on the nanometre length scale ( $l$ ). While X-rays are scattered by the electrons in the macromolecules and the buffer, neutrons are scattered by the respective atomic nuclei. The SAS (small-angle scattering) signal of a solubilized biomacromolecule is a Fourier transform, averaged over all possible orientations, of its X-ray (neutron) scattering length density (SLD) difference with respect to the solvent. The signal interpreted is usually an azimuthally averaged 1D curve (intensity vs. scattering angle or  $q$ -value), that informs about the molecular weight of a solubilized particle (via the intensity

---

\* Corresponding author: [frank.gabel@ibs.fr](mailto:frank.gabel@ibs.fr)

scattered at zero angle,  $I(0)$  and its linear dimensions (via the radius of gyration,  $R_G$ ). Low-resolution shapes can be generated from the 1D scattering curves or, if atomic-resolution models are available, their back-calculated SAXS/SANS curve be compared with the experimental data (2).



**Figure 1. Scattering length densities (SLD) of neutrons and X-rays for several biomolecules.** (adopted after (3)).  $\Delta\rho$  indicates the contrast of a d-protein at 42%  $\text{D}_2\text{O}$  where the h-protein is matched. *Inset:* Neutron and X-ray scattering lengths for major light atoms found in biomolecules. DDM (n-dodecyl- $\beta$ -D-maltopyranoside) is shown as an example of detergent used to solubilize membrane proteins. SLDs of DNA are not shown explicitly but are close to those of RNA. The deuteration state of the molecules is not indicated for the SAXS SLDs since X-rays do not distinguish hydrogenated and deuterated molecules.

Since the X-ray scattering length of an atom depends linearly on the number of electrons, X-rays are particularly sensitive to heavier atoms and DNA/RNA molecules scatter, per unit volume, stronger than a protein in an aqueous buffer, due to their sugar-phosphate backbones. It is therefore in principle possible for SAXS to distinguish proteins from RNA/DNA (and lipids/detergents in the case of solubilized membrane proteins) by modifying the solvent electron density by the addition of small, electron-rich molecules such as sugar (Fig. 1). Even though there have been some recent advances using medical contrast media (4), classical contrast agents such as sugar reach, in general, hardly the electron density of proteins (5). They can therefore not match the electron density of RNA/DNA in order to mask their

scattering contribution and to focus on the signal of proteins in a protein-RNA/DNA complex. Moreover, the high (molar) concentrations of most contrast agents pose problems of solubility and integrity of many biomacromolecular systems and practical issues due to high viscosities (4). Finally, SAXS cannot distinguish between different proteins in solution since they have, in general, a comparable average electron density and contrast with the solvent, i.e. their intrinsic electron density cannot be chemically modified in a global and homogeneous way. SAXS data on multi-protein complexes therefore represents an overall envelope (or shape) of the complex but does not, in general, reveal the internal arrangement and conformations of the protein partners.

In contrast to X-rays, neutron scattering lengths depend in an irregular way on the atomic number, but also on the isotope state of a given atom ((1), inset Fig. 1). The most important difference between neutron scattering lengths in biological samples is the one between protons ( $^1\text{H}$ ) and its isotope deuterium ( $^2\text{H}$  or D). Indeed, a continuous change from a 100%  $\text{H}_2\text{O}$  buffer solution to a 100%  $\text{D}_2\text{O}$  (heavy water) solution allows to cover the entire range of natural (i.e. hydrogenated) SLDs of all relevant biomolecules (including proteins, detergent/lipids, and RNA/DNA) (Fig. 1). Even though the exchange of hydrogen against deuterium in the solvent increases the tendency of certain macromolecules to aggregate, it is in general more “gentle” than the one with small electron-rich molecules in SAXS.

Moreover, selective deuteration of bio(macro)molecules (6) allows to increase the natural contrast between chemically different molecules, or create an artificial contrast between chemically similar molecules. **In this way, it is possible to separate the signal from protonated/hydrogenated (h) and deuterated (d) proteins in a complex, and focus on the structural signal from either of them separately. E.g., at about 42%  $\text{D}_2\text{O}$ , the buffer SLD is identical to that of hydrogenated protein, while a perdeuterated (i.e. fully deuterated) protein has a strong contrast** (Fig. 1). Finally, neutrons do not induce radiation damage (as do X-rays) and biological samples can be exposed for several hours and remain biochemically active (7).

Here, we compare experimental SANS data with synthetic SAXS data, calculated for the same experimental conditions, to illustrate the unique capacity of SANS to focus on the structural information of specific molecules in a mixture of several protein species of extremely variable size, shape, and oligomeric state.

As a concrete example, we use SANS data from the PAN-20S proteasome system (Fig. 6), in the presence of an ssrA-tagged GFP which is recognized by PAN-20 as a substrate (8). The proteasome is a large, hetero-oligomeric macromolecular complex, responsible for specific, ATP-driven degradation of dysfunctional, misfolded, or superfluous proteins in the cell (9). In the case of the hyperthermophilic archeon *Methanocaldococcus jannaschii*, it consists of a hexameric, 298 kDa unfoldase PAN, which recognizes and unfolds substrates, and transfers them into the protease 20S particle (751 kDa) where substrates are proteolyzed into small oligopeptide products. We used this robust and thermostable archaeal system to degrade a Green Fluorescent Protein (GFP) variant (28 kDa) containing an ssrA-tag, which allows PAN to recognize it as a substrate (10, 11). By using selective per-deuteration (i.e. full deuteration) of either PAN or GFP in a 42%  $\text{D}_2\text{O}$  buffer, we were able to focus selectively on either of both proteins during the active degradation process: under these labelling and solvent contrast conditions, the signal of the hydrogenated protein species in a reaction mixture are completely masked and the perdeuterated one has a strong contrast (Fig. 1). We demonstrate that this is even true if the hydrogenated species or oligomerization states are several orders of magnitude bigger than the deuterated ones. By thermo-activation at 55 °C, the protein degradation reactions were followed at a time-resolution of 30 seconds during 45 minutes, coupled with online fluorescence spectroscopy (8).

## 2 Material and methods

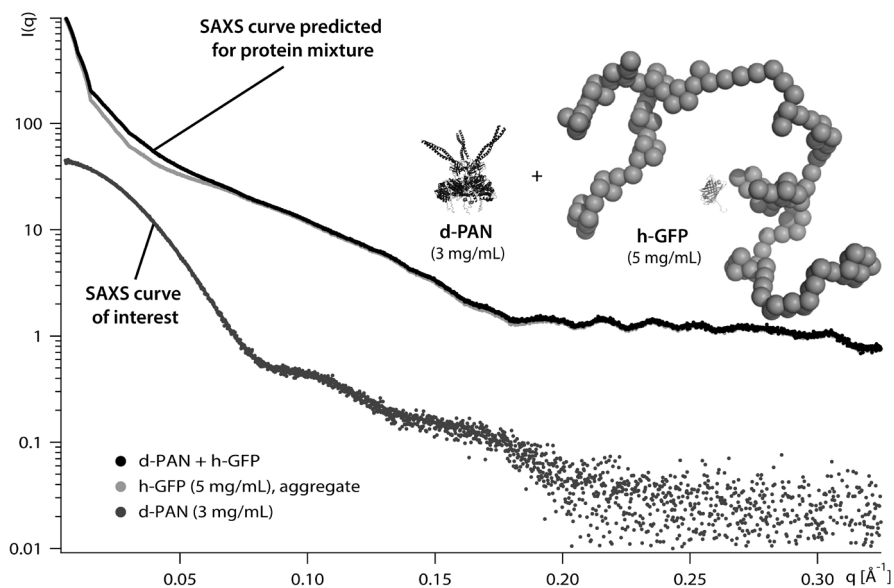
Synthetic SAXS data were calculated for native GFP (8), PAN (PDB ID 6HEA), 20S (PDB ID 6HEA), and a GFP aggregate by using the program *IMSIM* from the ATSAS 3.0.0 suite (12). Theoretical intensities, including errors, were theoretically calculated for 1 second exposure times on a Pilatus 6M detector at the P12 SAXS beamline at PETRAIII for protein concentrations as indicated in the legends of Figures 2-7, corresponding to the ones in the respective SANS experiments.

Experimental SANS data were recorded on D22 at the European Neutron Source Institut Laue-Langevin (Grenoble, France) on selectively perdeuterated (d) GFP or PAN, following sample preparation and deuteration protocols described elsewhere (8).

## 3 Results

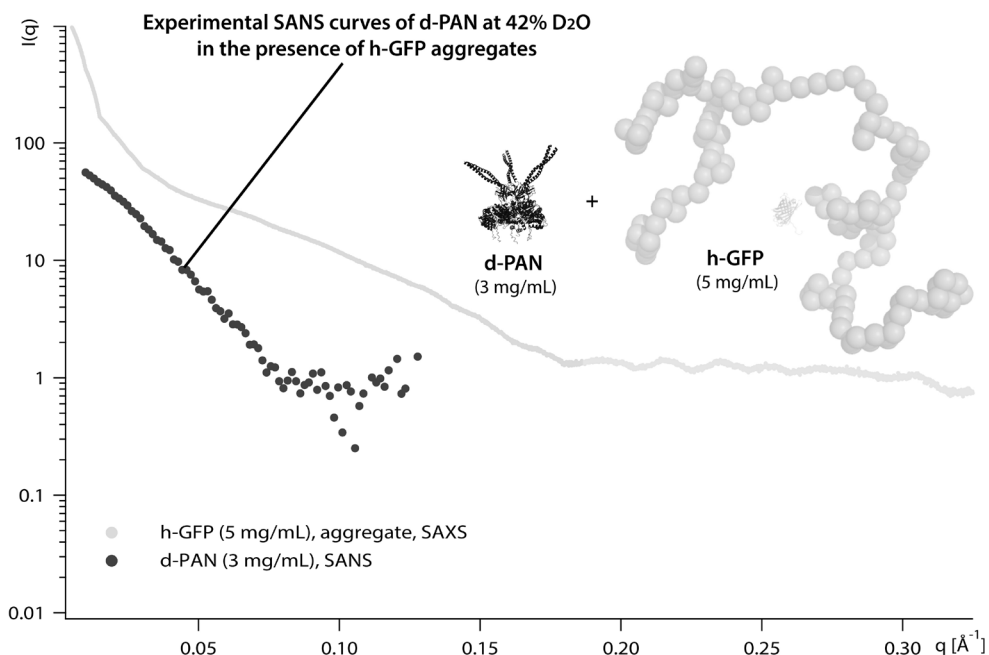
### 3.1 The PAN-GFP system in the absence of 20S

Isolated PAN unfolds *ssrA*-tagged GFP when ATP is supplied (13). In the absence of the proteolytic 20S partner, the released unfolded products instantly form large-scale aggregates in solution (8). Is it possible to extract structural information on PAN and follow potential conformational changes during ATP-hydrolysis in the presence of large-scale GFP aggregates? Figure 2 shows the expected SAXS curves of the individual partners as well as the overall signal, simulated for a 1 second exposure time on a high-performance SAXS synchrotron beamline. The overall signal (black) is completely dominated by the GFP aggregates and cannot be interpreted in terms of PAN structures or conformational changes. Indeed, the  $I(0)$  intensities of the aggregates are more than 30 times stronger (please note the logarithmic scale) than those of PAN for the given concentrations.



**Figure 2.** SAXS curve predicted for the d-PAN + h-GFP reaction mixture. The SAXS curve of the reaction mixture (black) is the sum of both species, and is strongly dominated by the h-GFP aggregates (light grey). The structural information on d-PAN (dark grey) is not accessible.

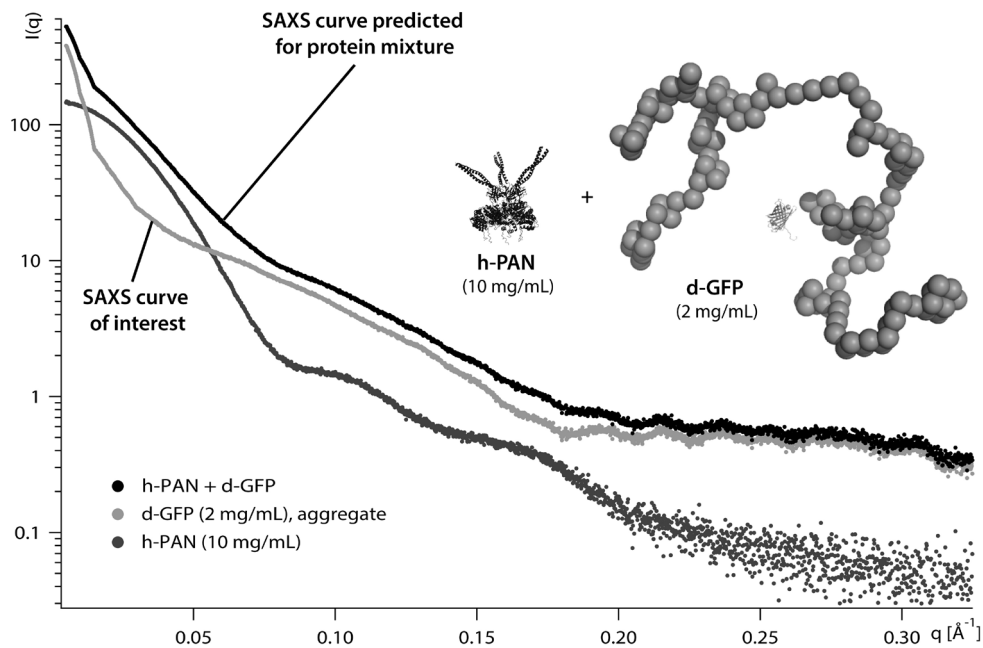
Experimental SANS data (in a 42% D<sub>2</sub>O buffer) from perdeuterated (d) PAN and protonated (h) GFP, recorded at the same concentrations as those imposed for the generation of the synthetic SAXS data (Fig. 2), are shown in Figure 3. Impressively, the signal from the very large GFP aggregates is almost completely masked and the overall signal from the mixture can be attributed almost exclusively to PAN. Indeed, in the previous original study, the change of this signal, over an experimental time of 45 minutes, was interpreted in terms of conformational changes of PAN due to the hydrolysis of ATP and the unfolding of GFP (8).



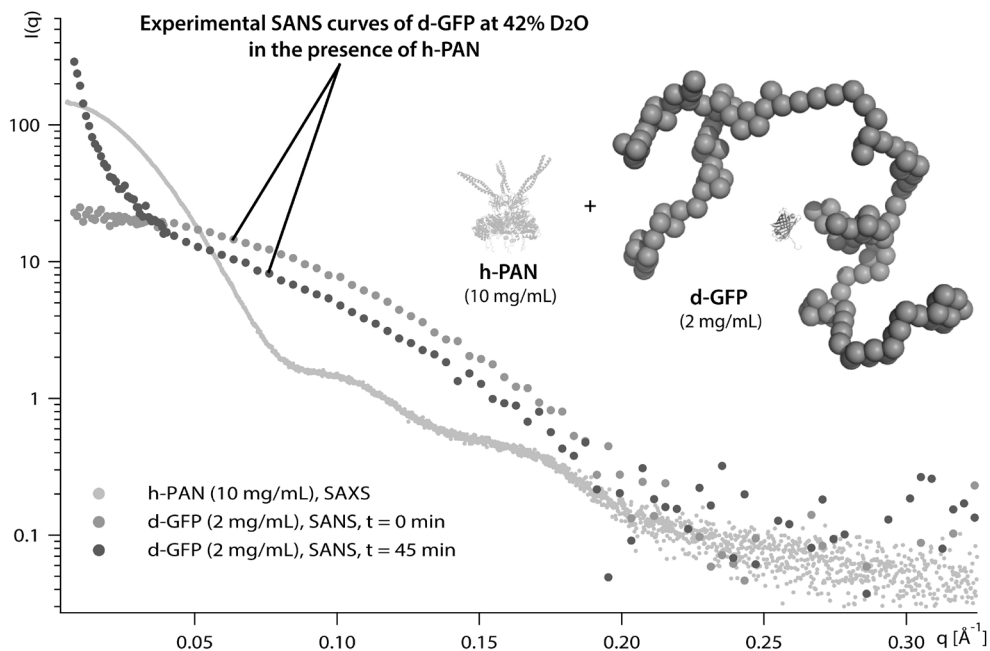
**Figure 3.** Experimental SANS curve of the d-PAN + h-GFP reaction mixture. In 42% D<sub>2</sub>O, the signal of the h-GFP aggregates (transparent grey) is matched and the final SANS curve (dark grey) can be attributed almost exclusively to d-PAN.

Is it possible, on the other hand, to interpret the signal of a mixture of GFP and PAN only in terms of GFP aggregates by neglecting the PAN scattering intensity? Figure 4 shows the expected SAXS curves of such a scenario. While GFP aggregates dominate the overall SAXS signal predicted for the mixture (black data points), there is a non-negligible contribution from PAN, especially at low angles ( $q < 0.07 \text{ \AA}^{-1}$ ) which, if neglected, would lead to an underestimation of the size and extension of the GFP aggregates and SAXS, again, does not yield the true structural values of the aggregates.

If the same reaction mixture and concentrations are measured by SANS in 42% D<sub>2</sub>O on h-PAN and d-GFP on the other hand, the signal from PAN (light grey) is matched out and the remaining signal can be attributed exclusively to the GFP states during the reaction (Fig. 5). It is therefore possible to follow the continuous evolution from native GFP at the beginning of the reaction (grey) until the build-up of the final, large aggregates (dark grey) at the end of neutron exposure after 45 min at 55 °C. Both the small native GFP molecules at the beginning of the reaction as well as the large-scale aggregates towards the end of the reaction are faithfully represented by the overall SANS curves and it is possible to analyse and interpret their evolution as a function of time (8).



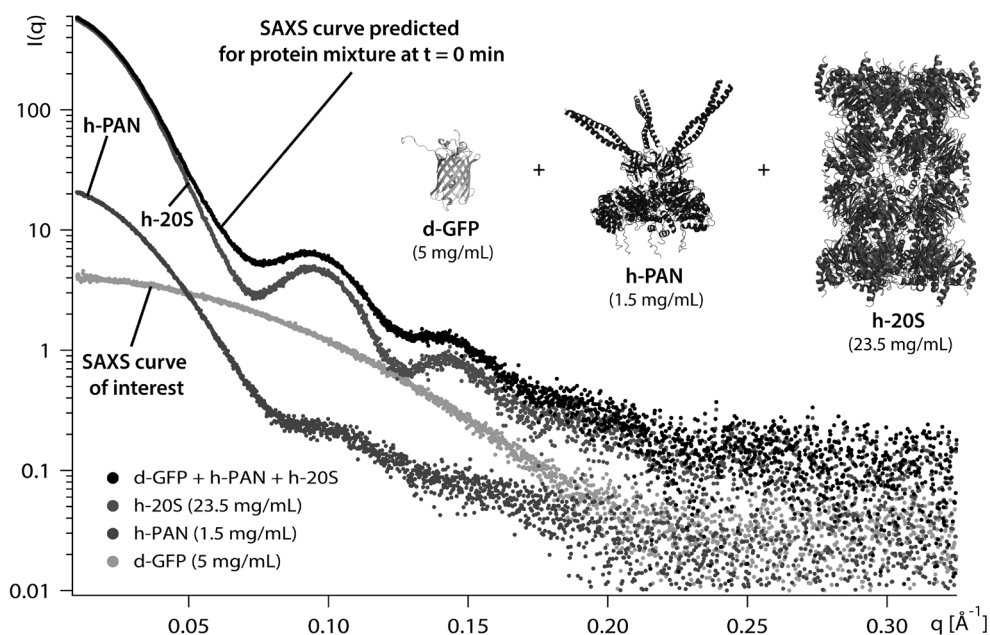
**Figure 4. Predicted SAXS curve of the h-PAN + d-GFP reaction mixture.** The SAXS curve of the reaction mixture (black) is the sum of the curves of both proteins. While it is dominated by the d-GFP aggregates (light grey), it is contaminated by the signal of d-PAN (dark grey) and can therefore not be interpreted exclusively in terms of the d-GFP aggregates.



**Figure 5. Experimental SANS curve of the h-PAN + d-GFP reaction mixture.** In 42%  $D_2O$ , the signal of h-PAN (transparent grey) is matched and the SANS curves (grey and dark grey) can be attributed exclusively to native d-GFP ( $t = 0$  min) and d-GFP aggregates formed ( $t = 45$  min).

### 3.2 The fully functional PAN-20S-GFP system

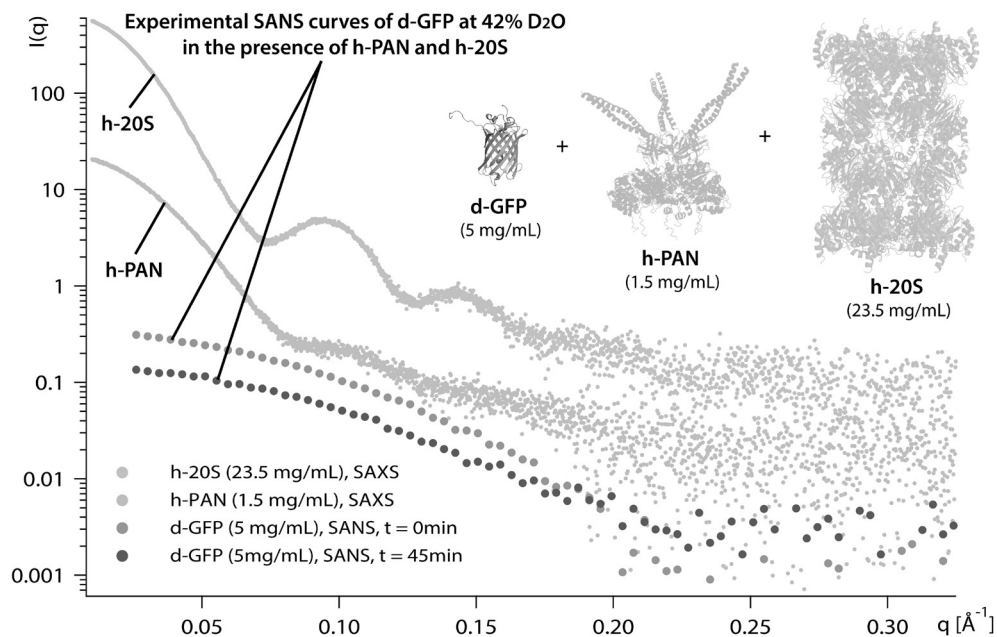
When a PAN solution is prepared in the presence of 20S and supplied with ATP, GFP molecules are efficiently proteolyzed into small oligopeptide products [unpublished data]. In this more complex, ternary system, how are the chances to observe the relatively small (28 kDa) GFP substrate in the presence of the very large oligomeric PAN and 20S particles (298 and 751 kDa, respectively)? Figure 6 shows the expected overall SAXS signal (black) from a mixture of all three partners, as well as the individual SAXS curves simulated for all molecules at the respective experimental SANS concentrations. As can be seen, the overall SAXS signal is largely dominated by, and representative of, 20S. PAN and GFP contributions are negligible in the global SAXS signal. Therefore, it would be very difficult to interpret the measured 1D SAXS curve of the mixture in terms of native GFP structural states or states of similar size.



**Figure 6. Predicted SAXS curve of the d-GFP + h-PAN + h-20S reaction mixture.** The SAXS curve of the reaction mixture (black) is the sum of the curves of all three proteins. It is strongly dominated by the signal of h-20S and it is not possible to extract structural information on d-GFP (light grey).

What would SANS data, measured on the same d-GFP/h-PAN/h-20S system with identical individual concentrations, look like when the hydrogenated, large partners PAN and 20S are masked in 42% D<sub>2</sub>O? Figure 7 shows such experimental SANS data in 42% D<sub>2</sub>O, measured at concentrations equivalent to those of the synthetic SAXS data in Figure 6. Very impressively, the overall experimental SANS curve represents the structure of native GFP molecules, and the signals of the much larger PAN and 20S particles are very efficiently masked. It is therefore possible, by SANS, in combination with selective deuteration and solvent contrast variation, to follow the state of the small GFP protein substrate during its degradation process by PAN and 20S, an achievement which would be completely impossible to accomplish by SAXS.





**Figure 7. Experimental SANS curve of the d-GFP + h-PAN + h-20S reaction mixture.** In 42% D<sub>2</sub>O, the strong signals of h-PAN and h-20S (both transparent grey) are matched out and the SANS curves (grey and dark grey) can be attributed exclusively to d-GFP derivatives (native proteins and oligomeric products).

## 4 General comparison of SAXS and SANS

Both SAXS and SANS allow to obtain structural information (at about the nanometer scale) on solubilized biomacromolecules and the complexes they form. Either technique has its intrinsic strengths and weaknesses:

- 1) The SAXS technique is more readily available and accessible through a broad geographical distribution of SAXS beamlines at synchrotrons worldwide but also due to widely spread benchtop SAXS instruments, fabricated by a number of companies, and available at various research facilities. Sample amounts required at synchrotron beamlines are modest (~5-50  $\mu\text{L}$  at ~mg/ml concentrations) and exposure times are short (~ms to ~s). No labelling (e.g. deuteration) is required for biomolecules, but samples must be monodisperse if a sophisticated structural interpretation in terms of single species is the aim of a study (14). Many biomacromolecules are sensitive to radiation damage by X-rays but protocols have been developed to quantify and deal with this effect (15). Generally, SAXS is a quick, and even high-throughput, method to obtain model-free structural parameters (molecular mass, radius of gyration), but also low-resolution envelopes of isolated biomacromolecules, complexes or mixtures of several species (1). SAXS is in general not very efficient in separating the structural information from individual partners in an assembled complex (protein-protein, protein-RNA/DNA, solubilized membrane protein). However, atomic models can be scored against SAXS data up to medium/high  $q$ -values (~0.5  $\text{\AA}^{-1}$ ) with good efficiency and minor conformational changes can be detected.
- 2) SANS is a “low-throughput” method requiring larger sample amounts (~200  $\mu\text{L}$  at ~mg/mL concentrations), H<sub>2</sub>O-D<sub>2</sub>O buffer exchange, and, often, the specific deuteration of bio(macro)molecules (3). Exposure times are of the order of a minute



up to hours, as a function of contrast, incoherent buffer background, sample concentration and molecular weight. However, neutrons do not induce radiation damage and samples can be modified (buffer exchange, titration of ligands/partners etc) for multiple use in an experimental series. Most importantly, as demonstrated by the data presented here, SANS allows to specifically extract structural information from selected macromolecules in a mixture, even if they are chemically similar (i.e. different proteins). SANS is therefore a technique that requires more investment in sample preparation and yields lower throughput. However, it allows to access and separate structural information of specific isolated macromolecules in a mixture or of partners within an assembled complex, even in the presence of much larger species that would completely dominate the signal if the sample was measured by SAXS.

## Acknowledgements

We acknowledge financial support from the French Agence Nationale de la Recherche (grant “PROTStretch” ANR-15-CE11-0026-01). This work used the platforms of the Grenoble Instruct-ERIC Centre (ISBG: UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INBS-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The authors thank Daniel Franke (EMBL Hamburg) for providing a script for the calculation of synthetic SAXS curves with *IMSIM*. The SANS raw data are available under the following DOIs: <https://doi.ill.fr/10.5291/ILL-DATA.8-03-935> and <https://doi.ill.fr/10.5291/ILL-DATA.8-03-836>.

## References

1. D. I. Svergun, M. H. J. Koch, P. A. Timmins, R. P. May, *Small Angle X-ray and Neutron Scattering from Solutions of Biological Macromolecules*. I. U. o. Crystallography, Ed., IUCr Monographs on Crystallography (Oxford University Press, Oxford, 2013).
2. C. D. Putnam, M. Hammel, G. L. Hura, J. A. Tainer, X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics* **40**, 191-285 (2007).
3. E. Mahieu, F. Gabel, Biological small-angle neutron scattering: recent results and development. *Acta Crystallographica Section D* **74**, 715-726 (2018).
4. F. Gabel, S. Engilberge, J. Pérez, E. Girard, Medical contrast media as possible tools for SAXS contrast variation. *IUCrJ* **6**, 521-525 (2019).
5. C. M. Jeffries *et al.*, Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat Protoc* **11**, 2122-2153 (2016).
6. M. Haertlein *et al.*, in *Methods in Enzymology*, Z. Kelman, Ed. (Academic Press, 2016), vol. 566, pp. 113-157.
7. M. Jasnin, M. Moulin, M. Haertlein, G. Zaccai, M. Tehei, Down to atomic-scale intracellular water dynamics. *EMBO Rep* **9**, 543-547 (2008).
8. Z. Ibrahim *et al.*, Time-resolved neutron scattering provides new insight into protein substrate processing by a AAA+ unfoldase. *Scientific Reports* **7**, 40948 (2017).
9. A. L. Goldberg, Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895-899 (2003).
10. E. U. Weber-Ban, B. G. Reid, A. D. Miranker, A. L. Horwich, Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature* **401**, 90-93 (1999).

11. S. Gottesman, E. Roche, Y. Zhou, R. T. Sauer, The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & development* **12**, 1338-1347 (1998).
12. D. Franke *et al.*, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography* **50**, 1212-1225 (2017).
13. N. Benaroudj, P. Zwickl, E. Seemüller, W. Baumeister, A. L. Goldberg, ATP hydrolysis by the proteasome regulatory complex PAN serves multiple functions in protein degradation. *Molecular Cell* **11**, 69-78 (2003).
14. J. Trewhella *et al.*, 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Crystallographica Section D* **73**, 710-728 (2017).
15. J. C. Brooks-Bartlett *et al.*, Development of tools to automate quantitative analysis of radiation damage in SAXS experiments. *Journal of Synchrotron Radiation* **24**, 63-72 (2017).