Australian Government ANSTO

THE UNIVERSITY OF SYDNEY

Australia's contribution to an International Project to Generate a Consensus Standard Set of SAS Data to Benchmark Methods for SAS profile Prediction

Anthony Duff[1], Nigel Kirby[2], Tim Ryan[1], Jill Trewhella[3], Andrew Whitten[1], Katy Wood[1]

[1] Centre for Neutron Scattering, Australian Nuclear Science and Technology Organisation, Lucas Heights, NSW

[2] The Australian Synchrotron, Australian Nuclear Science and Technology Organisation, Clayton, Victoria

[3] School of Life and Environmental Sciences, The University of Sydney, NSW.

Introduction

This international project is aimed at generating a consensus set of SAS data sets from well-characterized biomolecules that can be used to benchmark different approaches to predicting SAS profiles from atomic coordinates (full project description and participants can be found at https://sas.wwpdb.org/?q=node/25). This project emerged from the deliberations of the International Union of Crystallography Commission (IUCr) on Small Angle Scattering (CSAS) and the Small-Angle Scattering validation task force (SASvtf) of the world-wide Protein Data Bank (wwPDB) that led to the 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution (https://scripts.iucr.org/cgi-bin/paper?jc5010).

Having established a consensus set of publication guidelines for biomolecular SAS and 3D modelling, the SASvtf considered what is an ongoing matter for research in regard to predicting SAS profiles from atomic coordinates. The program CRYSOL (Svergun (1995) *J. Appl. Cryst.* **28**:768), as the first method to rapidly calculate SAXS profiles from atomic coordinates of biomolecules, was a major breakthrough in that it provided the missing link between high-resolution structures and solution SAS data. The importance of this breakthrough is evident in the fact that decades later, alternate methods continue to be published with various claims of improvement.

To date, each of the alternate methods published have been validated using different data sets and models. A consensus set of high quality data agreed would be of considerable value in benchmarking the different approaches. In order to evaluate different approaches to including the hydration layer contribution to the SAS profile, it is desirable to have data obtained using X-rays (SAXS) and neutrons (SANS), the latter in H_2O and D_2O, as the hydration layer contribution differs significantly for each of these measurements.

The current project involves the efforts of 37 researchers with participants from 11 X-ray and 3 neutron scattering facilities across Asia, Europe and North America (**see table 1**). To date 7 SAXS and 3 SANS data sets have been submitted for preliminary evaluation.

Table 1: Participating Facilities	
SAXS	SANS
Advanced Light Source/SIBYLS, Berkeley, CA, USA	ANSTO/QUOKKA
Advanced Photon Source/BioCAT, Chicago, IL, USA	ILL/D22
Australian Synchrotron/SAXS-WAXS, Clayton, Vic. Australia	NIST/30-m SANS
Diamond/B21, Didcot, Oxfordshire, UK	
Macromolecular X-ray science at the Cornell High Energy Synchrotron Source/BioSAXS, Ithaca, NY, USA	
National Institute of Standards Technology/U. Maryland-Institute for Bioscience and Biotechnology Research SAXS	
PETRA III-DESY/P12, Hamburg, Germany	
Shanghai Synchrotron/ BL19U2, Shanghai, China	
Soleil/SWING, Gif-sur-Yvette Cedex, France	
Spring 8-RIKEN/BL40B2, Japan	
Stanford Synchrotron Radiation Laboratory/ BL4-2, Menlo Park, CA, USA	

Project Aim

Generate a set of SAS data that can be used to benchmark different approaches to predicting SAS profiles from atomic coordinates.

Specific Objectives

Measure up to 5 proteins with known structures at different SAS beam lines using a common batch for each protein and appropriate standard buffer(s).

Compare data sets for consistency.

Agree upon a set of data to be made available to the research community to benchmark methods for predicting scattering profiles, along with the measurement protocol and sources of material via a publicly accessible Website(s).

Published a consensus white paper on the results of the exercise.

Criteria for Selected Proteins (see Table 2)

- Relatively rigid structures, *i.e.* no complications due to flexible regions and structural inhomogeneity.
- High resolution crystal structures available, also good if NMR solution structures available.
- Readily available in high purity.
- Known buffer conditions for optimal SAS data collection (aggregation free, no interparticle interference).
- Cover a range of sizes.

Table 2: Selected Proteins		
	Source	"Best" available structures
RNaseA Bovine MW 13690 (monomer) pI 8.64	Sigma Aldrich R6513	7RSA 1.26 Å resolution (X-ray) 5RSA 2.0 Å resolution (neutron)
Lysozyme (Hen egg white) Gallus gallus MW 14313 (monomer) pI 9.32	Sigma L4919 or L6876	2VB1, 0.65 Å resolution
Xylanase Tricoderma reesei MW 20844 (monomer) pI 8.14	Hampton Research HR7-104	2DFC, 1.19Å resolution
Urate oxidase Aspergillus flavus MW 137215 (tetramer w. inhibitor) pI 7.16	Available from Sanofi (Thierry Prangé)	3L8W, 1.0 Å resolution w. xanthin
Glucose Isomerase Streptomyces rubiginosus MW 172909 (tetramer, w/o Mn/Mg) pI 5.0	Hampton Research HR7-102	1MNZ, 0.99 Å resolution

Progress with Measurements

The five selected proteins, with standard buffers, were distributed to all participants in June 2019. Seven SAXS and three SANS data sets have been submitted for preliminary evaluation to identify potential outliers. Due to the logistical challenges with beamline schedules and geography, some samples degraded between shipment and measurement. Nevertheless, there is good agreement among multiple facilities for each protein. Overall, the availability of in-line size exclusion chromatography (SEC) has proven to be highly valuable in eliminating potential aggregates from some samples where delays in shipment or schedules for measurement were interrupted.

Results to Date from ANSTO

Table 3 provides the Guinier R_g values determined from SAXS and SANS measurements from ANSTO facilities compared to those expected based on the high-resolution crystal structures calculated using the program CRYSOL.

Table 3: Theoretical and Experimental R_g Values from ANSTO Measurements					
Protein	CRYSO(L/N) R_g (Å): SAXS, SANS H_2O, SANS D_2O	SEC-SAXS R_g (Å) mg/mL	Batch-SAXS R_g (Å) mg/mL	SANS H_2O R_g (Å) mg/mL	SANS D_2O R_g (Å) mg/mL
RNase A	15.68, 14.82, 14.04				
		15.09±0.05			
			15.37±0.06 1		
			15.35±0.09 2		
			15.40±0.08 4[a]	14.38±0.76 3.9	13.87±0.06 4
			15.52±0.05 8[a]	14.77±0.21 7.7	14.82±0.11 8.1
Lysozyme	14.94, 13.97, 11.48				
		14.28±0.20 [d]			
			14.30±0.21 0.75[b]		
			14.50±0.14 1.5[b]		
			14.56±0.06 3[b]		
			14.98±0.07 6[b]	14.71±0.74 4.1	13.68±0.23 5
				15.25±0.36 8.2	14.69±0.18 9
Xylanase[c]	16.39, 15.15, 13.92				
		15.98±0.11			
			severe aggregate	16.44±0.38 5.4	15.30±0.06 5.2
				17.20±0.20 10.6	16.49±0.30 10.3
UOX	31.55, 30.74, 29.82				
		32.26±0.10			
			not measured	32.79±2.19 1.7	31.15±0.161.8
				32.48±0.59 3.4	31.39±0.11 3.6
GI	32.66, 31.79, 30.43				
		33.10±0.24			
			32.79±0.14 0.9[e]	32.07±2.11 1.0	31.01±0.70, 1.0
				32.14±0.48 1.9	30.92±0.13, 2.0

[a] aggregation, analysis removed first 54,58,56,32 data points resp. with increasing concentration.
[b] interparticle interference, removed first 61, 45, 45, 44 data points resp. with increasing concentration.
[c] all batch-SAXS measurements of Xylanase showed significant dimer/aggregation.
[d] interparticle interference evident, analysis removed first 64 data points.
[e] all higher concentrations showed interparticle interference prohibiting analysis for this sized protein.

Conclusions

Preliminary analysis of the SAXS and SANS data from ANSTO provides the following:

- Measured R_g values are in the ball park of expectations based on those obtained from scattering profiles predicted from crystal structure coordinates using CRYSOL (SAXS) and CRYSON (SANS). Alternate methods for scattering profile prediction are yet to be assessed and compared.
- The trends observed for R_g progressing from SAXS, H_2O SANS, D_2O SANS measurements are generally as expected with some exceptions for proteins that had issues with aggregation or interparticle interference.
- Lysozyme at low pH values can be susceptible to interparticle interference, and this is evident for the SAXS data but not for SANS, suggesting some unanticipated difference in solvents.
- Xylanase suffers from dimerization and possibly higher order aggregates, and these effects are most evident in the batch-SAXS data. Concentration, time-dependent, and radiation effects may be impacting these results.
- RNase A is the most promising of the small proteins to be a good standard, which is consistent with data sets submitted from other facilities.
- The larger proteins (urate oxidase (UOX)) and glucose isomerase (GI) have proven to be significantly better standards than the smaller proteins.
- GI is the most robust of the proteins tested here, but is no longer available from Hampton.
- UOX is less stable and has limited commercial availability.