

Towards the standardized presentation and publication of small-angle scattering data from biomolecules in solution

Mirjam Czjzek*

Sorbonne Université, CNRS, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), 29680 Roscoff, Bretagne France. *Correspondence e-mail: czjzek@sb-roscoff.fr

Keywords: commentary; small-angle scattering; data presentation; standardized descriptions; templates.

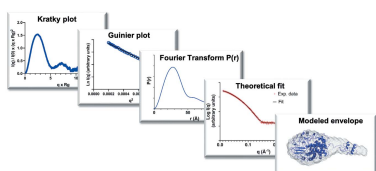
Today Findable Accessible Interoperable Reusable, or FAIR, data principles are on everyone's lips (Wilkinson *et al.*, 2016). But how do you ensure that these principles are easy to fulfil, widespread and adapted to *your* data? It is all about standardizing experiments, data treatment, data reporting and deposition. In this context, the macromolecular crystallography community, through engagement via the International Union of Crystallography (IUCr), has been exemplary and an early pioneer, as shown by the recent celebration of the 50th anniversary of the wwPDB (Worldwide Protein Data Bank) (Burley *et al.*, 2022). Through the decades, and in correlation with the technical and methodological advances in structural methods dedicated to macromolecules such as NMR or cryo-EM, the PDB coordinate repertoire has expanded to include as many as possible experimentally determined three-dimensional (3D) structural data from biological macromolecules (<https://www.rcsb.org/pages/about-us/history>).

In contrast to the above-mentioned structural methods that (potentially) give access to close-to-atomic resolution, small-angle scattering (SAS) profiles, whether obtained by neutron (SANS) or X-ray (SAXS) scattering of macromolecules in solution, only give access to low resolution, and do not allow the determination of atomic coordinates that could be compatible with data deposition at the PDB. Albeit with this limitation, small-angle scattering is complementary to other methods, and is particularly powerful in combination with these other methods, leading to the emerging field of integrative/hybrid structure determination (Sali, 2021). Thanks to considerable advances in instrumentation and computational methods, the amount of SAS data has also increased dramatically over the past two decades (Fig. 1), and the need for standardized data treatment and deposition has become equally urgent and important for SAXS and SANS data. Since the wwPDB initiative that created the SAS validation task force, publication guidelines have been extended to those of SAS data (Trehella *et al.*, 2013; Jacques *et al.*, 2012*a,b*), finally resulting in the Small Angle Scattering Biological Data Bank, SASBDB, (Valentini *et al.*, 2015) that today is connected to the wwPDB through PDB-dev (Vallat *et al.*, 2021).

The article by Trehella *et al.* (2023) in this issue depicts the history of the actions commissioned by the IUCr and the SAS community, leading to best practice and publication guidelines, including template tables being published by Trehella *et al.* in 2017. With the aim of providing an updated template, following the IUCr announcement of requirement of data deposition (including SAXS and SANS data) prior to editorial review (Baker *et al.* 2022), and of keeping pace with the recent methodological and technological advances, Trehella *et al.* (2023) now provide a 2023 update of template tables for reporting biomolecular structural modelling of small-angle scattering data.

Integrating the experience gained by five years of using the template tables developed in 2017, combined with practical experience, the updated tables improve the presentation, while trying not to impose unnecessary work on researchers. While the original templates were mainly based on data from four SAXS examples, the new template tables also better integrate data coming from more complex samples (*i.e.* glycosylated proteins, DNA and RNA) and experiments, such as those from SAS contrast variation experiments (SAS-cv).

The article transparently describes the process by which the new consensus template tables were elaborated, highlighting the major changes and proposing the standardization of nomenclature to describe the components. The article then describes the reasons for



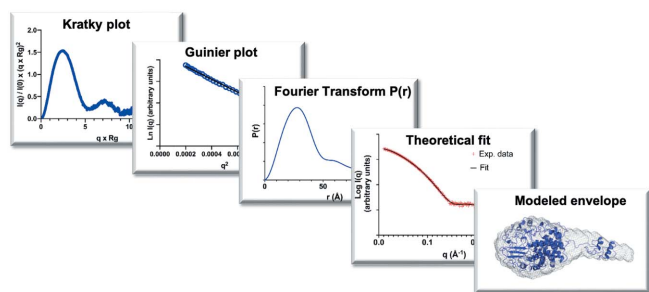


Figure 1
Schematic flow of data representations for SAXS experiments.

reorganization of some of the sample details, or changes in reporting specific data-collection parameters, such as wavelength or beam geometry.

Finally, the newly developed table is tested for its utility using a published complex DNA–protein SAXS study (Pozner *et al.*, 2018), for which all reported items populate the template and are subsequently commented on in the context of how to analyse and interpret the given values.

A second objective of Trehella and coworkers was the creation of a SAS-cv template table, to include the required additional parameters for this type of experiment. Again, Trehella *et al.* (2023) describe which items have been added and why, with the aim of allowing a broadened applicability, without increasing the workload. To test the relative ease of use and utility of the template, it was populated with data and modelling from a combined SAXS/SANS-cv experiment deposited in SASBDB, of a protein complex consisting of a histidine kinase with bound protein inhibitors that were partially deuterated (Whitten *et al.*, 2007). The reader is then guided through the resulting table, indicating how to read and interpret the given data, how the figures can complete the information obtained, and what assessments can be made by a reader/reviewer given the quality of the data presentation. The examples discussed nicely demonstrate the utility of such a standardized presentation, and the template tables will not only facilitate a critical assessment to readers and reviewers, but also largely support the experimenter to prepare, perform and present their SAS data in a concise but complete manner.

As a conclusion, the authors point out that it is evident that a single template table cannot account for all the different kinds of SAS experiments involving biomolecules that can be set up or thought of. However, discussions are already ongoing

in the SAS community to follow the examples presented here and develop templates for SAS experiments involving nanoparticle/micelle/bicelle-type structures (Trehella *et al.*, 2023).

Taken together, these efforts are of great importance to enhance the confidence with which the scientific community can rely on biomolecular SAS data and build up upon them for further investigations. And I sincerely believe that the structural biology community will strongly welcome the template tables and use them extensively for future publications.

References

- Baker, E. N., Bond, C. S., Garman, E. F., Newman, J., Read, R. J. & van Raaij, M. J. (2022). *IUCrJ*, **9**, 1–2.
- Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., Duarte, J. M., Dutta, S., Fayazi, M., Feng, Z., Flatt, J. W., Ganesan, S. J., Goodsell, D. S., Ghosh, S., Kramer Green, R., Guranovic, V., Henry, J., Hudson, B. P., Lawson, C. L., Liang, Y., Lowe, R., Peisach, E., Persikova, I., Piehl, D. W., Rose, Y., Sali, A., Segura, J., Sekharan, M., Shao, C., Vallat, B., Voigt, M., Westbrook, J. D., Whetstone, S., Young, J. Y. & Zardecki, C. (2022). *Protein Sci.* **31**, 187–208.
- Jacques, D. A., Guss, J. M., Svergun, D. I. & Trehella, J. (2012a). *Acta Cryst.* **D68**, 620–626.
- Jacques, D. A., Guss, J. M. & Trehella, J. (2012b). *BMC Struct. Biol.* **12**, 9.
- Pozner, A., Hudson, N. O., Trehella, J., Terooatea, T. W., Miller, S. A. & Buck-Koehntop, B. A. (2018). *J. Mol. Biol.* **430**, 258–271.
- Sali, A. (2021). *From Integr. Struct. Biol. Cell. Biol. J. Biol. Chem.* **296**, 100743.
- Trehella, J., Duff, A. P., Durand, D., Gabel, F., Guss, J. M., Hendrickson, W. A., Hura, G. L., Jacques, D. A., Kirby, N. M., Kwan, A. H., Pérez, J., Pollack, L., Ryan, T. M., Sali, A., Schneidman-Duhovny, D., Schwede, T., Svergun, D. I., Sugiyama, M., Tainer, J. A., Vachette, P., Westbrook, J. & Whitten, A. E. (2017). *Acta Cryst.* **D73**, 710–728.
- Trehella, J., Hendrickson, W. A., Kleywegt, G. J., Sali, A., Sato, M., Schwede, T., Svergun, D. I., Tainer, J. A., Westbrook, J. & Berman, H. M. (2013). *Structure*, **21**, 875–881.
- Trehella, J., Jeffries, C. M. & Whitten, A. E. (2023). *Acta Cryst.* **D79**, 122–132.
- Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M. & Svergun, D. I. (2015). *Nucleic Acids Res.* **43**, D357–D363.
- Vallat, B., Webb, B., Fayazi, M., Voinea, S., Tangmunarunkit, H., Ganesan, S. J., Lawson, C. L., Westbrook, J. D., Kesselman, C., Sali, A. & Berman, H. M. (2021). *Acta Cryst.* **77**, 1486–1496.
- Whitten, A. E., Jacques, D. A., Hammouda, B., Hanley, T., King, G. F., Guss, J. M., Trehella, J. & Langley, D. B. (2007). *J. Mol. Biol.* **368**, 407–420.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. *et al.* (2016). *Sci. Data*, **3**, 160018.