Project Plan: Preparation of Benchmark SAS Data Set using a Round Robin Approach to Developing a Consensus Standard Set

Aim

Generate a set of SAS data sets 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 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.

Analysis and Publication

- Project coordinator Jill Trewhella will receive, initially in confidence, individual data sets as $I(q)$ vs q with $I(q)$ on an absolute scale (cm^{-1}) and associated errors for [Sample+Solvent], [Solvent], and [Sample+Solvent] - [Solvent] and perform an initial comparison.
- This initial comparison would include direct comparison of scaled $I(q)$ vs q, Guinier R_g, $P(r)$ profiles with R_g, Porod volume and relative $I(0)$ compared to expected based on known molecular weight within each set from a given instrument.
- Any outliers identified in the above comparison will be discussed in confidence with the contributor to provide the opportunity to address any issues that might be revealed.
- With the agreement of each contributor, their data will be part of a set that will be made available to all project participants to develop a set of consensus findings.
- Individual participants will decide if they want to contribute their data to the benchmarking set or subset; the aim is to include data over the widest q-range possible with accurate error propagation, hence minimal data manipulation, on all four proteins.
- Data in the benchmarking set or subset will be deposited in a public archive with complete information on their source and measurement protocol and may include some SAXS-WAXS profiles developed by combining different measurements. In this latter case, the contributing measurements to generate the SAXS-WAXS profile will also be deposited.
- The consensus findings of the exercise will be published as white paper, all data contributing will be fully documented per the 2017 publication guidelines http://scripts.iucr.org/cgi-bin/paper?S2059798317011597.

Selection of Proteins

Criteria:

- 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.

Final measurement set and suppliers:

- From Tim Ryan and Anthony Duff, Australian Synchrotron and National Deuteration Facility, ANSTO
  - RNase A, Bovine, Bos taurus, Sigma Aldrich R6513
  - Xylanase, Tricoderma reesei, Hampton Research HR7-104
  - Lysozyme, chicken egg white, Gallus gallus, Sigma Aldrich L6876.
  - Glucose isomerase (a.k.a Xylose isomerase) tetramer, Streptomyces rubiginosus, Hampton Research HR7-102
- Optional additional protein
  - Urate oxidase tetramer, Aspergillus flavus, from Thierry Prangé (Université René Descartes, Paris)

(**Table 5** appended below provides some of the detail leading to the selection of the above 5 and references to high-quality crystal structures.)

Protein and Buffer Supply

Proteins are to be supplied to each participating lab from a single batch source with standardized buffers. **Note**: **xylanase, glucose isomerase and urate oxidase cannot be frozen** and it is also not advised to freeze RNase A or lysozyme once they are in solution. Upon receipt, proteins and buffers should be stored at 4°C. **Table 1** provides detail for each protein regarding pI values, A280 extinction coefficients, amounts, buffers, temperature and desired concentrations for SAS/WAS measurement.

Xylanase, glucose isomerase, RNase A and lysozyme

- Due the age of the preparations, xylanase and glucose isomerase were checked by mass spectrometry and gel electrophoresis to confirm they have the expected molecular weight and high purity.
- Each protein and buffer prepared using the recommended protocol has been checked by SEC-SAXS at the Australian Synchrotron prior to shipment and the scattering profile/parameters are as expected for their folded, intact forms.
- Approx. **15 mg amounts**, on ice, are being sent from Australia by special courier to each participating lab accompanied by 10x buffer solutions for SAS measurement.
- The **xylanase** will be supplied as a 43% glycerol stock, 0.5 mL of 36 mg/mL protein.
  - Immediately prior to SAS sample preparation and measurement, the xylanase stock needs to be dialyzed with 3 x 2 hour changes and 1:50 volume ratio, against homemade buffer of 50 mM Tris pH 7.5, 150 mM NaCl with progressively decreasing glycerol: 20%, 10%, and 0% (w/v). Finally, dialyze 2 x 6 hour, 1:100 volume ratio against the measurement buffer provided together with the protein. This will ensure sufficient removal of glycerol. Preserve sufficient last step dialysate for batch measurements.
- The **glucose isomerase** as an ammonium sulphate precipitate (120 mg/mL or 0.91 M Ammonium sulphate), 0.5 mL of 33 mg/mL protein.

- Immediately prior to SAS sample preparation and measurement, the glucose isomerase stock will need to be dialyzed with 3 x 2 hour changes, 1:50 volume ratio, against homemade buffer of 50 mM Tris pH 7.5 with progressively decreasing salt: 500 mM NaCl, 250 mM NaCl, 150 mM NaCl. Finally, dialyze 2 x 2 hour, 1:100 against the measurement buffer provided together with the protein. Preserve sufficient last step dialysate for batch measurements.

- **RNase A** and **lysozyme** will be powders (as supplied by Sigma).
  - Immediately prior to SAS sample preparation and measurement dissolve RNase A and lysozyme directly into the buffer provided for SAS measurements (per some labs you may spin samples in an Eppendorf centrifuge (or equivalent) for 5 mins to remove dust/particles, make a note in the reporting if you perform any such "standard practices" that you regard as critical for your set up) and dialyze against the measurement buffer with 2 x 2 hour changes, 1:100 volume ratio. Preserve sufficient last step dialysate for batch measurements. (Note the exception for lysozyme in D_2O buffer detailed below in special circumstances for SANS samples in D_2O)
  - Lysozyme Sigma Aldrich L6876 has been found to give an R_g value that is too high when measured in batch mode, though it appears to be correct when measured in SEC-SAXS mode. Sigma Aldrich L4919 was the preferred lysozyme, but delivery was delayed. Suggest to either run the supplied lysozyme through a SEC column prior to batch SAXS measurement and/or if you have locally available any Sigma Aldrich L4919, use this material and make relevant notes in the reporting.

Buffers for xylanase, glucose isomerase, RNase A and lysozyme

- For all SAS measurements, 10x buffer solutions will be provided, having been prepared in autoclaved bottles with filtering (0.22µm filter) and transferred to sterile 50 mL falcon tubes for transport as follows:
  - For SAXS measurements, a set of ~40 mL aliquots of 10x buffer solutions (one for each protein sample) for each SAXS lab where they will need to be diluted 1:10 using 18 MOhm water for SAXS.
  - For SANS measurements one set each of ~40 mL H_2O and ~40 mL D_2O 10x buffer solutions will be provided to be diluted 1:10 with 18 MOhm water and Sigma D_2O, respectively.
  - Each diluted buffer should be checked for pH and adjusted to desired values (per Table 1) if needed.
  - The free radical scavenger NaN_3 (0.1% w/v, recommend ReagentPlus, 99.5%, Cat # S2002, from Sigma-Aldrich) will need to be added to buffers just prior to SAS sample preparation (due to its time-dependent degradation).
- **Special circumstances for SANS samples in D_2O**:
  - Lysozyme will not dissolve in D_2O solutions, rather it forms an insoluble gel. Lysozyme must be first dissolved in H_2O and then can be dialyzed directly into D_2O or subjected to exchange on a column.
  - For ILL/D22, the exchange of H_2O buffer to D_2O buffer will be done by SEC, as describe below for the urate oxidase sample.
  - pH values measured in D_2O and H_2O will be as measured (*i.e.* no adjustment for D_2O measurements per modern practice).

Urate oxidase and buffer

- For SAS measurement, ~5 mgs (0.5 mL of 10 mg/mL) of protein in the measurement buffer plus sufficient buffer for SEC-SAS and batch SAS measurement will be shipped at 4-5°C with ice to each lab.
- In the case of SANS/QUOKKA at NIST/30-m SANS, half of the supplied buffer will be lyophilized and reconstituted with Sigma D_2O and used to dialyze half the supplied protein. For ILL/D22, as they will run both batch and SEC-SANS the exchange of H_2O buffer to D_2O buffer will be done by SEC, directly in SEC-SANS by recovering the sample after SEC-SANS, re-concentrating as needed and then measuring in batch mode to acquire data with better statistics.
- In the measurement buffer (see Table 1) and at 4°C the protein is stable for weeks or months. The enzyme is provided in complex with its very high affinity inhibitor 8-azaxanthine (MW=153, all four sites bound, no need for free inhibitor in the buffer) as it is even more stable in this form.
- Prior to shipment, HPLC was performed to confirm that the urate oxidase was pure tetramer with no significant higher order oligomers.
- For batch measurement, ideally, a SEC run on 10µL aliquot could be performed to evaluate the monodispersity of the sample following transport.

SEC-SAS measurements

Equilibrate column with prepared 1x buffer (+ 0.1% azide) and load concentration and volume to achieve concentrations for measurement as indicated in Table 1.

Batch measurements

Ideally two concentrations of each protein will be measured (see table for desired concentrations). Generally, last step dialysate will be used for solvent measurements. Alternatively, solvent exchange could be achieved by a SEC purification step (as is planned for SANS at the ILL).

Data to be Supplied

Initial evaluation and identification of outliers

For each protein the following will be provided:

- For batch measurements:
    - $I(q)$ vs q with $I(q)$ on an absolute scale (cm^{-1}) and associated errors for [Sample+Solvent], [Solvent], and [Sample+Solvent] - [Solvent]
    - Protein concentration value(s)

- For SEC-SAXS
    - Combined/averaged $I(q)$ vs q from experimenter's selected frames/elution times with $I(q)$ on an absolute scale (cm^{-1}) and associated errors for [Sample+Solvent], [Solvent], and [Sample+Solvent] - [Solvent]
    - R_g and $I(0)$ vs frame number/elution time
    - UV/Vis trace (A280) correlated with frame number/elution time, or in the event this is not available any available alternate method for estimating MW of the scattering particle beyond methods dependent on the scattering invariant (*i.e.* Porod volume) as it is desirable to be able to cross check.

For inclusion in the benchmark data set

The reporting on each data set will be per the 2017 guidelines http://scripts.iucr.org/cgi-bin/paper?S2059798317011597.

Table 1: Protein and buffers* to be sent for measurement

Protein	Quantity protein supplied	Buffer and temperature for run	SEC-SAS loading conc./vol.	Batch SAS /WAS
RNase A MW 13690.29 Monomer seq pI 8.64 A_{280} 0.1% 0.690	15 mgs salt free lyophilized powder, Sigma R6513	50 mM Tris pH 7.5, 100 mM NaCl, 25C	Sufficient for peak concentration ~5 mg/mL at measurement	5 & possibly 10 mg/mL for SAS 10 mg/mL for WAS
Xylanase MW 20843.58 Monomer seq pI 8.14 A_{280} 0.1% 2.798	0.5 mL 36 mg/mL glycerol stock Hampton HR7-104, mass spec and SDS gel MW verified	50 mM Tris pH 7.5, 100 mM NaCl, 25C	Sufficient for peak concentration ~5 mg/mL at measurement	5 & possibly 10 mg/mL for SAS 10 mg/mL for WAS
Lysozyme MW 14313.14 monomer seq pI 9.32 A_{280} 0.1% 2.653	10 mgs lyophilized powder >90% protein, Sigma L4919 or L6876	50 mM sodium citrate pH 4.5, 150 mM NaCl, 25 C	Sufficient for peak concentration ~5 mg/mL at measurement	5 & and possibly 10 mg/mL for SAS 10 mg/mL for WAS
Glucose Isomerase MW 43227.37 monomer seq w/o Mg (stable tetramer 172909.48) pI 5.0 A_{280} 0.1% 1.074	0.5 mL 33mg/mL ammonium sulphate precipitate Hampton HR7-102, mass spec and SDS gel MW verified	50 mM Tris pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, 25C	Sufficient for peak concentration of ≤1 mg/mL at measurement	1 mg/mL for SAS 5 & 10 mg/mL for WAS
Urate Oxidase** MW 34150.7 Stable tetramer MW 136603 pI 7.16 A_{280} 0.1% 1.569 psv = 0.734 $cm^3.g^{-1}$	0.5 mL at 10 mg/mL in Buffer at 5°C.	100 mM Tris pH 8.0, 150 mM NaCl, 25C	Sufficient for ~1.5 mg/mL at measurement	5 & possibly 10 mg/mL for SAS 10 mg/mL for WAS
* 0.1% NaN_3 (ReagentPlus, 99.5%, Cat # S2002, from Sigma-Aldrich) to be added to each buffer immediately prior to preparation for SAS measurement as radical scavenger and as it is categorized as a hazard waste, the amount of hazardous waste can be minimized in the dialysis procedures by adding it only for the final dialysate. ** inhibitor: MW=153 Tetramer complex: MW: 136603 + 612 = 137612				

Table 2: Desired measurements with suggested q-ranges ($Å^{-1}$) *

Protein	SEC-SAS	SAS batch mode	WAS**
RNase A	≤ 0.01 – 0.5	≤ 0.01 – 0.5	0.01 – 1
Xylanase	≤ 0.01 – 0.5	≤ 0.01 – 0.5	0.01 – 1
Lysozyme	≤ 0.01 – 0.5	≤ 0.01 – 0.5	0.01 – 1
Urate Oxidase	≤ 0.01 – 0.5	≤ 0.01 – 0.5	0.01 – 1
Glucose Isomerase	≤ 0.006 – 0.35	≤ 0.006 – 0.35	0.01 – 1
* Participant groups can contribute in ranges as they see fit – the suggested ranges are based on what would be of value in a benchmark data set.			

Table 3: Participating beamlines*

SAXS	SANS
ALS/SIBYLS	ANSTO/QUOKKA
APS/BioCAT	ILL/D22
Australian Synchrotron/SAXS-WAXS	NIST/30-m SANS
Diamond/B21	
MacCHESS/BioSAXS	
NIST-UMD/IBBR SAXS	
Petra III/P12	
Shanghai Synchrotron/ BL19U2	
Soleil/SWING	
Spring 8/BL40B2	
SSRL/ BL4-2	
* Table 4 below lists contact information for participants.	

Table 4: Project participants

Name	Affiliation	email contact
Clement Blanchet	Petra III/P12	clement.blanchet@embl-hamburg.de
Srinivas Chakravarthy	APS/BioCAT	schakrav11@gmail.com
Emre Brookes	University of Montana	emre.brookes@mso.umt.edu
Paul Butler	NIST/30-m SANS	paul.butler@nist.gov
Thomas Cleveland	NIST-IBBR	clevelandt@ibbr.umd.edu
Nathan Cowieson	Diamond	nathan.cowieson@diamond.ac.uk
Anthony Duff	ANSTO/National Deuteration Facility	
Frank Gabel	Institut de Biologie Structurale, Grenoble (SASvtf member)	frank.gabel@ibs.fr
Richard Gillilan	MacCHESS/ BioSAXS	reg8@cornell.edu
Melissa Graewert	Petra III/P12	graewert@embl-hamburg.de
Alex Grishaev	NIST-UMD/ IBBR SAXS	agrishaev@ibbr.umd.edu
J. Mitchell Guss	University of Sydney	mitchell.guss@sydney.edu.au
Jesse Hopkins	APS/BioCAT	jhopkins1@iit.edu
Greg Hura	ALS/SIBYLS	glhura@lbl.gov
Thomas Irving	APS/BioCAT	irving@iit.edu
Cy Jeffries	Petra III/P12	cy.jeffries@embl-hamburg.de
Nigel Kirby	Australian Synchrotron/SAXS-WAXS	Nigel.Kirby@synchrotron.org.au
Susan Krueger	NIST/30-m SANS	susan.krueger@nist.gov
Anne Martel	ILL/D22	martela@ill.eu
Li Na	Shanghai Synchrotron/ BL19U2	nli@sibcb.ac.cn
Javier Pérez	Soleil/SWING	javier.perez@synchrotron-soleil.fr
Lois Pollack	Cornell University (SASvtf member)	lp26@cornell.edu
Lionel Porcar	ILL/D22	porcar@ill.eu
Thierry Prangé	Université René Descartes, Paris	Thierry.Prange@univ-paris5.fr
Mattia Rocco	IRCCS AOU San Martino IST	mattia.rocco@hsanmartino.it
Tim Ryan	Australian Synchrotron/SAXS-WAXS	tim.ryan@synchrotron.org.au
Dina Schneidman	Hebrew University (SASvtf member)	dina.schneidman@mail.huji.ac.il
Hiroshi Sekiguchi	JSRI/Spring 8 BL40B2	sekiguchi@spring8.or.jp
Masaaki Sugiyama	Kyoto University (SASvtf member)	sugiyama@rri.kyoto-u.ac.jp
Dmitri Svergun	Petra III/P12 (SASvtf member)	svergun@embl-hamburg.de
Susana Teixeira	NIST/30-m SANS	susanat@udel.edu
Aurélien Thureau	Soleil/SWING	aurelien.thureau@synchrotron-soleil.fr
Jill Trewhella	University of Sydney, Australia (SASvtf Chair)	Jill.trewhella@sydney.edu.au
Patrice Vachette	Université Paris-Sud (SASvtf member)	patrice.vachette@i2bc.paris-saclay.fr
Thomas Weiss	SSRL/ BL4-2	weiss@slac.stanford.edu
Andrew Whitten	ANSTO/QUOKKA	andrew.whitten@ansto.gov.au
Kathleen Wood	ANSTO/QUOKKA	kathleen.wood@ansto.gov.au

Table 5: Five proteins that appear to meet all selection criteria				
	Source	Prior SAXS buffers/conc	"Best" available structures	Issues
RNaseA Bovine MW 13690.29 Monomer seq pI 8.64 Abs 0.1% 0.690	Sigma Aldrich R6513	TR-data Generally PBS, SEC-SAXS data collected with 5% glycerol 50mM TRIS, 150mM NaCl, 0.05% NaN3, 5mM DTT, pH 7.4, Conc. up to 10 mg/mL PV-data Tris 100mM pH8 NaCl 100mM Concentrations up to 10 mg/mL	7RSA 1.26 Å res. (X-ray, no phosphate in binding pocket, excellent stats) 5RSA 2.0 Å res. (neutron, phosphate binding pocket occupied) 6RSA 2.0 Å res. (neutron, w. uridine vanadate) 2E3W 1.06 Å res. (X-ray, synthetic native sequence, modest stats, no phosphate but 3 Cl ions) 2AAS, NMR structure (0.92 Å rmsd w. Xtal) This structure was also refined by RECOORD.	Phosphate binding site occupancy needs to be controlled, available neutron structure means we have explicit H atoms. 5% glycerol was used in the SEC-SAXS expt with co-flow cell to minimize buffer damage due to very high intensity exposure. Not needed to protect protein generally. Best data fit so far in terms of agreement with Xtal str 2E3W is with PV data from Soleil -
Xylanase Tricoderma reesei MW 20843.58 Monomer seq pI 8.14 Abs 0.1% 2.798	Hampton Research HR7-104	SEC-SAXS data set: 10mM Na2HPO4, 1.8 mM KH_2PO_4, 137 mM NaCl and 2.7 mM KCl with 5% glycerol, pH 7.5 Loading conc. 3.5 mg/mL	2DFC, 1.19Å res.	Very high-quality X-ray structure, excellent validation stats. 5% glycerol was used in the co-flow cell to minimize buffer damage due to very high intensity exposure. Not needed to protect protein generally.
Lysozyme Hen egg white Gallus gallus MW 14313.14 monomer seq pI 9.32 Abs 0.1% 2.653	Sigma L4919 or L6876	40 mM NaOAc pH 4.2-4.4, 150mM NaCl, 1-2 mM TCEP, 0.1% NaN_3 Conc. At least to 10 mg/mL also 100 mM NaCl and 100 mM sodium citrate buffer at pH 2.9	2VB1, 0.65 Å res. also 2BLX, 1.4 Å res. 193L, 1.33 Å res. 194L, 1.4 Å res.	2VB1 has superior validation stats, other are modest.
Glucose Isomerase Streptomyces rubiginosus MW 43227.37 monomer seq (w/o Mn/Mg) Tetramer 172909.48 pI 5.0 Abs 0.1% 1.074	Hampton Research HR7-102	Most SAXS data seem to be at pH~7.5, with 150 mM salt. The one ipi free data set for: 10mM HEPES, with 1 mM $MgCl_2$, pH 7.5.	1MNZ, 0.99 Å, only 1 missing amino acid	Very high-quality X-ray structure, excellent validation stats. Mg/Mn occupancy needs to be controlled. Most published SAXS profiles show ipi evidence, even when concentration series is extrapolated. One data set from Australian Synchrotron < 1 mg/ml shows no indication of ipi, and it is in low salt.
Urate oxidase Aspergillus flavus Uniprot Q00511 34150.7*4 =136603 Note Mass spec: 34149±2 Tetramer	Available from Thierry Prangé (per Patrice Vachette)	Tris 50mM pH 8.3, NaCl 100 mM	4POE, 1.07 Å w. uric acid and azide 3L8W, 1.0 Å w. xanthin	Stable for years at 5°C, cannot be frozen, must be shipped cold. Validation stats pretty good for 4POE, excellent for 3L8W excellent