Jill Trewhella
Professor Emeritus

11 June 2019

To the wwPDB Leadership Group
John L. Markley, BMRB
Genji Kurisu, PDBj
Sameer Valenkar, PDBe
Stephen K. Burley, PDB RSCB
Helen Berman, PDB RSCB

Transmitted by email

Dear Colleagues

I write to provide a brief progress report on the project outlined in my letter to you of January 31, 2019 regarding our plans to take a set of well known, relatively rigid structures to test and benchmark different approaches for modelling scattering profiles from PDB coordinates.

The attached project plan has been refined and agreed among 33 participating researchers (Table 4) and samples are in transit for measurements on five standard proteins at 10 X-ray and 3 neutron scattering facilities across the US, Europe and Asia (Table 3). While the participating groups do not include every possible facility, we have endeavoured to achieve as much geographical diversity as possible within our somewhat limited means.

We expect all measurements will be completed over the coming summer, at which time we will begin writing up the results. We trust that these results will be very helpful as we consider data and model validation in regard to small-angle scattering data and its future in integrative/hybrid structural biology.

Yours sincerely,

Jill Trewhella, PhD

cc SASvtf members
Gaetano Montelione, Michael Nilges, Co-Chairs, NMRvtf

Biochemistry Bldg, G08
Faculty of Science
The University of Sydney
NSW 2006 Australia

T +61 (0)4 3275 3933
E jill.trewhella@sydney.edu.au
http://sydney.edu.au/science/people/jill.trewhella.php

ABN 15 211 513 464
CRICOS 00026A

Draft Plan for comment

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 at least 3 proteins with known structures at different SAS beam lines using a common batch for each protein and appropriate standard buffer(s).
- Make a set of these data sets available, along with the measurement protocol and sources of material, to the research community via a suitable Website (Emre Brookes has offered to host).

Criteria for Selection of Proteins

- 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, commercial sources.
- Known buffer conditions for optimal SAS data collection (aggregation free, no interparticle interference).
- Cover a range of sizes.

Table 1 provides complete summary of research for suitable candidates. There are four proteins that appear to meet all criteria:

	Source	Buffer conditions	"Best" available structures	Issues
RNaseA Bovine MW 13682 monomeric	Sigma Aldrich also Worthington Biochemicals	TR-data Generally PBS, SEC-SAXS data collected with 5% glycerol PV-data Tris 100mM pH8 NaCl 100mM	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	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. Apparently 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 -

			also refined by the RECOORD program.	
Xylanase Tricoderma reesei MW 20831.8 monomeric	Hampton Research – high purity	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	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.
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
Glucose Isomerase Streptomyces rubiginosus MW 172912 (w/o Mn/Mg) Tetrameric	available from Aust. Synchrotron originally purchased from Hampton Research (per Tim Ryan)	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.

Protocol:

- Proteins are to be supplied to each participating lab from a single batch source. Suggested amounts per the table. There will be multiple protein shippers; Australian Synchrotron, Thierry Prangé, and TBD for RNase and Xlanase.

Approximate amounts for each protein to send for measurement*			
	SEC-SAXS	SAXS batch mode	WAXS
RNase A	100 µL at 5 mg/mL	400 µL at 5 mg/mL	400 µL at 5 mg/mL
Xlanase	100 µL at 5 mg/mL	400 µL at 5 mg/mL	400 µL at 5 mg/mL
Urate Oxidase	100 µL at 5 mg/mL	400 µL at 5 mg/mL	400 µL at 5 mg/mL
Glucose Isomerase	100 µL at 5 mg/mL	400 µL at 1 mg/mL	400 µL at 5 mg/mL
*Need advice from each beam line as to their specific needs, available columns, etc.			

- Buffers need to be standardized to the extent possible.
  - Both the protein and buffer solutions to be sent from the same lab. Send x10 buffers to be diluted for SEC-SAXS elution using 18 MOhm water. Also perform a dialysis and send a few mLs of buffer to be sent for the batch measurement where buffer matching is critical.

- SAXS-WAXS experiments desired:
  - SEC-SAXS on each protein
  - SAXS measurement in batch mode
  - WAXS-range measurement

Suggested *q*-range ($Å^{-1}$) for each protein/measurement*			
	SEC-SAXS	SAXS batch mode	WAXS
RNase A	0.01 – 0.5	0.01 – 0.5	0.01 – 1
Xlanase	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
* Need advice from each beam line re-specific capabilities for each measurement			

- Where possible, mass spec to be done on batch samples, before and after SAXS, to check for proteolysis. Note: Not all facilities will have easy access to Mass Spectrometry and GI may be a difficult candidate.
- Experiment to be fully documented per the 2017 publication guidelines http://scripts.iucr.org/cgi-bin/paper?S2059798317011597

List of participants

Contacted to date:

Patrice Vachette, Université Paris-Sud, patrice.vachette@i2bc.paris-saclay.fr
Mattia Rocco, IRCCS AOU San Martino IST, mattia.rocco@hsanmartino.it
Jill Trewhella, University of Sydney, jill.trewhella@sydney.edu.au
Emre Brookes, UT Health San Antonio, emre.brookes@mso.umt.edu
Javier Pérez, Soleil, javier.perez@synchrotron-soleil.fr
Nigel Kirby/Tim Ryan, Australian Synchrotron, Nigel.Kirby@synchrotron.org.au; tim.ryan@synchrotron.org.au
Dmitri Svergun/Cy Jeffries, EMBL-Hamburg, svergun@embl-hamburg.de; cy.jeffries@embl-hamburg.de
Frank Gabel, IBS/ILL, frank.gabel@ibs.fr
Alex Grishaev, NIST/UMD, and 12ID at APS, agrishaev@ibbr.umd.edu
Thomas Weiss, Stanford/SSRL, weiss@slac.stanford.edu

Potential Measurements (to date) at:

Soleil - SWING
Australian Synchrotron – SAXS/WAXS
Petra III – P12
Lab-based SAXS instrument – NIST via Alex G.
ILL (neutrons) – via Frank Gabel, inst. tbd, perhaps D22
NIST (neutrons) – NG7 30m SANS, via Alex G. and NIST contacts
APS – 12ID via Alex G.
SSRL – BL4-2