ARPHA Preprints

PENSOFT

Grant Proposal

Author-formatted document posted on 05/12/2023

Published in a RIO article collection by decision of the collection editors.

DOI: https://doi.org/10.3897/arphapreprints.e116670

Following biological processes combining small angle neutron and x-ray scattering and modelling techniques

Miguel Angel Gonzalez, Irina Safiulina, Arianna D'Angelo, Paolo Mutti, Jordi Bodera, Gianluca Santoni, Rudolf Dimper, Jayesh Wagh, Patrick Fuhrmann, Paul Millar, Krisztian Pozsa, Leonardo Sala, Alun Ashton, Giuseppe La Rocca

ARPHA Preprints *Author-formatted document posted on 05/12/2023*. DOI: https://doi.org/10.3897/arphapreprints.e116670

Following biological processes combining small angle neutron and x-ray scattering and modelling techniques.

Authors: Gonzalez, M., Safiulina, I., D'Angelo,A., Mutti, P., Bodera,J., Santoni, G., Dimper,R., Wagh, J., Fuhrmann, P., Millar, P., Pozsa, K., Sala, L., Ashton, A., La Rocca, G.

Abstract

Small-angle scattering techniques are widely used in scientific communities to determine the shape, distribution, and uniformity of particles in solution. New developments and faster acquisition will also allow for tracking the dynamics of the particles themselves. Small-Angle X-ray or Neutron Scattering (SAXS or SANS, respectively) can be very effective tools for studying, for example, the time dependence of genome release from phages, investigating entire viral life cycles, or the assembly of macromolecular complexes, providing deep insights into infection pathways. Neutrons and X-rays can be applied in a complementary mode, as is the case for the joint SANS-SAXS proposal between the ESRF and the ILL.

This strategic plan aims to advance the field by providing an EOSC-based platform, enabling FAIR data and software, unified data processing pipelines featuring robust scaling algorithms for the two different sources, supporting reproducibility and automated validation, and integrating with other relevant structural databases (e.g., electron microscopy/tomography or protein structural and ligand databases).

Keywords

EOSC Future; Science Clusters; Science Projects; PaNOSC; Small angle neutron scattering (SANS); Small angle x-ray scattering (SAXS)

Description

Existing situation

The Photon and Neutron (PaN) user facilities serve a large variety of scientific user communities using x-rays and/or neutrons in their research projects to understand the structure and function of matter. Several beamlines available in these facilities specialize in determining the structure of complex macromolecules, such as ribosomes, protein-ligand complexes, enzymes, etc. However, the classical approach based on crystallography requires the existence of crystals or microcrystals of the corresponding system. An alternative approach involves measuring the macromolecule in an aqueous or buffer solution using small angle scattering techniques, either with neutrons (SANS) or with x-rays (SAXS). The two main advantages of this approach are that it can be applied to systems where obtaining single crystals is impossible or extremely difficult, and that biomolecules can be measured under conditions much more similar to physiological media. Therefore, molecular motion is not hindered by the crystal structure, and all possible conformations available to the biological complex contribute to the measured signal. On the other hand, the

trade-off for the lack of order is that standard crystal structure solving techniques cannot be applied, and data analysis must rely heavily on modeling the possible conformations. This approach provides a final set of the most likely structures and conformations that the macromolecule can adopt, along with their relative frequencies.

As the analysis and modeling approaches are not unique and results can vary depending on initial predictions, the variety of models tested, and available computational resources, it has been recognized by the community the need to keep a data base of instrument-corrected SANS and SAXS results. This allows other groups to validate published results, and test new approaches and algorithms. As a result, a small angle scattering biological data bank (SASBDB) has been created and is open for public access (Kikhney et al. 2019, SASBDB 2023). In combination with the structural models available in another public data base, the well known RCSB Protein Data Bank (PDB 2023, Berman 2000), both repositories offer a powerful and comprehensive resource for researchers in the field of structural biology.

Objectives

This SP will implement a pipeline hat combines SASBDB and PDB data, providing a powerful framework for elucidating the structural dynamics and functional mechanisms of complex biological systems. While various tools already exist, their usage typically requires very specific domain knowledge and, furthermore, quite often different software is used for SANS and SAXS data. Therefore our main goal is to create a simple workflow that enables non-expert users to examine and work with neutron and x-ray small angle scattering data of biological data, collected at neutron and synchrotron facilities. This workflow will provide a simple interface along with the necessary computational resources.

As only reduced data are accessible in SASBDB, another goal is to make the data portals of the neutron and X-ray facilities involved accessible through the EOSC portal. This will grant EOSC users access to the raw data of many of the final data sets published in SASBDB, allowing them to retreat the original data if needed.

Compliance to criteria developed by EOSC Future

This SP is part of the contribution of the PaNOSC cluster to the EOSC platform.

Eligibility

This SP has a cross-disciplinary character and involves two large scale facilities providingEU scientists with access to numerous x-ray and neutron beamlines. The proposed SP will serve as a practical demonstration of the tools and organization developed during the previous PaNOSC/ExPaNDS projects.

Contribution to EOSC

This SP will facilitate the re-use of open SANS and SAXS data already published in freely available databases. The necessary services (data analysis notebooks, virtual computational resources, etc.) will be onboarded into EOSC and made available to the entiry community through the EOSC AAI federation.

Quality

The facilities involved are committed to complying with FAIR principles (Wilkinson et al. 2016). High-quality data and tools will be provided, ensuring that expert domain knowledge is accessible for multidisciplinary research.

Relevance

SANS and SAXS are highly complementary techniques. Open SAS data are already provided to the entire research community through the SASBDB effort (SASBDB 2023), while the US RCSB PDB offers a vast database containing atomistic description of more than 200,000 biological structures. This SP will facilitate the combined use of these resources, enhancing the utility of structural biology research by offering a holistic approach to studying macromolecular structures. This integration empowers researchers to explore both the solution-state behavior of macromolecules and the precise atomic details of their structures, enabling deeper insights into the fundamental mechanisms of life and facilitating the development of novel therapeutics and biotechnological applications.

Implementation, Plan of work

Tasks

- Identify existing software and components to be used in the pipeline.
- Umbrella AAI federated with EOSC AAIs.
- Extend the software to handle both SANS and SAXS data.
- Identify suitable data sets to benchmark the proposed workflow.
- Jupyter notebook and PaN data analysis portal integrated with EOSC AAI.
- PaN learning platform with interactive content.
- Jupyter notebook direct integration into EOSC portal.

Use of resources

Initially, the focus will be on leveraging internal resources, taking advantage of the newly established infrastructure within the facilities to allow remote experiments and data analysis. Additionally, internal scientific expertise will be used to define the necessary pipelines for implementation. Technical support from EGI and EOSC partners will be sought to assist with on boarding local services and integrating any potential web services that could be developed during the project.

Partners

Institut Laue-Langevin (ILL); European Synchrotron Radiation Facility (ESRF).

Impact

Strategic

This SP aims to enhance accessibility of open Small-Angle Neutron Scattering (SANS)/Small-Angle X-ray Scattering (SAXS) data to a broader community, highlighting the transformative impact of open-science data analysis within the EOSC framework and promoting the application of FAIR principles. By simplifying and integrating highly specialized modeling tools into practical pipelines, this SP seeks to make them accessible to non-experts, thereby reducing the domain-specific knowledge required to harness the valuable data sets stored in SASBDB. Through the provision of user-friendly, readily available, and state-of-the-art data processing pipelines, the proposed SP will facilitate cross-disciplinary collaboration and foster innovative advancements, thereby accelerating the adoption of emerging technologies. This approach aligns with FAIR-enabling services, many of which are available through the EOSC marketplace, ultimately fostering greater engagement among researchers in the EOSC implementation.

Scientific/User communities

This SP aims to drive advancements in the field by establishing an EOSC-based platform that facilitates the principles of FAIR data and software. The platform will incorporate unified data processing pipelines with robust scaling algorithms, valid for neutron and synchrotron sources, while prioritizing reproducibility and automated validation.

Additionally, integration with other relevant structural databases (e.g., electron microscopy/ tomography or protein structural and ligand databases) will be pursued. Naturally, the primary scientific community directly impacted by this SP comprises structural biologists already engaged in SAS techniques. As a result of this SP, they will be able to test models and ideas against available experimental data in a much more straightforward manner. However, the benefits of this SP extend to numerous other research domains as well, as the reduced barrier to access and use SAS data will significantly enhance the reuse of SANS/SAXS

data for modelling purposes across multiple fields. For example, the information contained in SASBDB could be routinely employed to validate force fields and models used in computer simulations.

Societal/Economic

One of the key objectives of this SP is to demonstrate the potential impact of SAS in the bio-medical and bio-chemical characterization of macromolecules and drugs. By highlighting its ability to expedite the discovery of novel drugs or healthcare treatments, this SP aims to underscore the valuable role that SAS can play in advancing these fields.

EU policies

The SP builds on the EU commitment to foster the advancement of neutron and synchrotron research, encouraging the complementarity between both techniques and the collaboration among member states and international partners, and establishing frameworks for access to these cutting-edge facilities by researchers from across Europe.

Engagement plan

Target groups

Users of photons and neutrons facilities, structural biology community in Europe, students and teachers for experimental data analysis and method developers. Pharmaceutical and biotechnology industries.

Dissemination measures

Scientific publications and reports. Conferences, workshops and user meetings. Training programs (e.g. through the PaN-learning platform PaN-learning 2023).

Acknowledgements

PaNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852. ExPaNDS has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641. The EOSC Future project is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020 - Grant Agreement Number 101017536.

Funding program

EOSC Future project, WP 6.3, co-funded by the EU Horizon Programme call INFRAEOSC-03-2020 – Grant Agreement Number 101017536.

Conflicts of interest

The authors have declared that no competing interests exist.

References

[Berman HM (2000] The Protein Data Bank. Nucleic Acids Research 28 (1): 235-242. https://doi.org/10.1093/nar/28.1.235.

[Kikhney A, Borges C, Molodenskiy D, Jeffries C, Svergun D (2019)] SASBDB: Towards an automatically curated and validated repository for biological scattering data. Protein Science 29 (1): 66-75. https://doi.org/10.1002/pro.3731.

[PaN-learning (2023)] Photon and Neutron eLearning. https://pan-learning.org/. Accessed on: 2023-6-30.

[PDB (2023)] RCSB PDB: Protein Data Bank. https://www.rcsb.org/ . Accessed on: 2023-6-29.

[SASBDB (2023)] SASBDB: Small Angle Scattering Biological Data Bank. https://www.sasbdb.org/ . Accessed on: 2023-6-29.

[Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016)] The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1). https://doi.org/10.1038/sdata.2016.18.