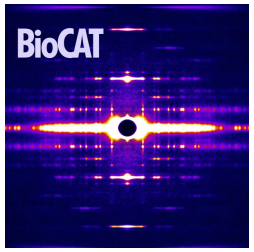
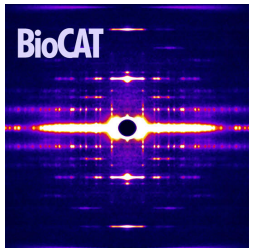


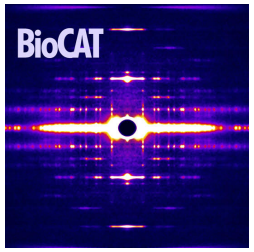
How to publish SAXS data

Jesse Hopkins, PhD
IIT/CSRRI
Staff Scientist, BioCAT
Sector 18, Advanced Photon Source



How to publish SAXS data

- SAXS data is extremely powerful, but there are many ways it can go wrong
- Need to take care during analysis to not fool yourself
- Need to present data correctly in publications to not fool your readers
- Accurately and honestly present strengths and limitations of data
  - Imperfect data can be used. For example, a small amount of aggregation could be accepted if you were simply using the SAXS to verify flexibility

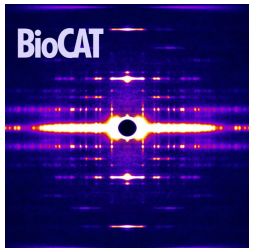


Publication guidelines

2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. J. Trewhella, A. P. Duff, D. Durand, F. Gabel, J. M. Guss, W. A. Hendrickson, G. L. Hura, D. A. Jacques, N. M. Kirby, A. H. Kwan, J. Pérez, L. Pollack, T. M. Ryan, A. Sali, D. Schneidman-Duhovny, T. Schwede, D. I. Svergun, M. Sugiyama, J. A. Tainer, P. Vachette, J. Westbrook and A. E. Whitten. Acta Cryst. (2017). D73, 710-728.

- https://doi.org/10.1107/S2059798317011597
- Open access

Provides guidelines for how to present data, what data to present so that readers can independently assess quality of data and models presented.



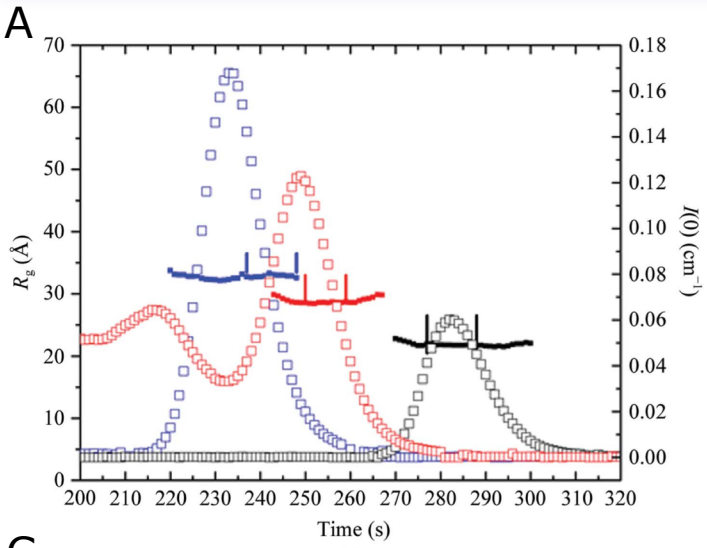
Publication guidelines

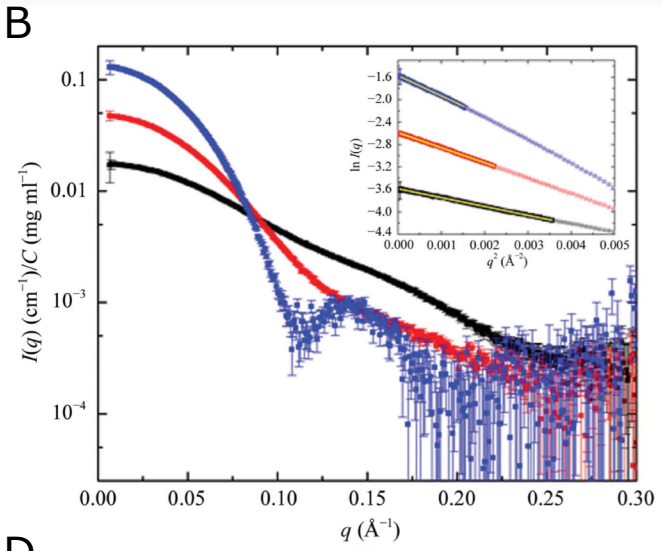
- Reporting guidelines with summary tables for:
  - Sample details
  - Data acquisition and reduction
  - Data presentation, analysis, and validation
  - Structure modeling

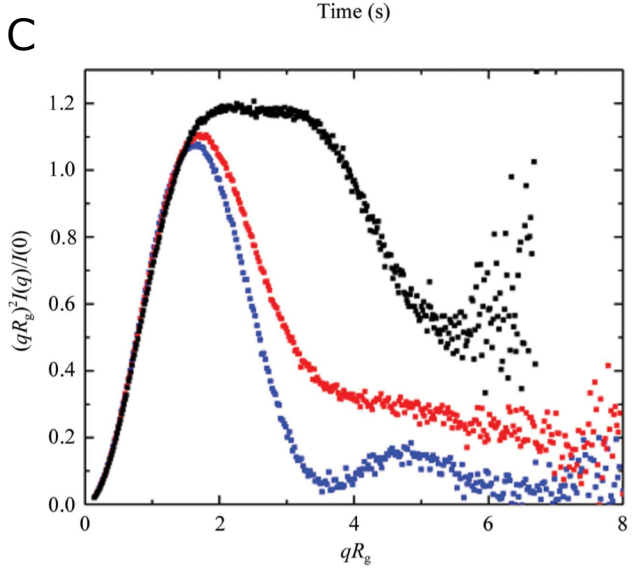
- Example report, including figures and tables, on SEC-SAXS from three well-known proteins

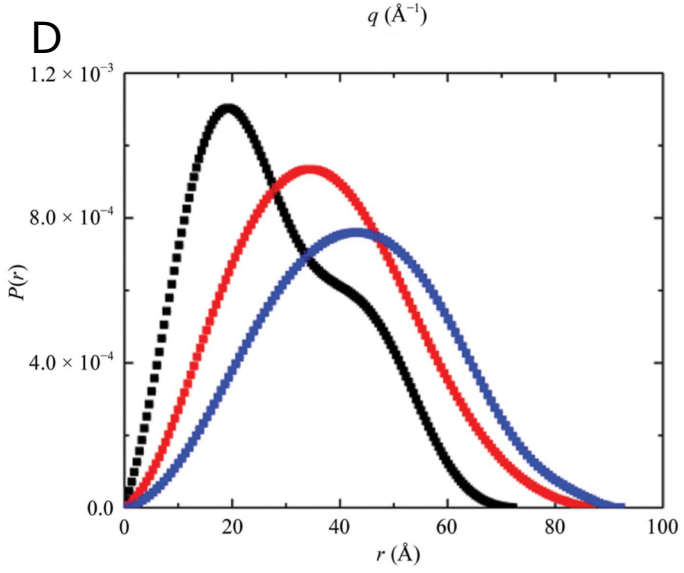
- Template reporting table

Publications guidelines









A. Intensity and Rg vs. frame number or time

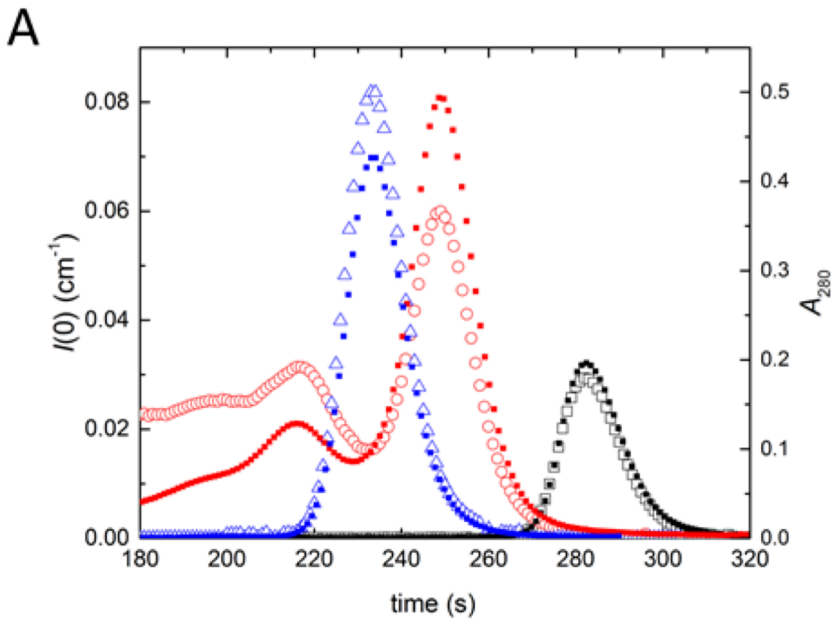
B. I(q) vs. q as log-lin, Guinier fits

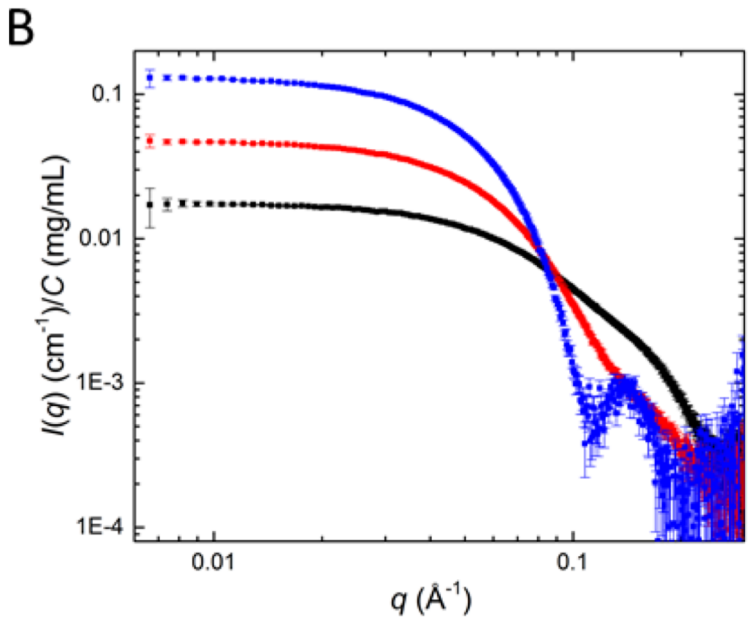
C. Dimensionless Kratky plots (puts all proteins on similar scale, highlights flexibility and shape)

D. P(r) vs. r profiles, normalized to equal area (i.e. by I(0))

Plot from Trewhella et al, Acta Cryst D73, 2017

Publication guidelines



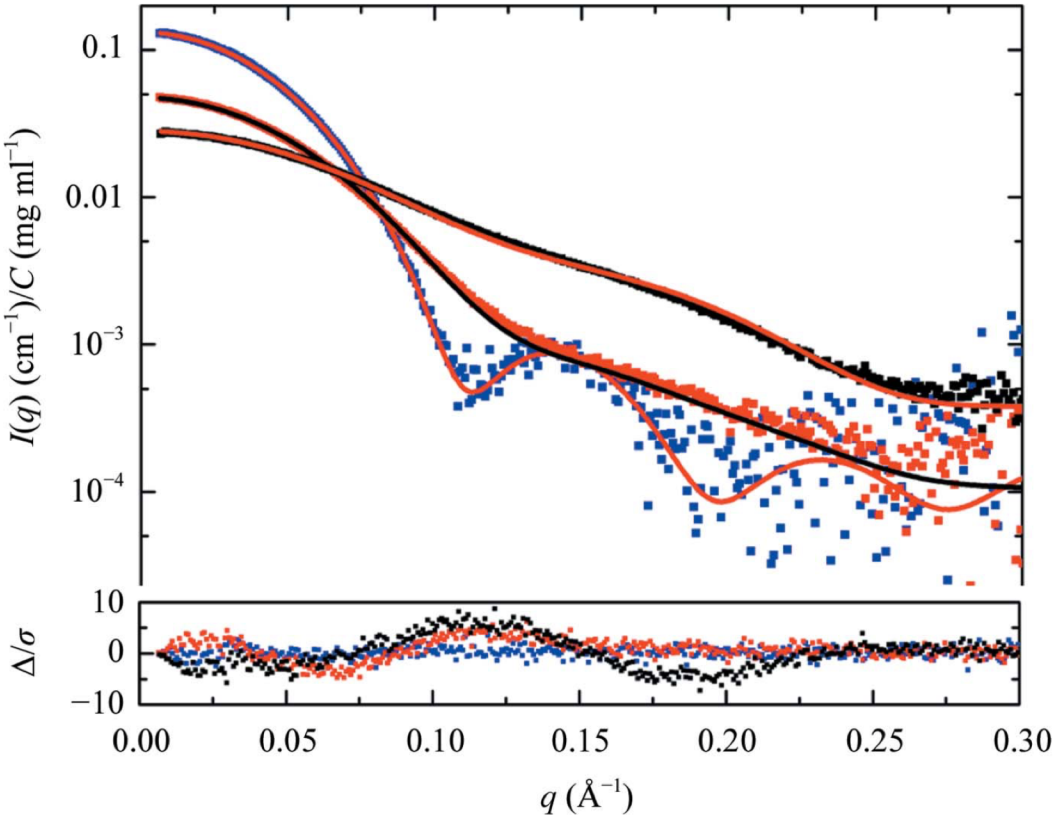


A. UV traces, which should correspond reasonably well to SAXS curve

B. Log-log plots showing appropriate low-q behavior

Plot from Trewhella et al, Acta Cryst D73, 2017

Publication guidelines

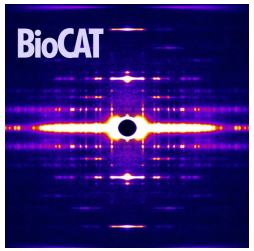


For modeling results, fits to data should be shown, as should residuals

Paper recommends plotting error weighted residuals

$$\frac{\Delta}{\sigma}(q) = \frac{I_{exp}(q) - cI_{mod}(q)}{\sigma(q)}$$

Plot from Trewhella et al, Acta Cryst D73, 2017

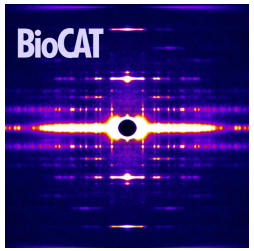


Publication guidelines

Report sample details

	GI (tetramer)	BSA	CaM
Organism	*Streptomyces rubiginosus*	*Bos taurus*	*Xenopus laevis*
Source (catalogue No. or reference)	Hampton Research (HR7-100)	Sigma–Aldrich (A3294)	*E. coli* expressed (Michie *et al.*, 2016)
UniProt sequence ID (residues in construct)	P24300 (2–388)	P02769 (25–607)	P62155 (2–149)
Extinction coefficient [A_{280}, 0.1%(*w*/*v*)]	1.075	0.646	0.178
$\overline{\nu}$ from chemical composition ($cm^3\ g^{-1}$)	0.732	0.732	0.716
Particle contrast from sequence and solvent constituents, $\Delta\overline{\rho}$ ($\rho_{protein} - \rho_{solvent}$; $10^{10}\ cm^{-2}$)	2.87 (12.39 − 9.52)	2.86 (12.38 − 5.92)	3.09 (12.61 − 5.92)
M from chemical composition (Da)	172912	66400	16842
SEC–SAXS column, 5 × 150 mm Superdex S200			
Loading concentration (mg ml^{-1})	6	25	20.2
Injection volume (µl)	30	35	35
Flow rate (ml min^{-1})	0.45	0.45	0.45
Average *C* in combined data frames (mg ml^{-1})	0.58 (0.20–1.09)	1.81 (1.01–2.45)	3.09 (2.38–3.55)
Solvent (solvent blanks taken from SEC flowthrough prior to elution of protein)	25 m*M* MOPS, 250 m*M* NaCl, 50 m*M* KCl, 2 m*M* TCEP, 0.1% NaN_3 pH 7.5		

Table from Trewhella et al, Acta Cryst D73, 2017

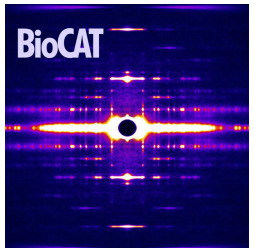


Publication guidelines

Report data collection parameters

Instrument/data processing	Australian Synchrotron SAXS/WAXS beamline with Dectris PILATUS 1M detector (Kirby *et al.*, 2013)
Wavelength (Å)	1.0332
Beam size (µm)	250 × 130
Camera length (m)	2.683
q measurement range ($Å^{-1}$)	0.00663–0.3104
Absolute scaling method	Comparison with scattering from 1 mm pure H_2O
Normalization	To transmitted intensity by beam-stop counter
Monitoring for radiation damage	X-ray dose maintained below 210 Gy, data frame-by-frame comparison
Exposure time	Continuous 1 s data-frame measurements of SEC elution
Sample configuration	SEC–SAXS with sheath-flow cell (Kirby *et al.*, 2016), effective sample path length 0.49 mm
Sample temperature (°C)	22

Table from Trewhella et al, Acta Cryst D73, 2017

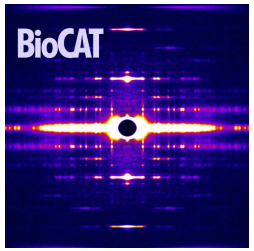


Publication guidelines

Report software used for reduction and analysis

SAXS data reduction	*I*(*q*) *versus q* using *ScatterBrain* 2.82 (http://www.synchrotron.org.au/aussyncbeamlines/saxswaxs/software-saxswaxs), solvent subtraction using *PRIMUSqt* (*ATSAS* 2.8.0; Petoukhov *et al.*, 2012)
Extinction coefficient estimate	*ProtParam* (Gasteiger *et al.*, 2005)
Calculation of $\Delta\overline{\rho}$ and $\overline{\nu}$ values	*MULCh* 1.1 (06/10/16; Whitten *et al.*, 2008)
Basic analyses: Guinier, $P(r)$, V_P	*PRIMUSqt* from *ATSAS* 2.8.0 (Petoukhov *et al.*, 2012)
Shape/bead modelling	*DAMMIF* (Franke & Svergun, 2009) and *DAMMIN* (Svergun, 1999) *via ATSAS* online (https://www.embl-hamburg.de/biosaxs/atsas-online/)
Atomic structure modelling	*FoXS* (Schneidman-Duhovny *et al.*, 2013) *via* web server (https://modbase.compbio.ucsf.edu/foxs/) *CRYSOL* from *PRIMUSqt* in *ATSAS* 2.8.1 (Svergun *et al.*, 1995) *MultiFoXS* (Schneidman-Duhovny *et al.*, 2016) *via* web server (https://modbase.compbio.ucsf.edu/multifoxs/) *EOM* (Bernadó *et al.*, 2007) *via ATSAS* online (https://www.embl-hamburg.de/biosaxs/atsas-online/)
Missing sequence modelling	*MODELLER* (https://salilab.org?modeller/; Webb & Sali, 2014)
Three-dimensional graphic model representations	*PyMOL* v.1.70.0.5 Win64

Table from Trewhella et al, Acta Cryst D73, 2017

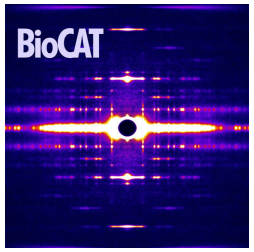


Publication guidelines

Report structural parameters from Guinier fits, P(r) functions, MW estimates

	GI (tetramer)	BSA	CaM
Guinier analysis			
$I(0)$ (cm^{-1})	0.0759 ± 0.0008	0.0861 ± 0.0008	0.0554 ± 0.00008
R_g (Å)	32.87 ± 0.13	28.33 ± 0.05	21.74 ± 0.06
q_{min} ($Å^{-1}$)	0.007	0.007	0.007
qR_g max (q_{min} = 0.0066 $Å^{-1}$)	1.3	1.3	1.3
Coefficient of correlation, R^2	0.999	0.999	0.999
M from $I(0)$ (ratio to predicted)	178312 (1.03)	65589 (0.99)	21944 (1.31)
$P(r)$ analysis			
$I(0)$ (cm^{-1})	0.0748 ± 0.00008	0.0850 ± 0.00006	0.0533 ± 0.00006
R_g (Å)	32.65 ± 0.04	28.32 ± 0.03	22.2 ± 0.06
d_{max} (Å)	92	87	72
q range ($Å^{-1}$)	0.007–0.243	0.007–0.282	0.0074–0.310
χ^2 (total estimate from *GNOM*)	0.929 (0.94)	0.858 (0.96)	0.855 (0.91)
M from $I(0)$ (ratio to predicted value)	180191 (1.04)	65354 (1.00)	21718 (1.29)
Porod volume ($Å^{-3}$) (ratio V_P/calculated M)	229000 (1.3)	101000 (1.5)	25200 (1.5)
V, M using the Fischer method (ratio of M to expected)	192400, 157.9 (0.91)	82440, 67.9 (1.02)	21550, 17.7 (1.05)

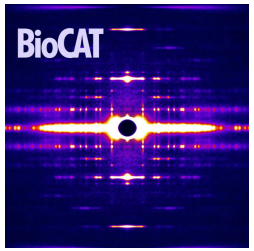
Table from Trewhella et al, Acta Cryst D73, 2017



Publication guidelines

Report modelling results

	GI (tetramer)	BSA	CaM
DAMMIF (default parameters, 20 calculations)			
q range for fitting (Å^{-1})	0.007–0.243	0.007–0.282	0.007–0.310
Symmetry, anisotropy assumptions	*P*1, none	*P*1, none	*P*1, prolate
NSD (standard deviation), No. of clusters	0.62 (0.01), 1	0.75 (0.63), 6	0.77 (0.02), 4
χ^2 range	2.25–2.29	0.96–0.99	1.30–1.37
Constant adjustment to intensities	Skipped, unable to determine	1.51×10^{-4}	1.48×10^{-4}
Resolution (from *SASRES*) (Å)	37 ± 3	32 ± 3	30 ± 3
M estimate as 0.5 × volume of models (Da) (ratio to expected)	134000 (0.77)	66700 (1.00)	16300 (0.97)
DAMMIN (default parameters)			
q range for fitting (Å^{-1})	0.007–0.243	0.007–0.282	0.007–0.310
Symmetry, anisotropy assumptions	*P*1	*P*1	*P*1
χ^2, *CORMAP* *P*-values	0.95, 0.04	0.85, 0.16	0.844, 0.53
Constant adjustment to intensities	2.697×10^{-5}	7.736×10^{-5}	1.877×10^{-4}



Publication guidelines

Report modelling results

Crystal structures	PDB entry 1oad	PDB entry 4f5s (chain *A*)	PDB entry 1cll+†
q range for all modelling	0.007–0.243	0.007–0.282	0.007–0.310
FoXS‡			
χ^2, *P*-value	1.02, 0.05	4.4, 0.00	9.2, 0.00
Predicted R_g (Å)	31.70	26.75	21.58
c_1, c_2	1.03, 0.81	0.99, 2.39	0.99, 2.94
CRYSOL§ (with default parameters)			
No constant subtraction			
χ^2, *P*-value	1.00, 0.05	2.78, 0.00	15.95, 0.00
Predicted R_g (Å)	32.69	27.89	22.51
Vol (Å), Ra (Å), Dro (e Å^{-3})	230987, 1.80, 0.0130	76791, 1.80, 0.035	20271, 1.40, 0.025
Constant subtraction allowed			
χ^2, *P*-value	1.01, 0.05	2.14, 0.00	12.62, 0.00
Predicted R_g (Å)	32.71	28.01	22.11
Vol (Å), Ra (Å), Dro (e Å^{-3})	226689, 1.40, 0.013	76791, 1.80, 0.037	22012, 1.40, 0.055
Multistate/ensemble models			
Starting crystal structures		PDB entry 4f5s (chain *A*)	PDB entry 1cll+†
Flexible residues		183–187 and 381–384	1–3 (ADQ), 77–87 (KDTDS)
MultiFoXS¶ (10 000 models in starting set)			
No. of states		1	1
χ^2, *CORMAP P*-values		1.05, 0.02	0.85, 0.31
c_1, c_2		0.99, 0.63	1.05, 0.99
R_g values of each state (Å)		27.59	21.03
Weights w_n		1	1
No. of states		2	2
χ^2, *CORMAP P*-values		0.96, 0.09	0.79, 0.79
c_1, c_2		1.02, 1.21	1.02, 1.50
R_g values of each state (Å)		26.42, 32.35	22.32, 19.47
Weights w_n		0.83, 0.17	0.70, 0.30
No. of states		3	3
χ^2, *CORMAP P*-values		0.82, 0.17	0.79, 0.79
c_1, c_2		1.02, 0.94	1.02, 1.52
R_g values of each state (Å)		26.42, 30.43, 29.80	22.32, 30.25, 19.00
Weights w_n		0.74, 0.08, 0.08	0.68, 0.13, 0.18
EOM (default parameters, 10 000 models in initial ensemble, native-like models, constant subtraction allowed)			
χ^2, *CORMAP P*-values			0.82, 0.79
Constant subtraction			0
No. of representative structures			13

Publication guidelines

- These guidelines are extremely thorough. Follow them as best you can and you "won't mislead or be misled" (J. Trewhella)

- The paper includes a supplemental word document with the tables that you can download and fill out

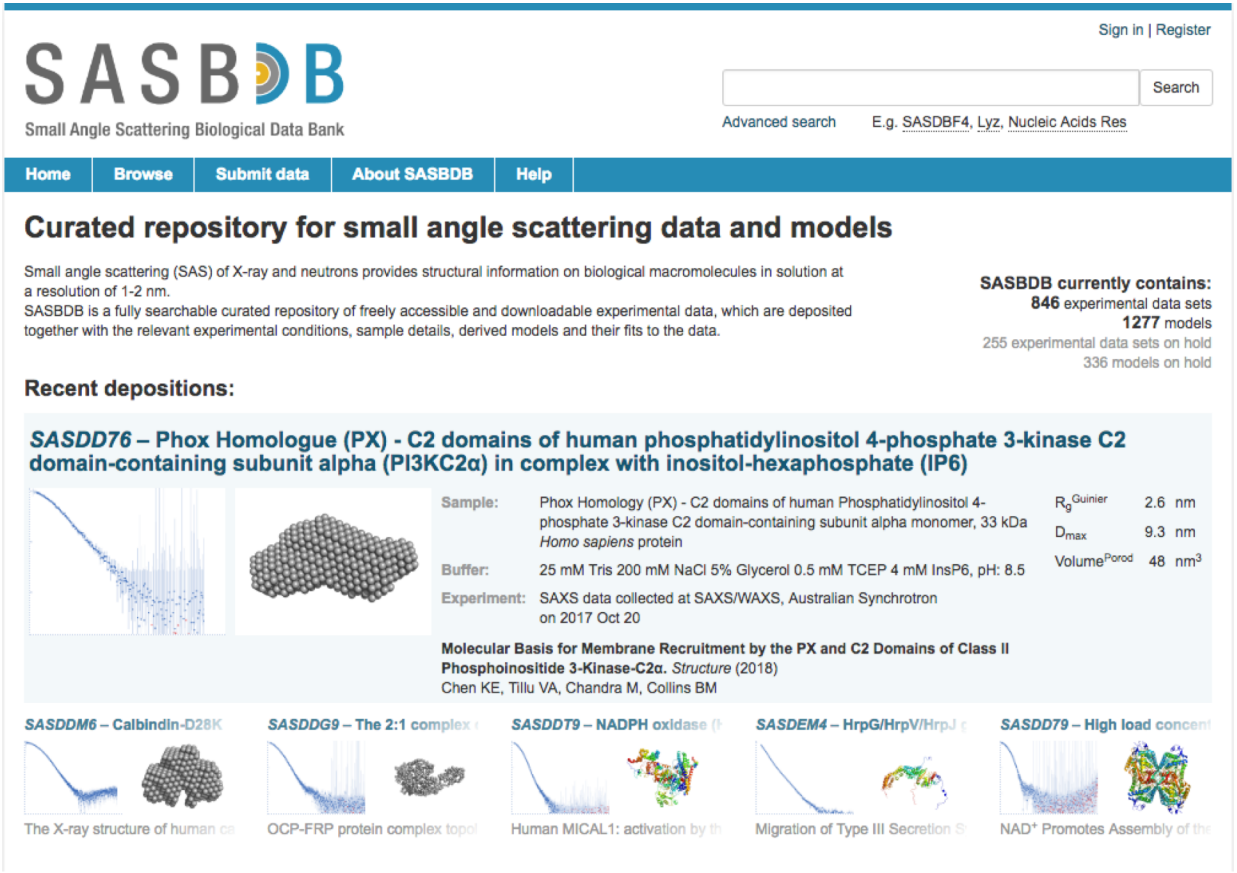
Table S1 Reporting template for tabulating essential SAS data acquisition, sample details, data analysis, modelling fitting and software used.

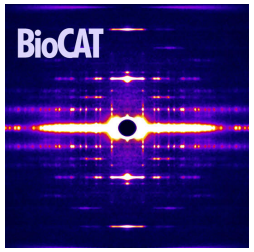
(*a*) Sample details			
	Sample 1	Sample 2	Sample 3, *etc.*
Organism			
Source (Catalogue No. or reference)			
Description: sequence (including Uniprot ID + uncleaved tags), bound ligands/modifications, *etc.*			
Extinction coefficient ε (wavelength and units)			
Partial specific volume $\bar{\upsilon}$ ($cm^3\ g^{-1}$)			
Mean solute and solvent scattering length densities and mean scattering contrast $\Delta\bar{\rho}$ (cm^{-2})			
Molecular mass *M* from chemical composition (Da)			
For SEC-SAS, loading volume/concentration, (mg ml^{-1}) injection volume (µl), flow rate (ml min^{-1})			
Concentration (range/values) measured and method			
Solvent composition and source			

Data deposition

It is now recommended (but not required) that you deposit your SAXS data in an online repository

- Most commonly the SASBDB (https://www.sasbdb.org/)





Publishing your BioCAT results

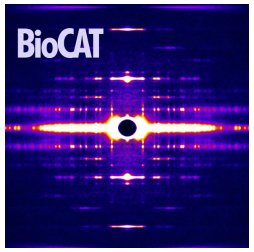
- In addition to getting the science right, we need you to help us

- When publishing results from BioCAT, you need to include the following acknowledgement (most facilities have similar):

  This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. "This project was supported by grant 9 P41 GM103622 from the National Institute of General Medical Sciences of the National Institutes of Health." Use of the Pilatus 3 1M detector was provided by grant 1S10OD018090-01 from NIGMS.

- User output is how we justify our existence to the NIH. So if you want to keep collecting SAXS data here, we need you to acknowledge us

- Publications with results from BioCAT also need to be submitted to PubMed and made open access according the the NIH public access policy



Publishing your BioCAT results

- We want you to get it right
- Questions about data analysis, contact us
- Want us to read over your methods section or check your analysis, contact us
- Any time you collect data, you should get a handout with the acknowledgements, sample methods section, and the relevant tables filled out. Take a look.