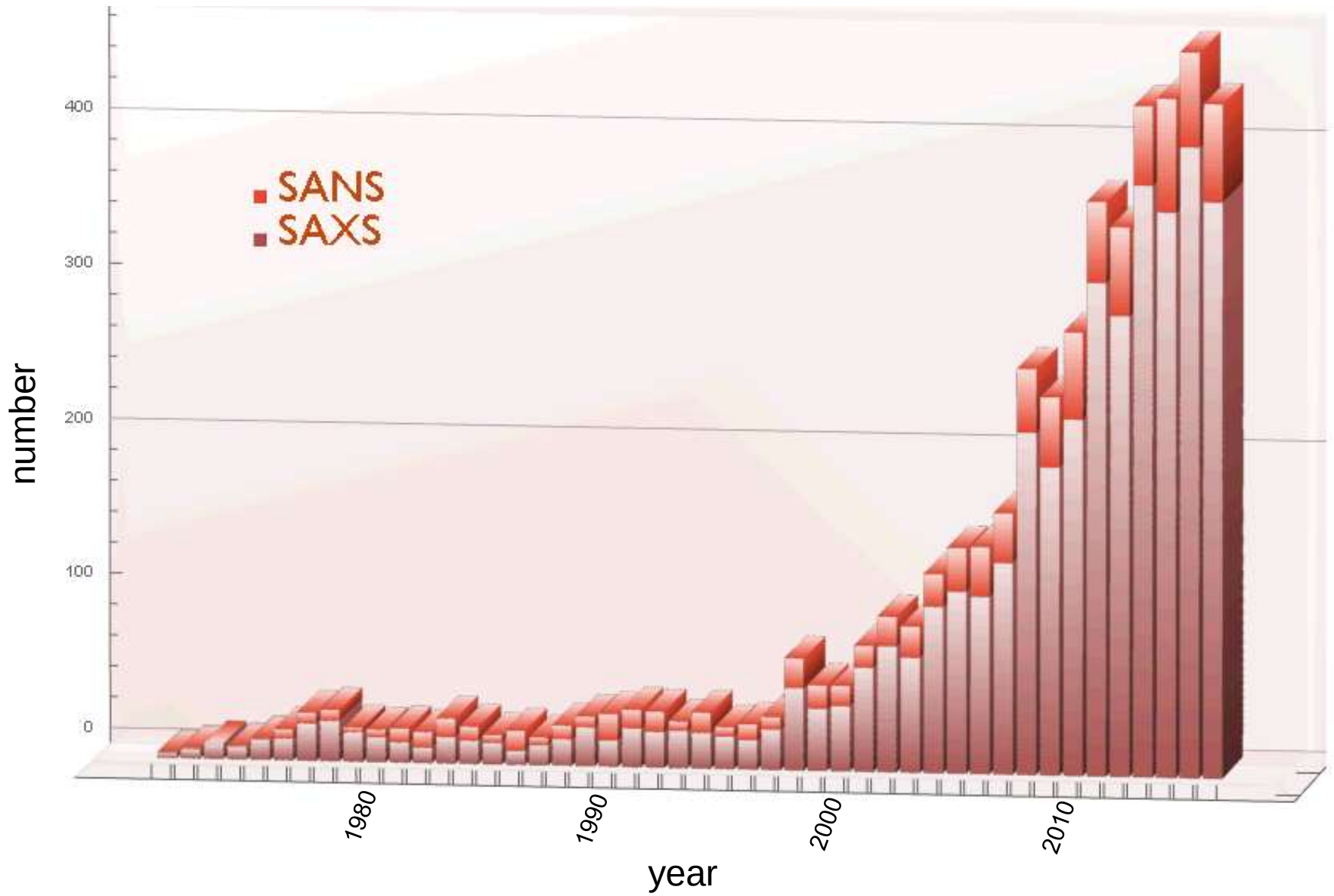
Publishing your SAXS Data

Thomas M. Weiss
Stanford University, SSRL/SLAC,
BioSAXS beamline BL 4-2

BioSAXS2019, Dec 2-5, 2019

Bio-SAS Publications in PubMed (1971 – 2016)



The Problem:

SAXS data provides only few simple cross checks of consistency and data quality

SAXS related structural models in the literature increases

Number of researchers using SAXS increases

(not all are SAXS experts)

Analysis algorithms more sophisticated and more divers

How can one judge othe accuracy of the analysis and the reliability of the structural model?

SAXS publication guidelines

research papers

STRUCTURAL BIOLOGY

ISSN 2059-7983

2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update

Jill Trewhella,[a]* Anthony P. Duff,[b] Dominique Durand,[c] Frank Gabel,[d] J. Mitchell Guss,[a] Wayne A. Hendrickson,[e] Greg L. Hura,[f] David A. Jacques,[g] Nigel M. Kirby,[h] Ann H. Kwan,[a] Javier Pérez,[i] Lois Pollack,[j] Timothy M. Ryan,[h] Andrej Sali,[k] Dina Schneidman-Duhovny,[l] Torsten Schwede,[m] Dmitri I. Svergun,[n] Masaaki Sugiyama,[o] John A. Tainer,[p] Patrice Vachette,[c] John Westbrook[q] and Andrew E. Whitten[b]

[a]School of Life and Environmental Sciences, The University of Sydney, NSW 2006, Australia, [b]ANSTO, New Illawarra Road, Lucas Heights, NSW 2234, Australia, [c]Institut de Biologie Intégrative de la Cellule, UMR 9198, Bâtiment 430, Université Paris-Sud, 91405 Orsay CEDEX, France, [d]Université Grenoble Alpes, Commissariat à l'Energie Atomique (CEA), Centre National de la Recherche Scientifique (CNRS), Institut de Biologie Structurale (IBS), and Institut Laue–Langevin, 71 Avenue des Martyrs, 38000 Grenoble, France, [e]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA, [f]Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [g]University of Technology Sydney, ithree Institute, 15 Broadway, Ultimo, NSW 2007, Australia, [h]Australian Synchrotron, 800 Blackburn Road, Clayton, VIC 3168, Australia, [i]Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP48, 91192 Gif-sur-Yvette CEDEX, France, [j]School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853-2501, USA, [k]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences (QB3), University of California San Francisco, San Francisco, California, USA, [l]School of Computer Science and Engineering, Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel, [m]Biozentrum, University of Basel and SIB Swiss

Received 16 June 2017
Accepted 7 August 2017

Edited by M. Czjzek, Station Biologique de Roscoff, France

Keywords: small-angle scattering; SAXS; SANS; biomolecular structure; proteins; DNA; RNA; structural modelling; hybrid structural modelling; publication guidelines; integrative structural biology.

Supporting information: this article has supporting information at journals.iucr.org/d

Trewella *et al.* Acta Cryst. D **73**, 710 (2017)

SAXS publication guidelines

should cover and describe in detail all aspects of the SAXS experiment (table):

- Sample preparation
- Data collection (instrument and configuration)
- Data reduction and modelling (software and methods)
- Derived structural parameters
- Details of ab-initio model fits
- Comparison with high-resolution structures
- If deposited (which is recommended): SASBDB ID

Should include also data (figures/plots):

- Log(I) vs q, Guinier and Kratky plot of the data
- If SEC-SAXS overlay with of SAXS data with R_g (and UV)
- Calculated P(r) function from the data
- If available comparison to X-tal structure (possibly with refinement by SREFLEX)
- If multiconformations or flexible parts also show fit and possible average dammin model

Sample Preparation and Data collection

(*a*) 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
$\bar{\nu}$ from chemical composition ($cm^3\ g^{-1}$)	0.732	0.732	0.716
Particle contrast from sequence and solvent constituents, $\Delta\bar{\rho}$ ($\rho_{protein} - \rho_{solvent}$; $10^{10}\ cm^{-2}$)	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		

(*b*) SAXS 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

Software used for processing

(*c*) Software employed for SAXS data reduction, analysis and interpretation.

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

Structural Parameters

(*d*) Structural parameters.

	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)

(*e*) Shape model-fitting 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}

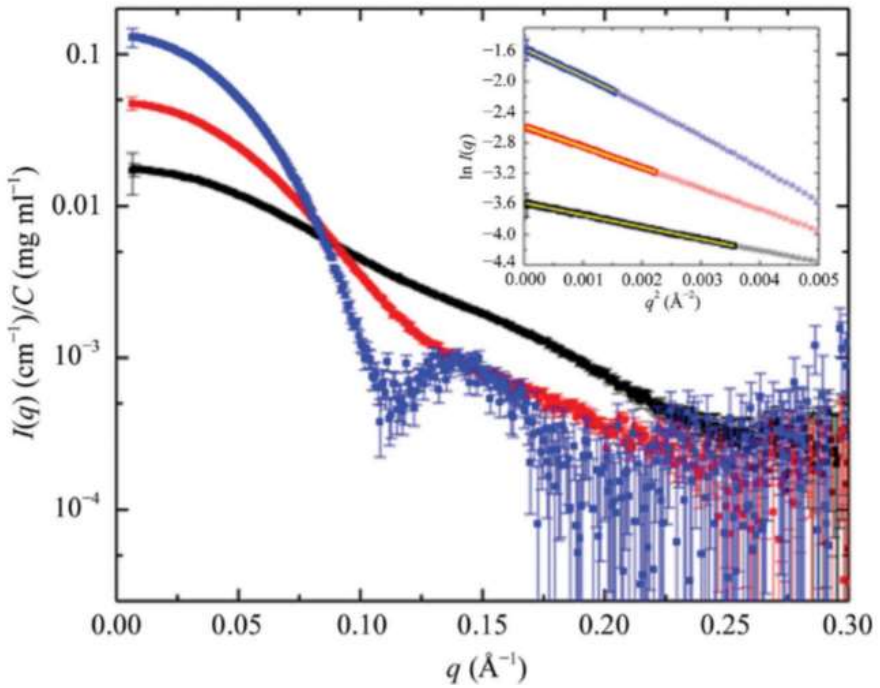
Do you have X-tal structures available?

(*f*) Atomistic modelling.

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

SAXS data plots

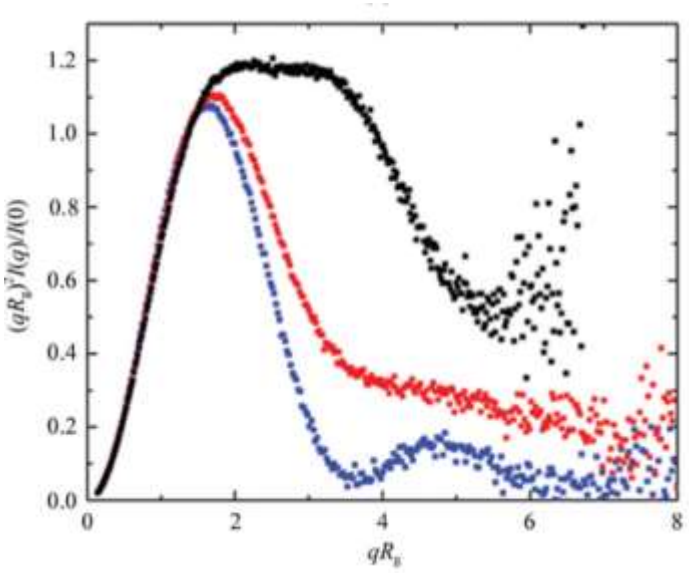
Data I vs q including Guinier Analysis



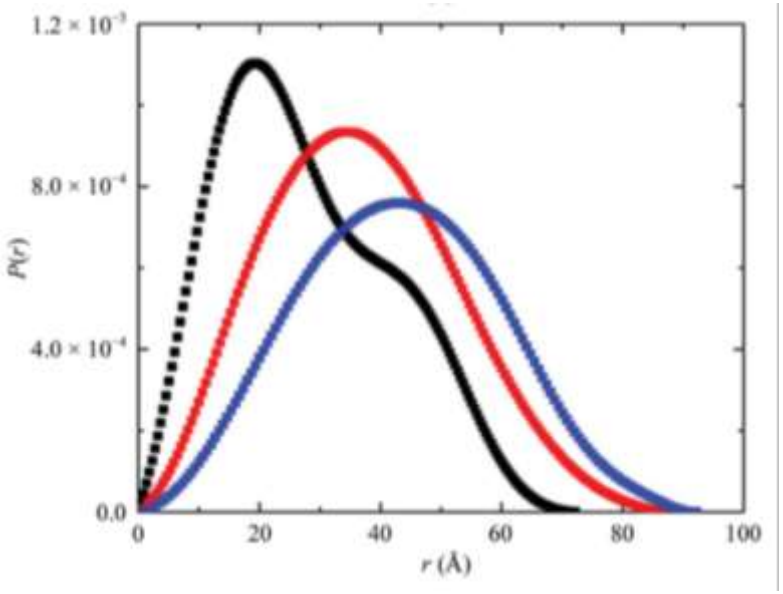
- Most basic report
- If exclude some low angle points should be mentioned in Table
- Should be on absolute scale (intensity)
- Here normalized by intensity (not necessary if conc. Is mentioned somewhere)
- If merged data is used concentration needs to be correctly adjusted

additional data plots

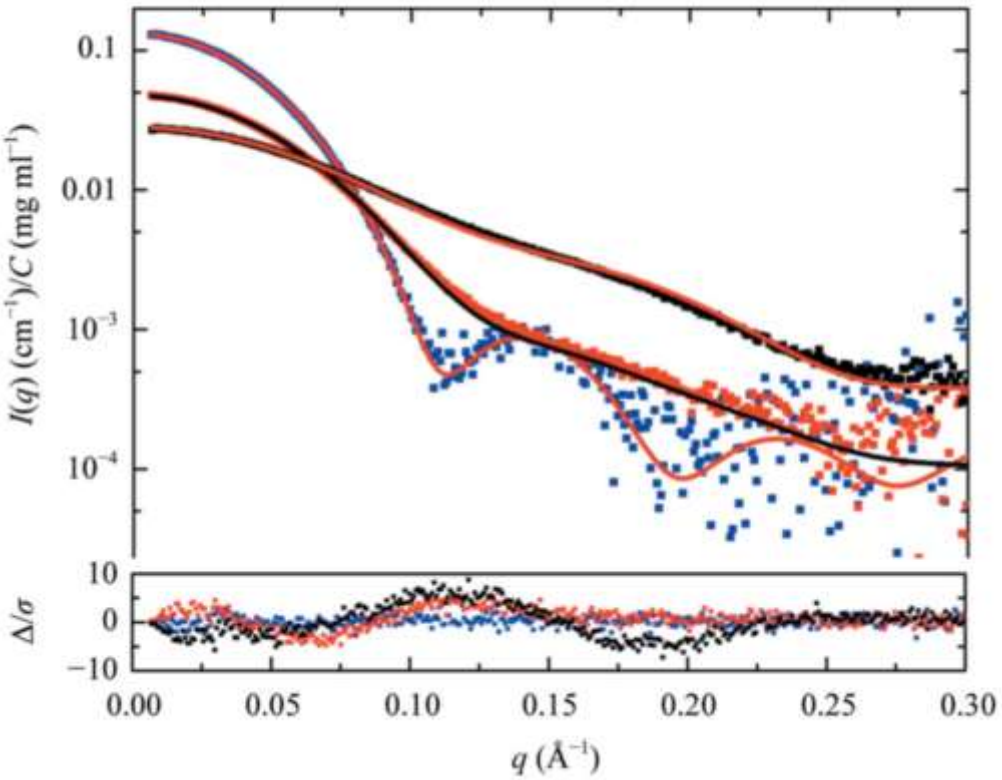
Kratky



P(r): pair distance distribution function (Gnom)



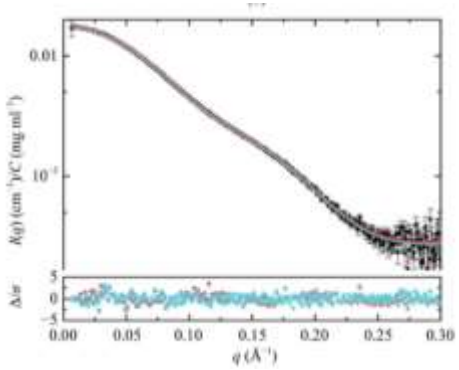
Comparison with X-tal structures

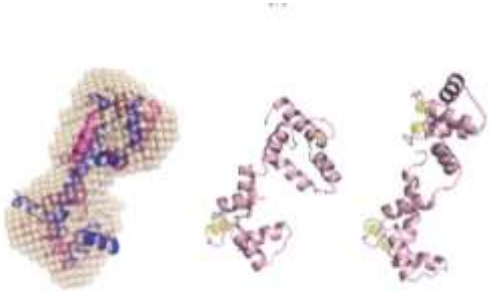


- using Crysol (ATSAS), FOXS (Sali Lab) or others to fit the data to the Xtal structure
- Report data and fit (chi^2 value and residuals)

Report on Multiple Conformations

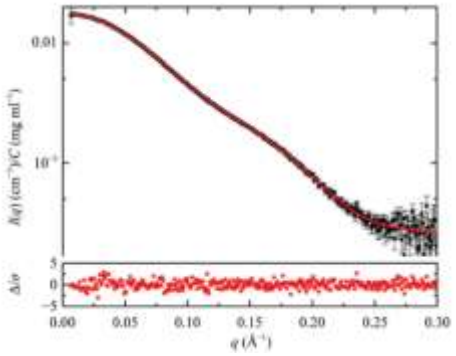
Flexible linker

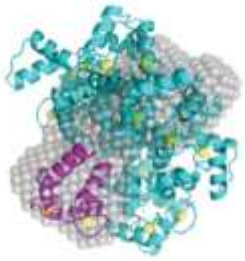


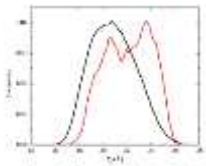


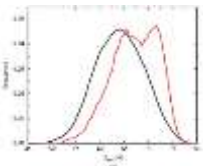
- If few conformations, report different conformers and population
- fit to data with residuals

Ensemble optimization method (EOM)









- Full ensemble method, show fit to data and residual
- Selected structures (probably DAMCLUST classes
- Show Rg and Dmax distribution before and after EOM

Summary

- Those are guidelines not requirements – but most of them are fairly reasonable and probably should be used
- Lots of it can go into the supporting information
- You can deposit data to data base (SASBDB) but not required
- Journals don't have settled on requirements (yet) but many are pushing for them

Use your own judgement

SAXS experiments sometimes do not fit is specific categories and might have other reporting requirements.