Biomolecular small-angle scattering: data reproducibility, benchmarking predictive methods, and best practice reporting

Jill Trewhella

School of Life & Environmental Sciences, U. Sydney, NSW, 2006, Australia, and Dept. of Chemistry, U. Utah, UT 84112, U.S.A.

jill.trewhella@sydney.edu.au

Keywords: small angle scattering, SAXS, SANS, structural biology

In 2019, a round robin study was initiated aimed at providing a quantitative assessment of the reproducibility of biomolecular SAS data and consensus profiles. Five proteins (RNaseA, lysozyme, xylanase, urate oxidase and xylose isomerase) were measured on twelve Small-Angle X-ray Scattering (SAXS) and four Small-Angle Neutron Scattering (SANS) instruments [1]. From these data, the solvent-subtracted protein scattering profiles were shown to be reproducible, with the caveat that an additive constant adjustment was required to account for small errors in solvent subtraction. Further, the major features of the obtained consensus SAXS data over the q-measurement range 0 – 1 Å⁻¹ are consistent with theoretical prediction, but with residual differences that can now be accounted for. The inherently lower statistical precision for SANS limited the reliably measured q-range to < 0.5 Å⁻¹, but within the limits of experimental uncertainty the major features of the consensus SANS data are also consistent with prediction for all five proteins measured in H₂O and in D₂O. Thus, a foundation set of consensus SAS profiles has been obtained for benchmarking scattering profile prediction from atomic coordinates. Additionally, two sets of SAS data measured at different facilities to q > 2.2 Å⁻¹ showed good mutual agreement, affirming that this region has interpretable features for structural modelling. SAS measurements with in-line SEC proved generally superior for eliminating sample heterogeneity, but with unavoidable sample dilution during column elution, while batch SAS data collected at higher concentrations and for longer times provided superior statistical precision. Careful merging of data measured using in-line SEC- and batch-modes, or low- and high-concentration data from batch measurements, was successful in eliminating small amounts of aggregate or interparticle interference from the SAS data while providing improved statistical precision overall for the benchmarking data set.

Data for the round robin study were assessed and analyzed using all the criteria and tools described in the 2017 publication guidelines for reporting biomolecular SAS data and 3D modelling [2] that were established through a consultive process spanning more than a decade and a half [3]. In 2022, the IUCr journal editors mandated deposition of SAS data used in biomolecular structure solution to a public archive [4], and notes for authors specify adherence to the 2017 reporting guidelines. The accompanying standard table templates therefore were reviewed and updated to include standard descriptions for proteins, glycosylated proteins, DNA and RNA, with some reorganization of data to improve readability and interpretation. A specialized template was also developed for reporting SAS-contrast variation (SAS-cv) data and models that includes the additional reporting required for these more complex experiments (Trewhella, Whitten & Jeffries accepted for publication Acta Cryst. D, Dec 2022). This framework of best practice reporting standards for the field has helped to position biomolecular SAS to achieve its full potential at the frontier of integrative structural biology, as evidenced by the fact that the prototype archiving system for structural models obtained using integrative modeling, PDB-Dev [5], includes validation reports for contributing SAS data with links to the SAS Biological Data Bank (SASBDB), which now has > 3000 experimental SAS profiles and > 4000 associated models [6,7].

This presentation will focus on the findings of the round robin study and work done since their publication utilizing the consensus profiles and briefly review the rationale for the updates to the template tables for biomolecular SAS and 3D modelling.