Full wwPDB Integrative Structure Validation Report

June 15, 2020 -- 10:00 PM

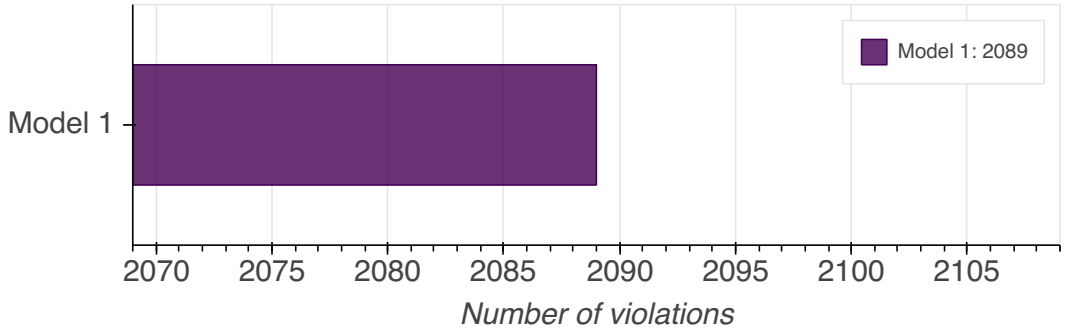
The following software were used in the production of this report:

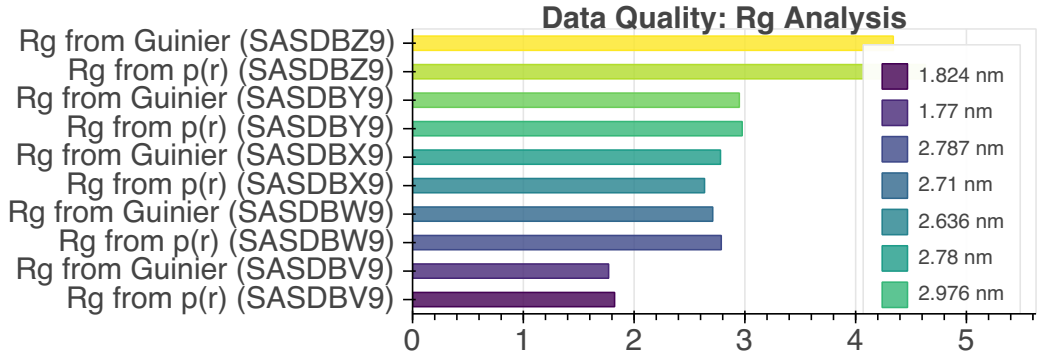
Integrative Modeling Validation Package : Version 1.0

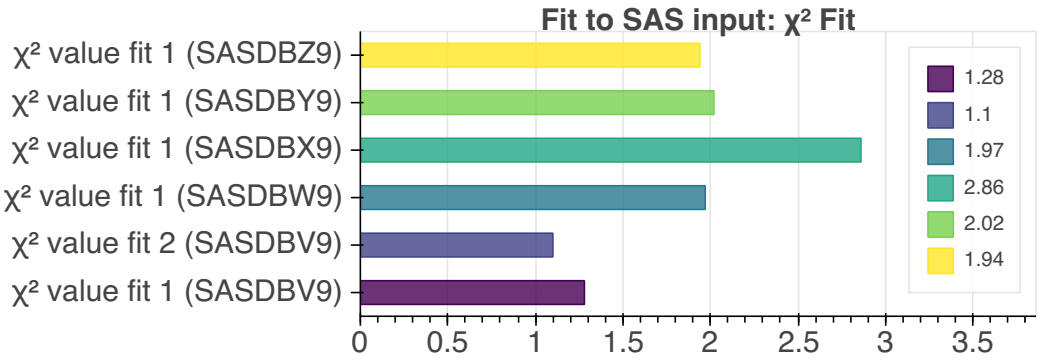
PDB ID	PDBDEV_00000017
Molecule Name	Molecular architecture of the major membrane ring component, Pom152, of the yeast nuclear pore complex
Title	Molecular Architecture of the Major Membrane Ring Component of the Nuclear Pore Complex.
Authors	Upla P;Kim SJ;Sampathkumar P;Dutta K;Cahill SM;Chemmama IE;Williams R;Bonanno JB;Rice WJ;Stokes DL;Cowburn D;Almo SC;Sali A;Rout MP;Fernandez-Martinez J

Overall quality

Model quality: Excluded Volume Analysis







Ensemble information

This entry consists of 1 distinct ensemble.

Entry composition

There is 1 unique type of model in this entry.This model is titled Cluster 0/None respectively.

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Total residues
1	1	1	pom152	A	1337

Methodology and software

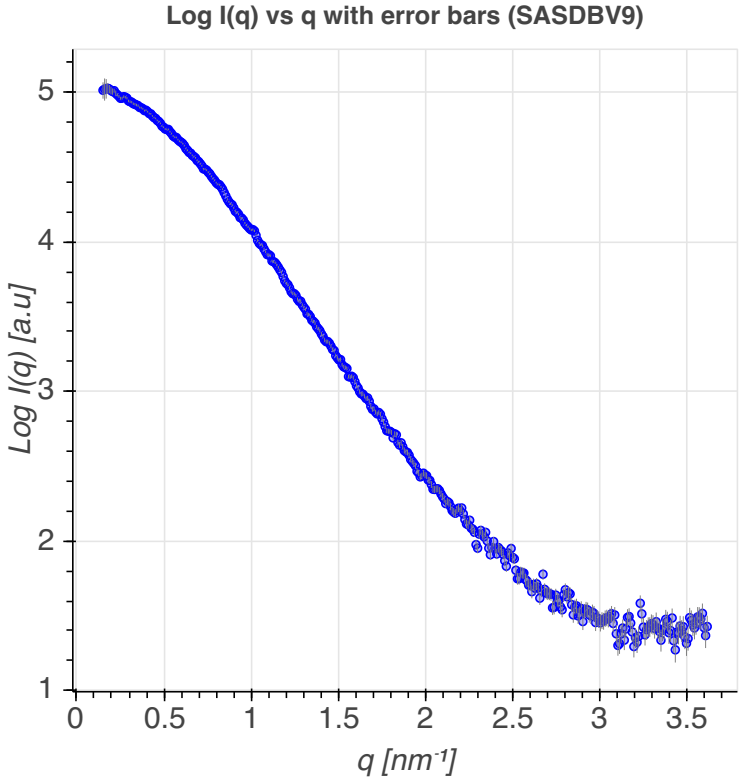
Step number	Protocol ID	Method name	Method type	Number of computed models	Multi state modeling	Multi scale modeling
1	1	Replica exchange monte carlo	Sampling	100000	False	True

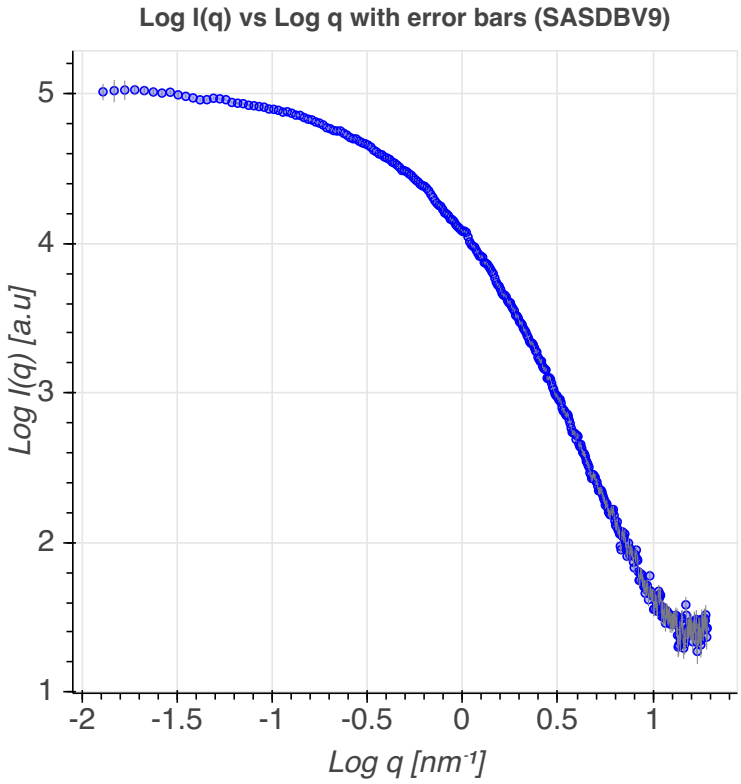
Data quality

Scattering profile

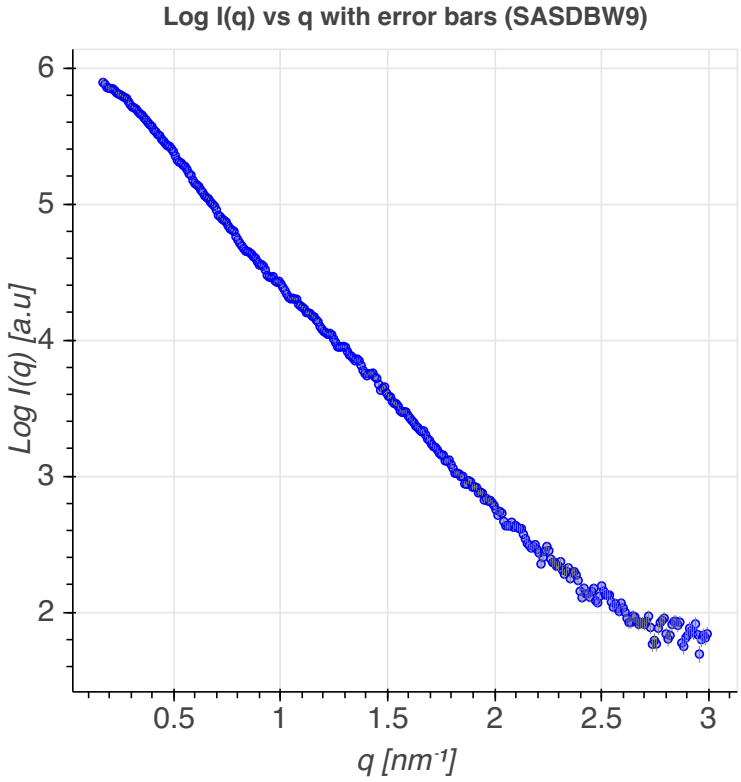
SAS data used in this integrative model was obtained from 5 deposited SASBDB entry (entries).

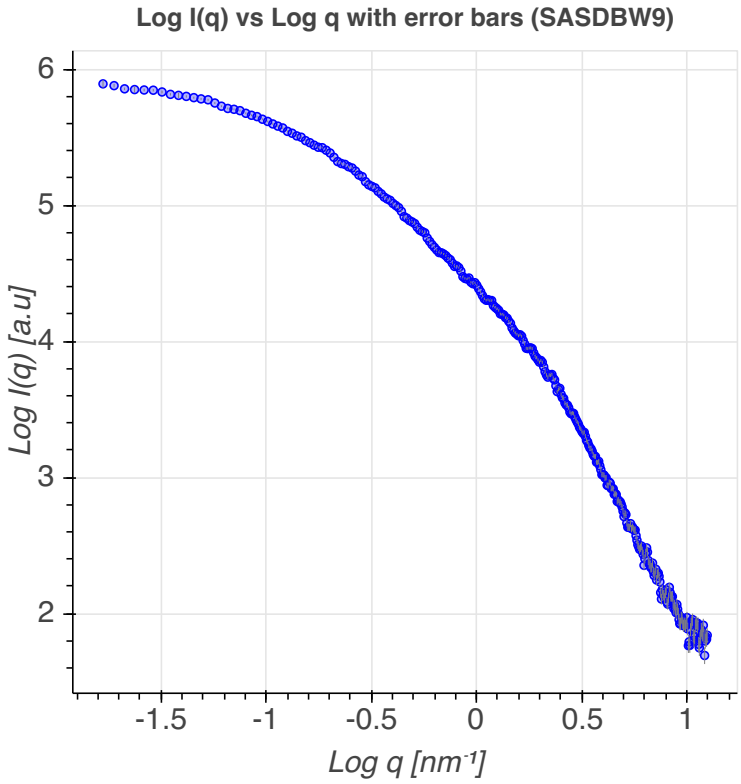
Scattering profile for SASDBV9: data from solutions of biological macromolecules are presented as both log I(q) vs q and log I(q) vs log (q) based on SAS validation task force (SASvtf) recommendations. I(q) is the intensity (in arbitrary units) and q is the modulus of the scattering vector.



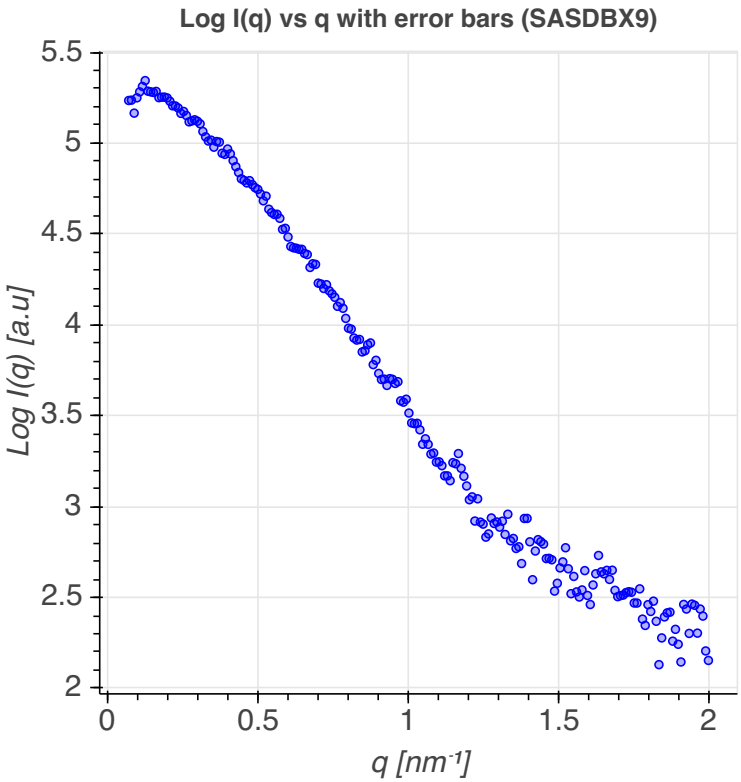


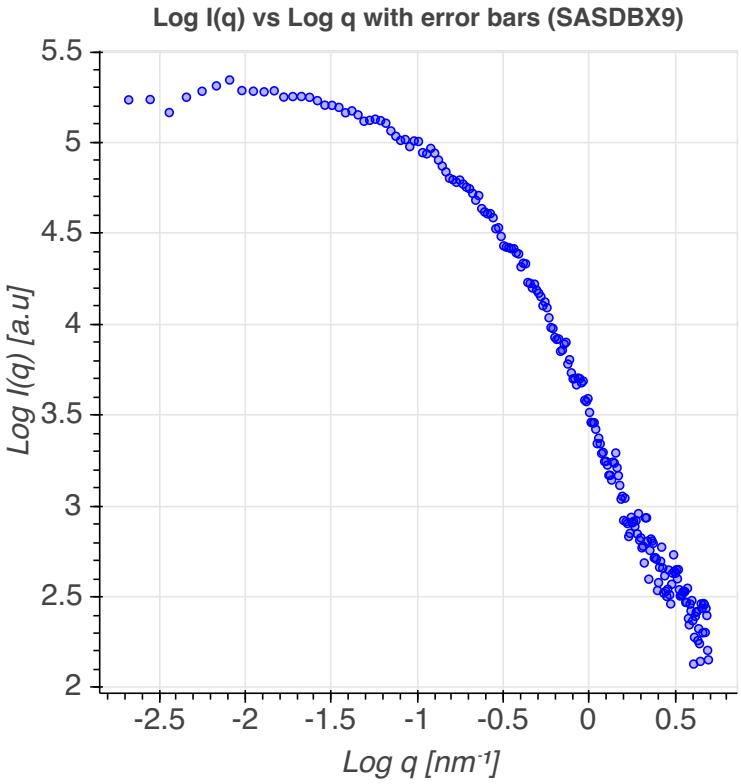
Scattering profile for SASDBW9: data from solutions of biological macromolecules are presented as both log I(q) vs q and log I(q) vs log (q) based on SAS validation task force (SASvtf) recommendations. I(q) is the intensity (in arbitrary units) and q is the modulus of the scattering vector.



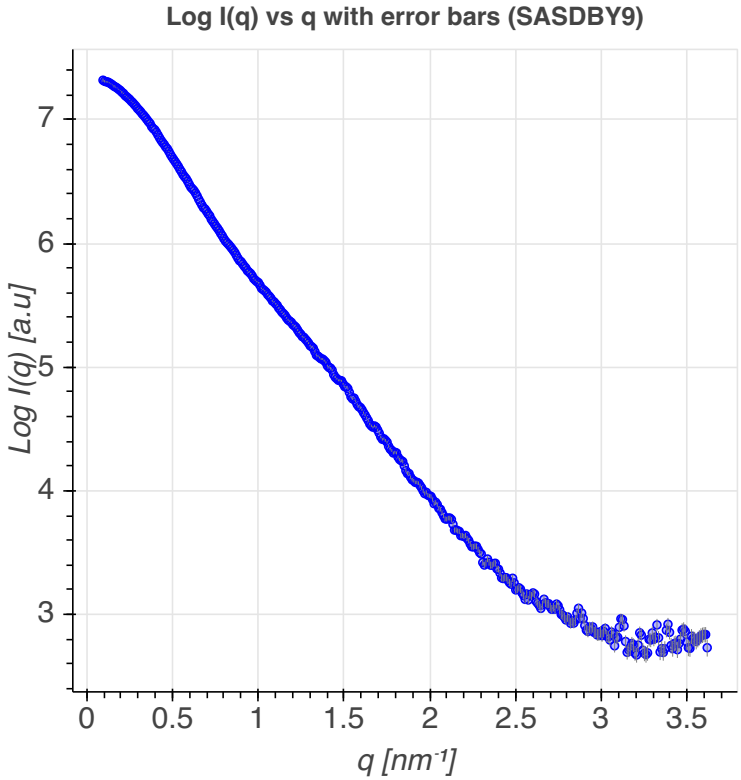


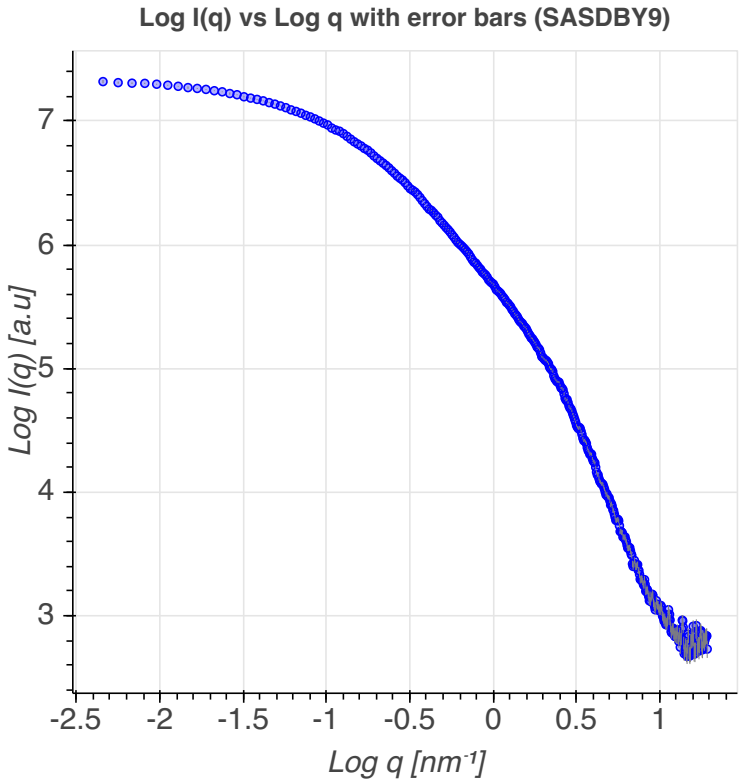
Scattering profile for SASDBX9: data from solutions of biological macromolecules are presented as both log I(q) vs q and log I(q) vs log (q) based on SAS validation task force (SASvtf) recommendations. I(q) is the intensity (in arbitrary units) and q is the modulus of the scattering vector.



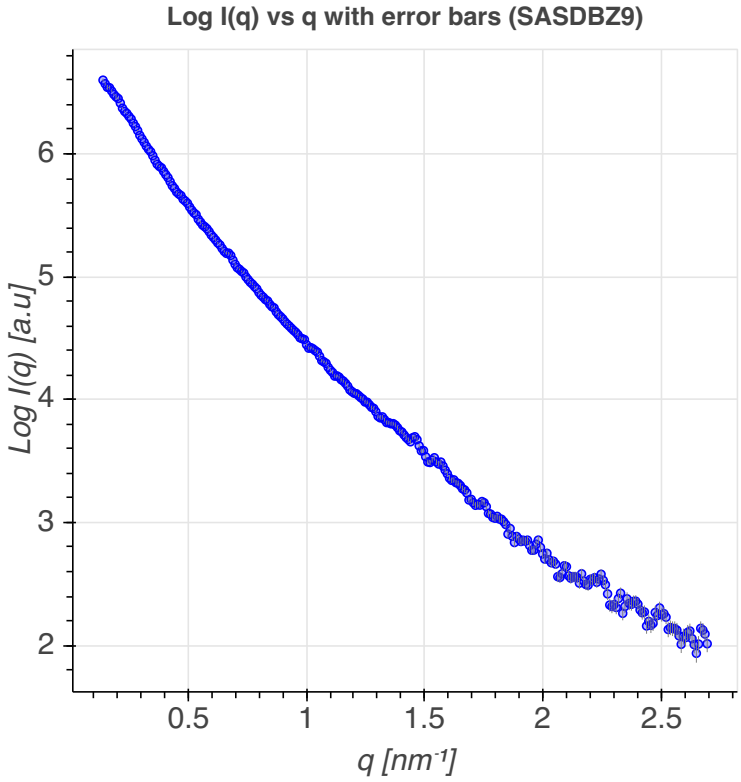


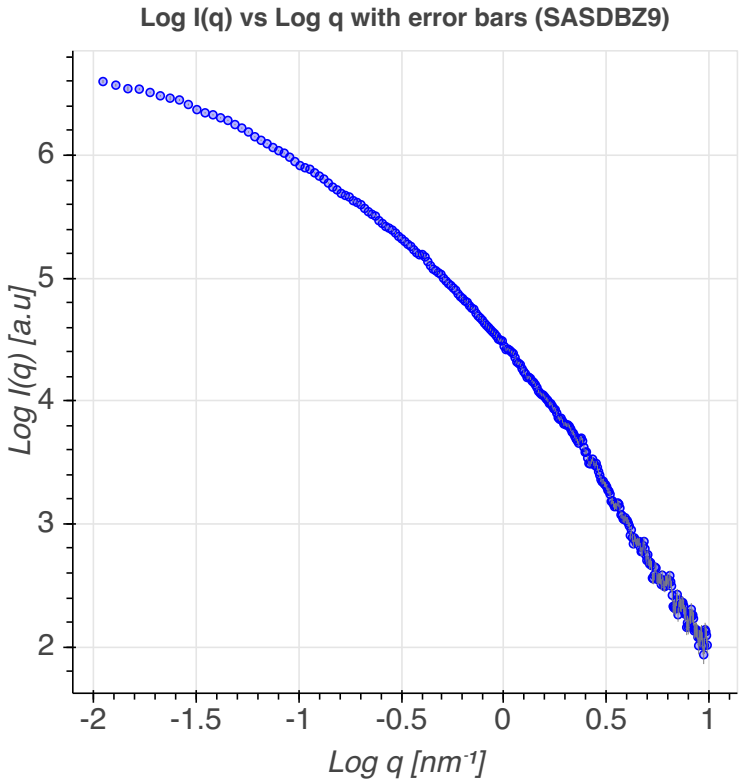
Scattering profile for SASDBY9: data from solutions of biological macromolecules are presented as both log I(q) vs q and log I(q) vs log (q) based on SAS validation task force (SASvtf) recommendations. I(q) is the intensity (in arbitrary units) and q is the modulus of the scattering vector.





Scattering profile for SASDBZ9: data from solutions of biological macromolecules are presented as both log I(q) vs q and log I(q) vs log (q) based on SAS validation task force (SASvtf) recommendations. I(q) is the intensity (in arbitrary units) and q is the modulus of the scattering vector.





Key experimental estimates

Molecular weight (MW) estimates from experiments and analysis: true molecular weight can be compared to the Porod estimate from scattering profiles.

SASDB ID	Chemical composition MW	Standard MW	Porod MW
SASDBV9	12.6 kDa	12.2 kDa	N/A
SASDBW9	24.1 kDa	25.2 kDa	N/A
SASDBX9	12.5 kDa	14.7 kDa	N/A
SASDBY9	25.9 kDa	25.2 kDa	N/A
SASDBZ9	49.4 kDa	48.3 kDa	N/A

Volume estimates from experiments and analysis: estimated volume can be compared to Porod volume obtained from scattering profiles.

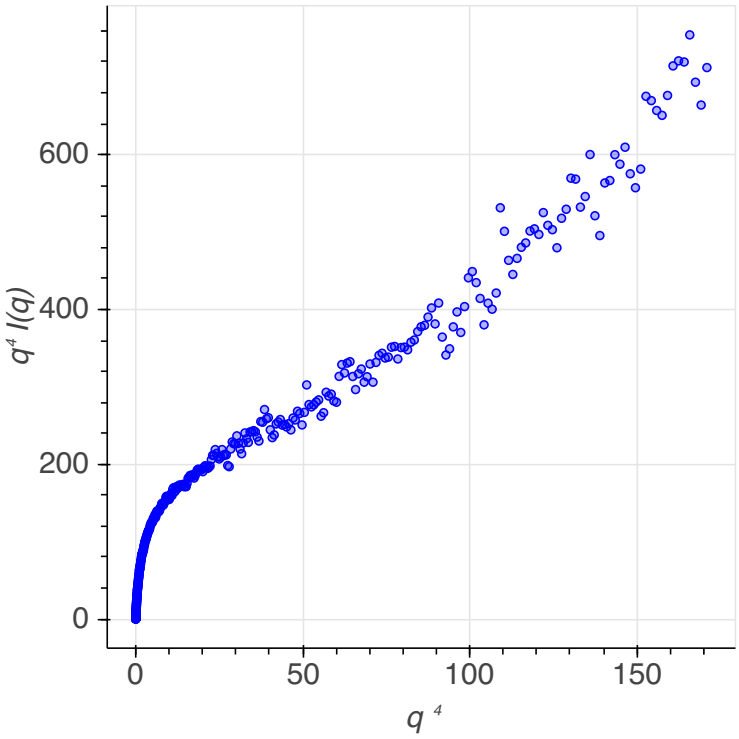
SASDB ID	Estimated volume	Estimated volume method	Porod volume
SASDBV9	None	None	17.94 nm³

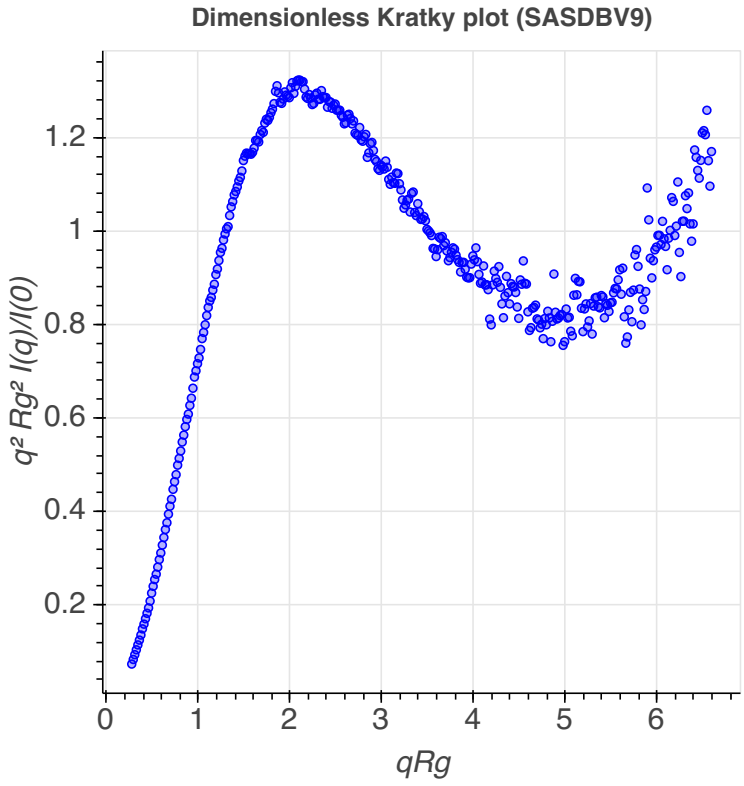
SASDB ID	Estimated volume	Estimated volume method	Porod volume
SASDBW9	None	None	22.50 nm³
SASDBX9	None	None	56.68 nm³
SASDBY9	None	None	27.97 nm³
SASDBZ9	None	None	66.59 nm³

Flexibility analysis

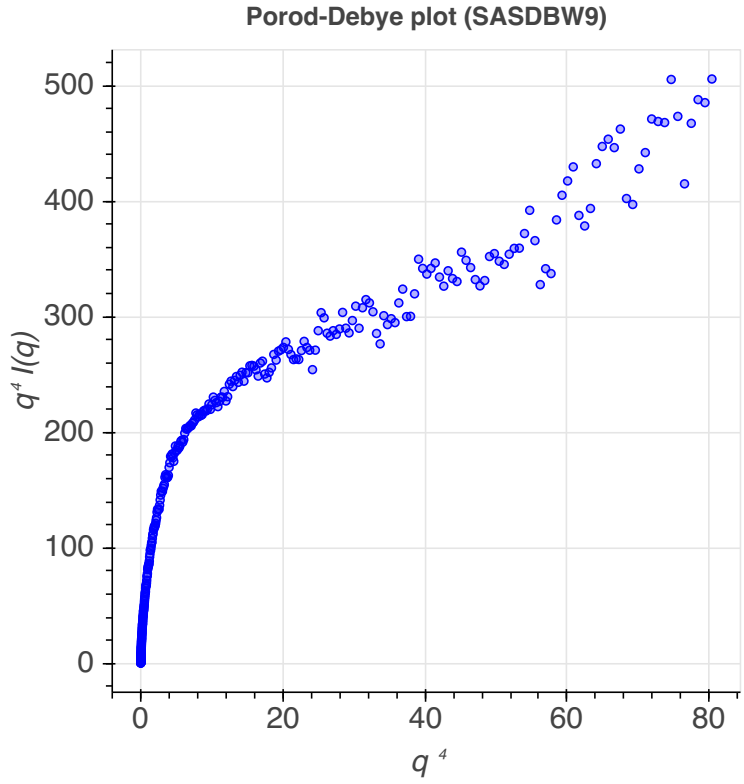
Flexibility analysis for SASDBV9: In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas, fully unfolded domains are devoid of any discernable plateau. For details, refer to Figure 5 in Rambo and Tainer, 2011 . In a Kratky plot, a parabolic shape is observed for globular (partial or fully folded) domains and a hyperbolic shape is observed for fully unfolded domains.

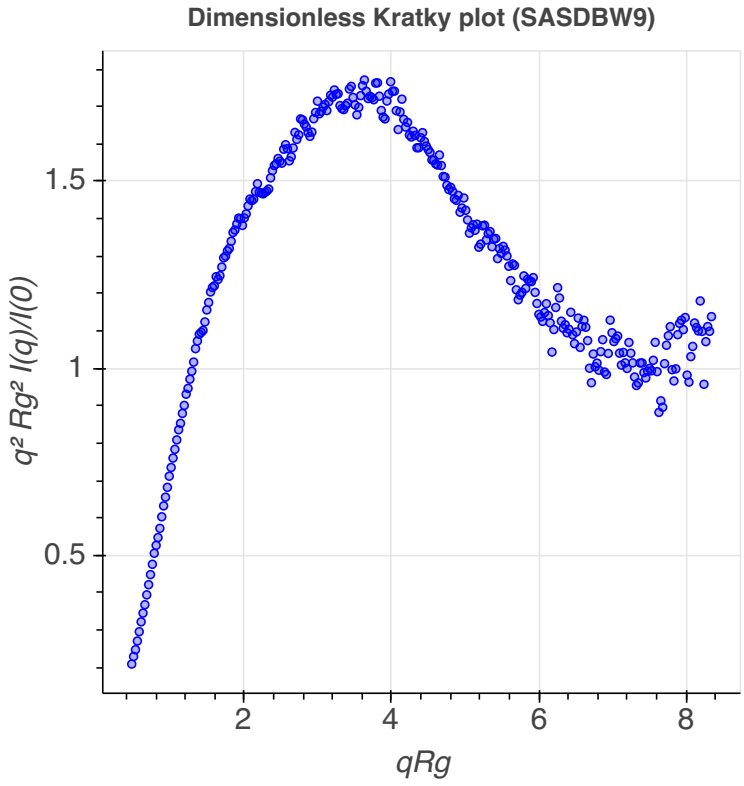
Porod-Debye plot (SASDBV9)



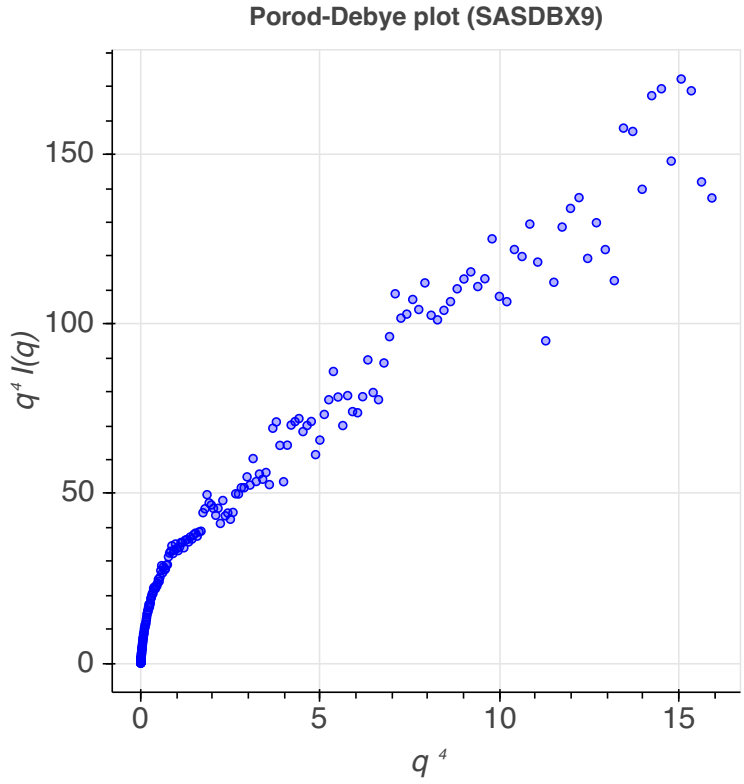


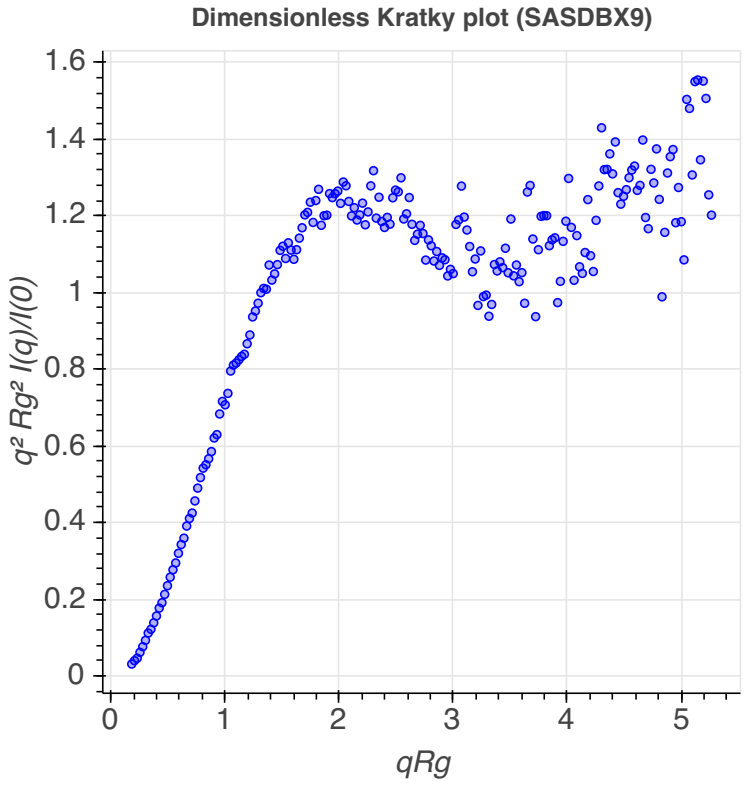
Flexibility analysis for SASDBW9: In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas, fully unfolded domains are devoid of any discernable plateau. For details, refer to Figure 5 in Rambo and Tainer, 2011 . In a Kratky plot, a parabolic shape is observed for globular (partial or fully folded) domains and a hyperbolic shape is observed for fully unfolded domains.



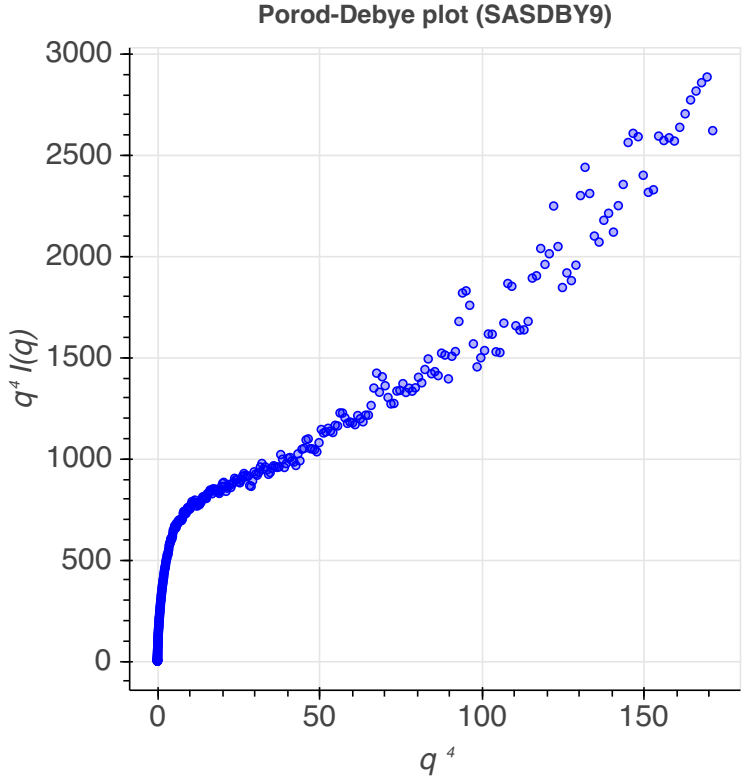


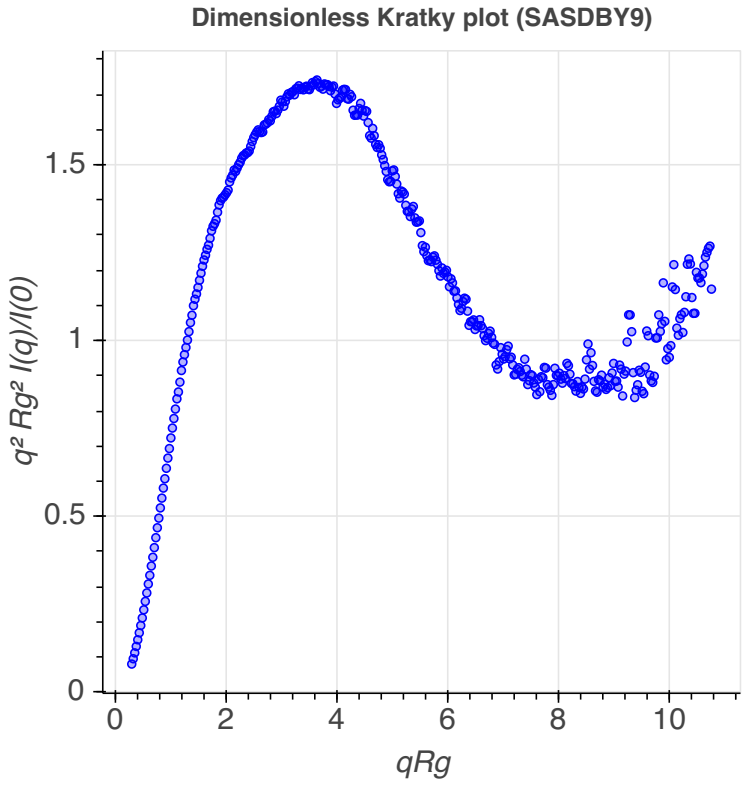
Flexibility analysis for SASDBX9: In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas, fully unfolded domains are devoid of any discernable plateau. For details, refer to Figure 5 in Rambo and Tainer, 2011 . In a Kratky plot, a parabolic shape is observed for globular (partial or fully folded) domains and a hyperbolic shape is observed for fully unfolded domains.



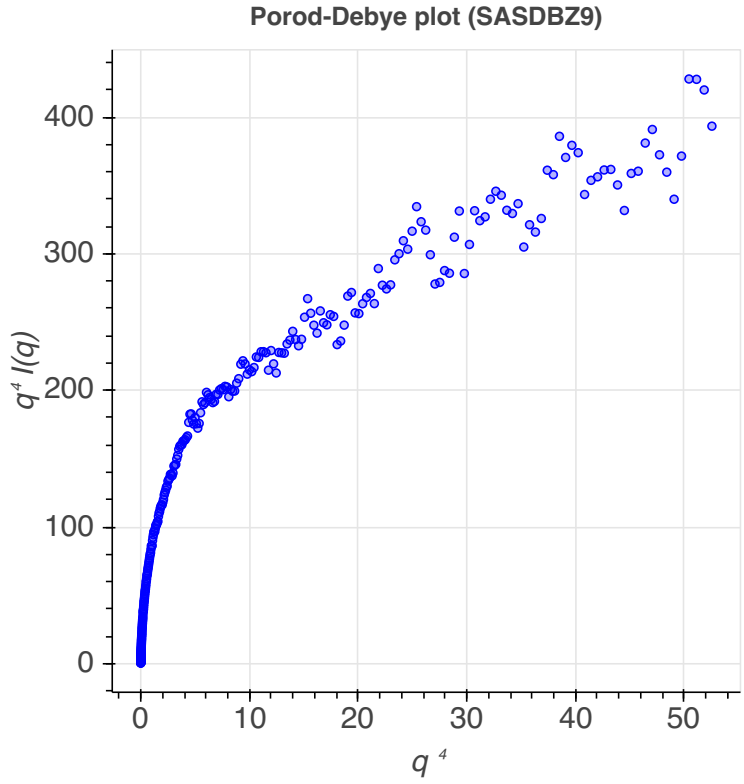


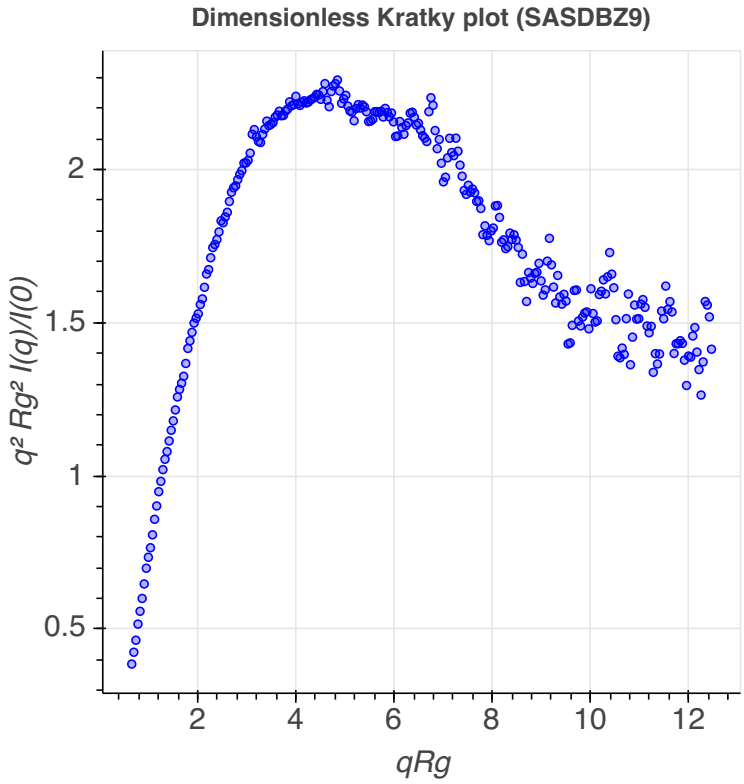
Flexibility analysis for SASDBY9: In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas, fully unfolded domains are devoid of any discernable plateau. For details, refer to Figure 5 in Rambo and Tainer, 2011 . In a Kratky plot, a parabolic shape is observed for globular (partial or fully folded) domains and a hyperbolic shape is observed for fully unfolded domains.





Flexibility analysis for SASDBZ9: In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas, fully unfolded domains are devoid of any discernable plateau. For details, refer to Figure 5 in Rambo and Tainer, 2011 . In a Kratky plot, a parabolic shape is observed for globular (partial or fully folded) domains and a hyperbolic shape is observed for fully unfolded domains.



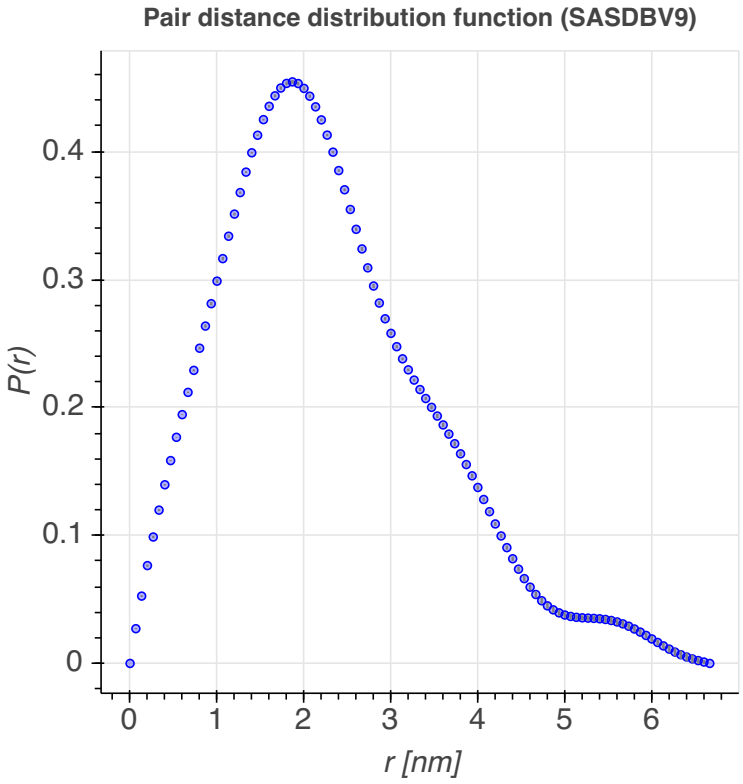


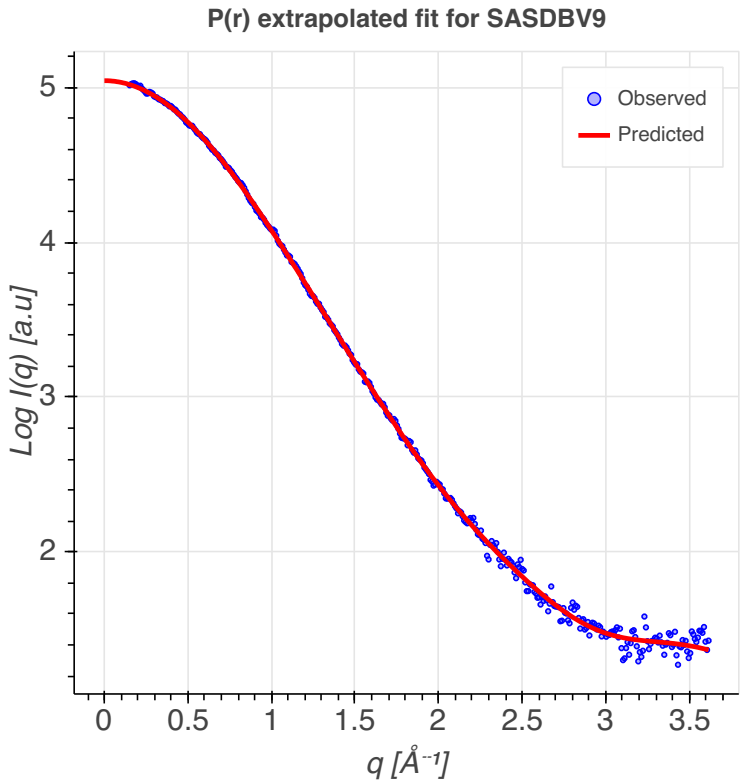
Pair-distance distribution analysis

P(r) analysis: p(r) represents the distribution of distances between all pairs of atoms within the particle weighted by the respective electron densities. p(r) is the Fourier transform of I(s) (and vice versa). Rg can be estimated from integrating the p(r) function. Agreement between the p(r) and Guinier-determined Rg (table below) is a good measure of the self-consistency of the SAS profile. Rg is a measure for the overall size of a macromolecule; e.g. a protein with a smaller Rg is more compact than a protein with a larger Rg, provided both have the same molecular weight (MW). The point where p(r) is decaying to zero is called Dmax and represents the maximum size of the particle.

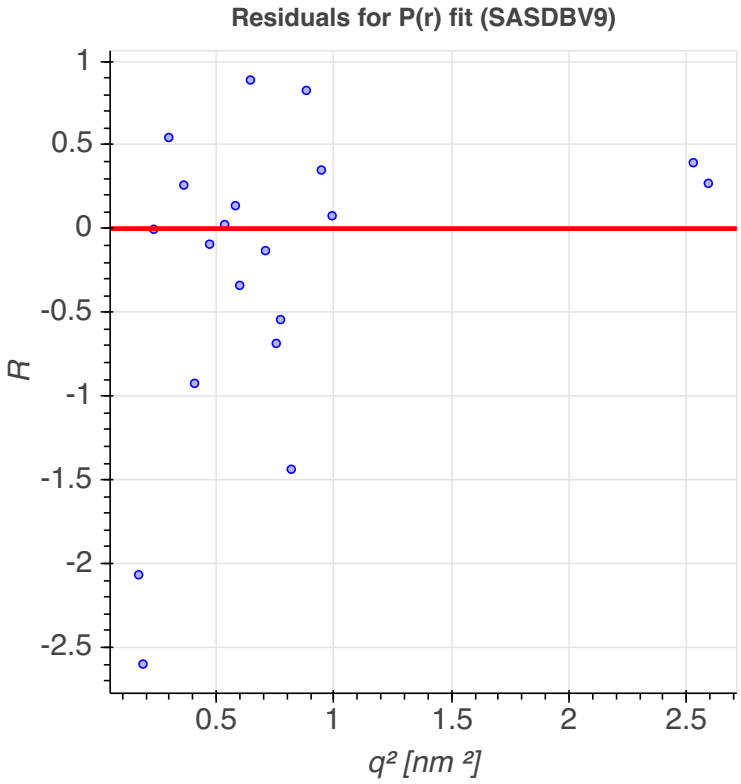
SASDB ID	Software used	Dmax	Dmax error	Rg	Rg error
SASDBV9	ATSAS GNOM	6.66 nm	N/A	1.824 nm	0.006 nm
SASDBW9	ATSAS GNOM	9.37 nm	N/A	2.787 nm	0.007 nm
SASDBX9	ATSAS GNOM	7.93 nm	N/A	2.636 nm	0.008 nm
SASDBY9	ATSAS GNOM	10.45 nm	N/A	2.976 nm	0.005 nm
SASDBZ9	ATSAS GNOM	15.43 nm	N/A	4.629 nm	0.011 nm

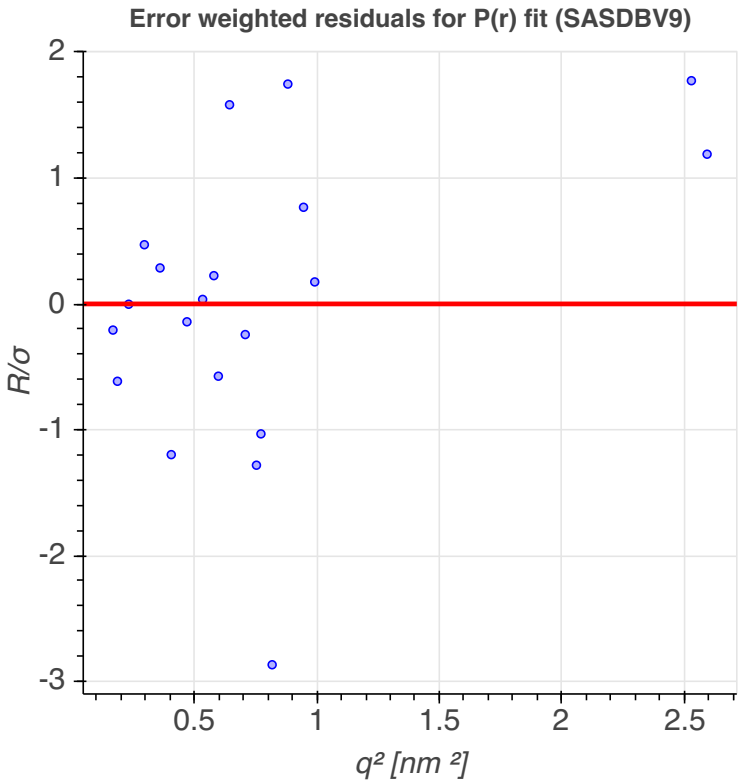
P(r) for SASDBV9: The value of P(r) should be zero beyond r=Dmax.



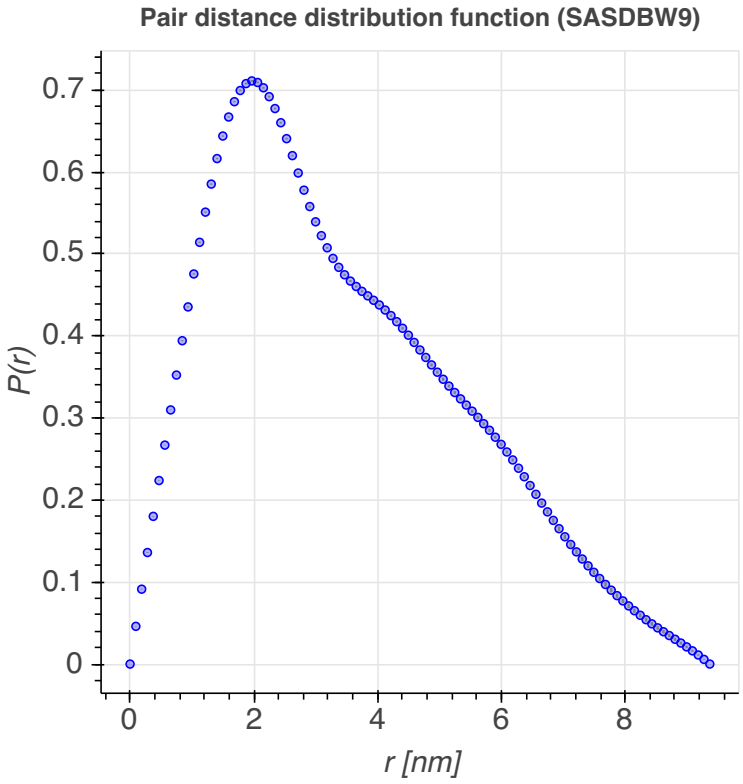


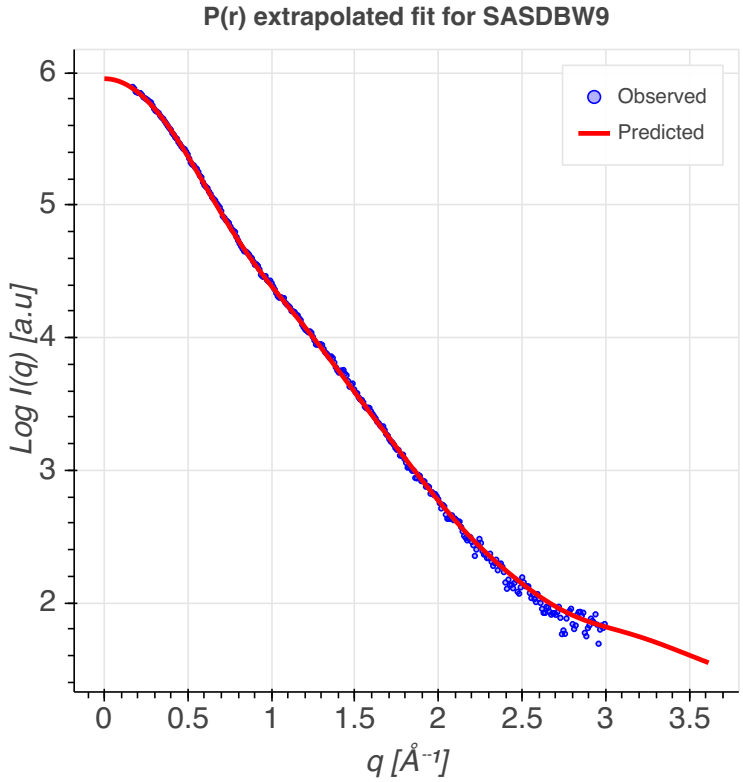
Residuals and error weighted residuals for P(r) analysis for SASDBV9: Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.



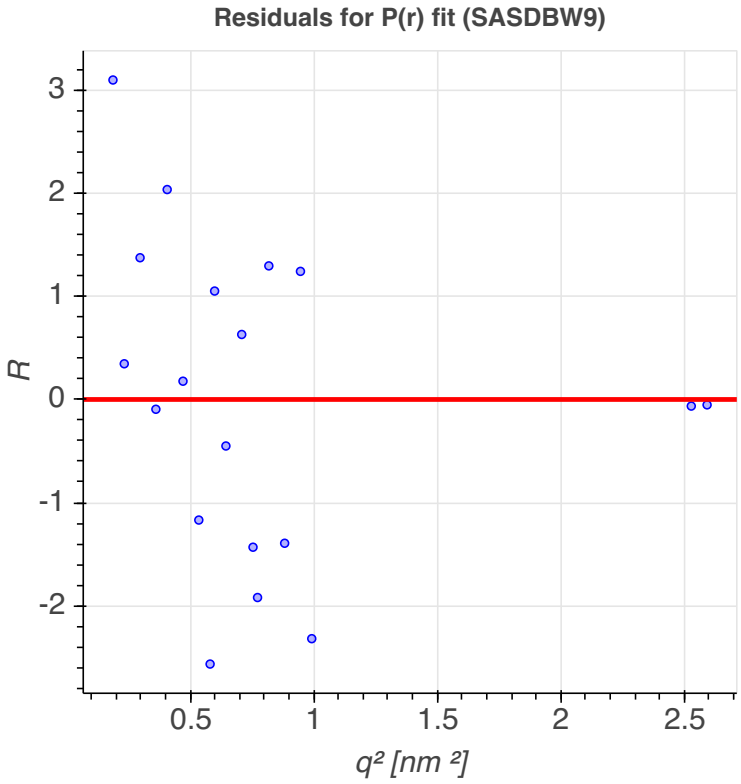


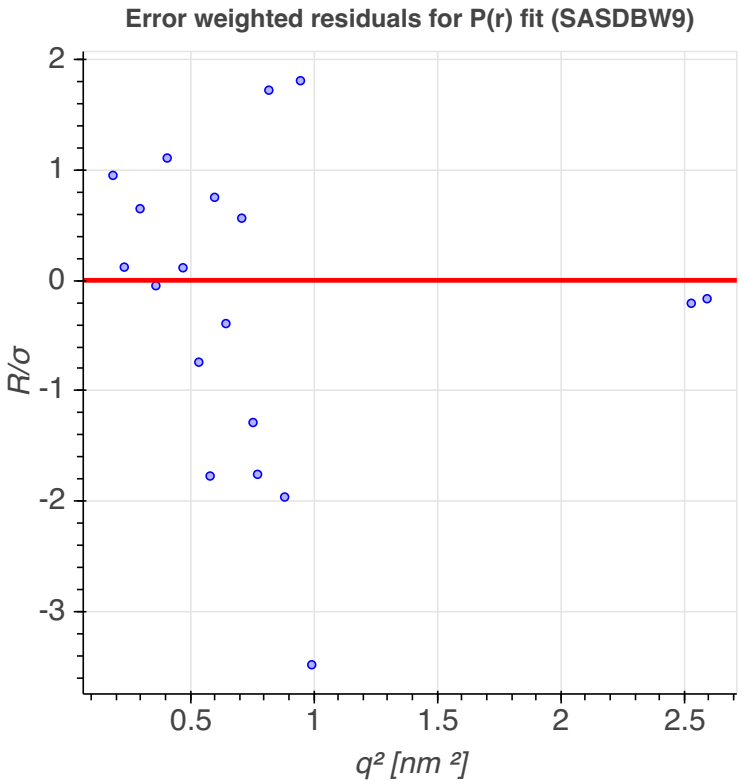
P(r) for SASDBW9: The value of P(r) should be zero beyond r=Dmax.



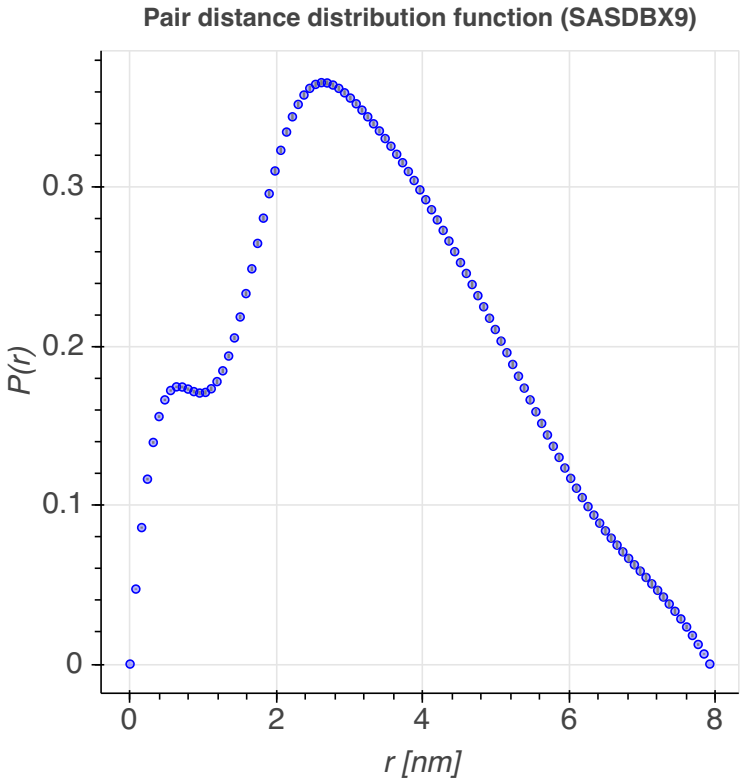


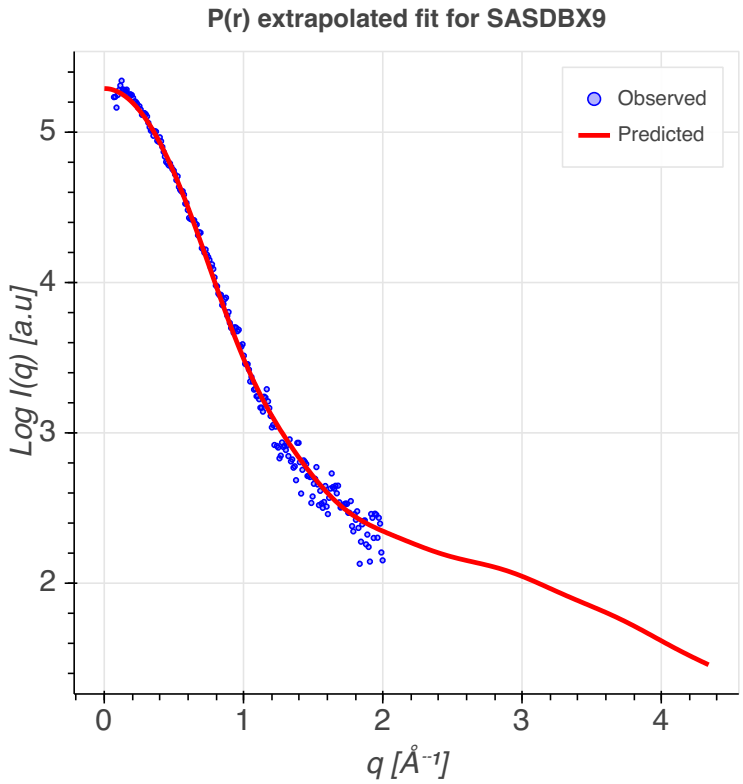
Residuals and error weighted residuals for P(r) analysis for SASDBW9: Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.



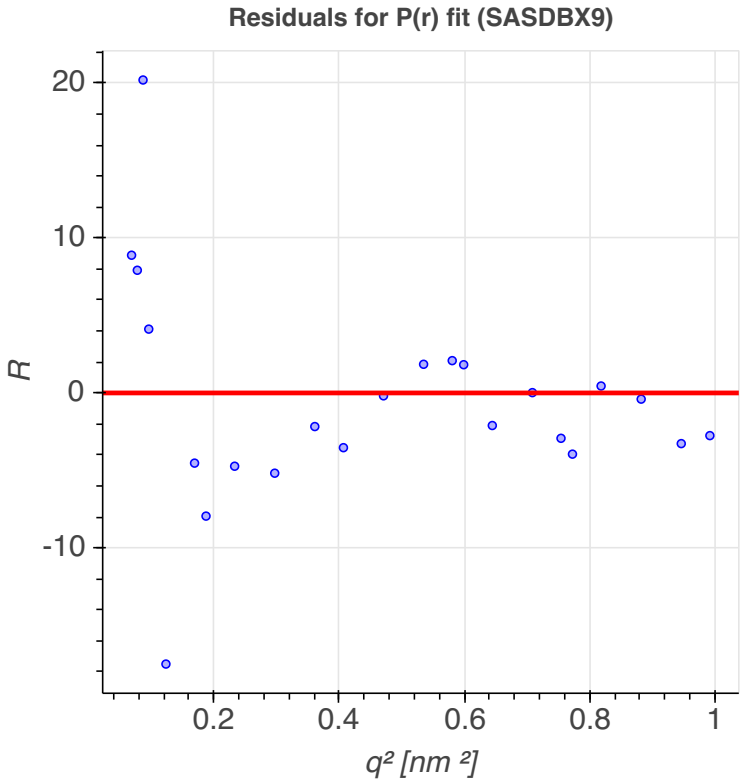


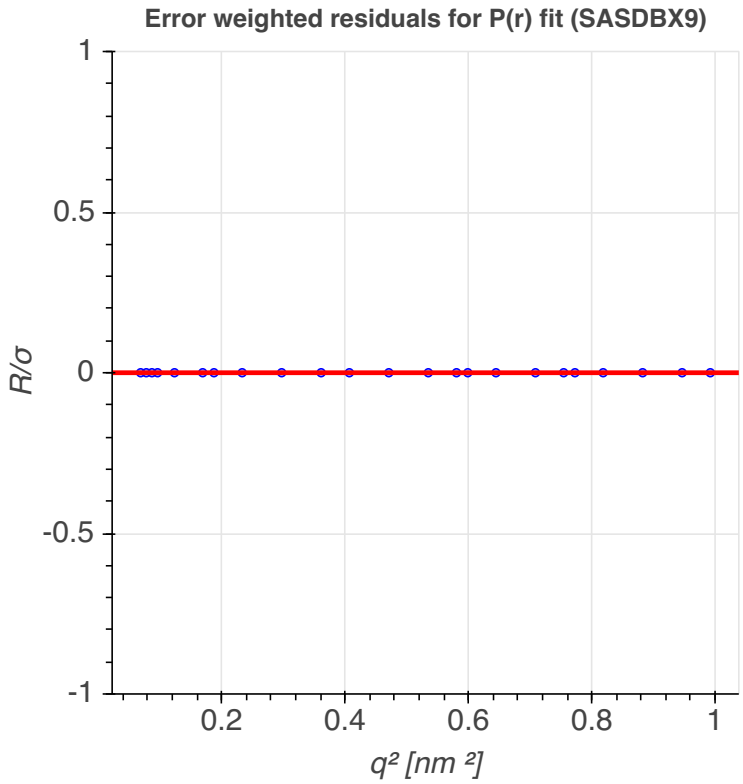
P(r) for SASDBX9: The value of P(r) should be zero beyond r=Dmax.



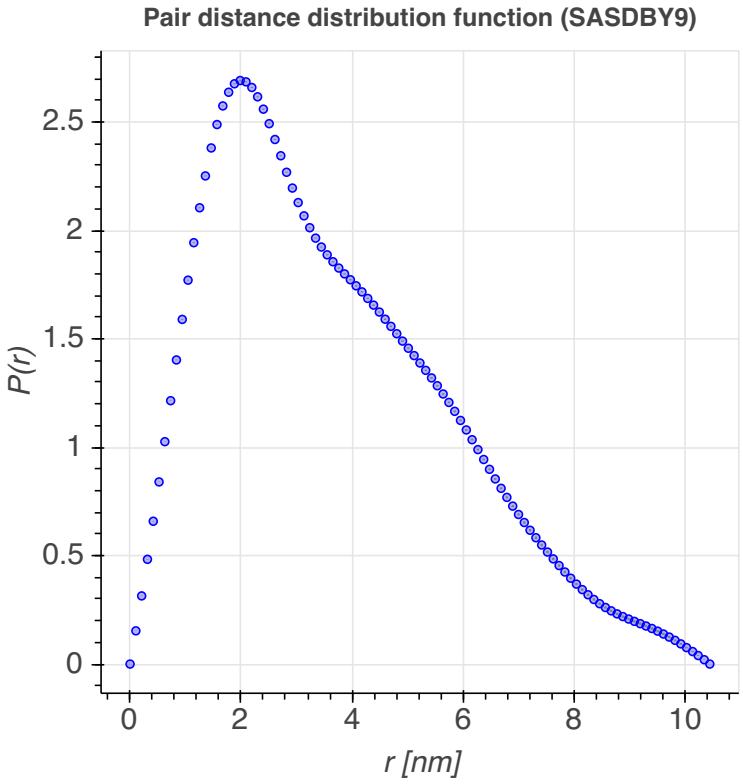


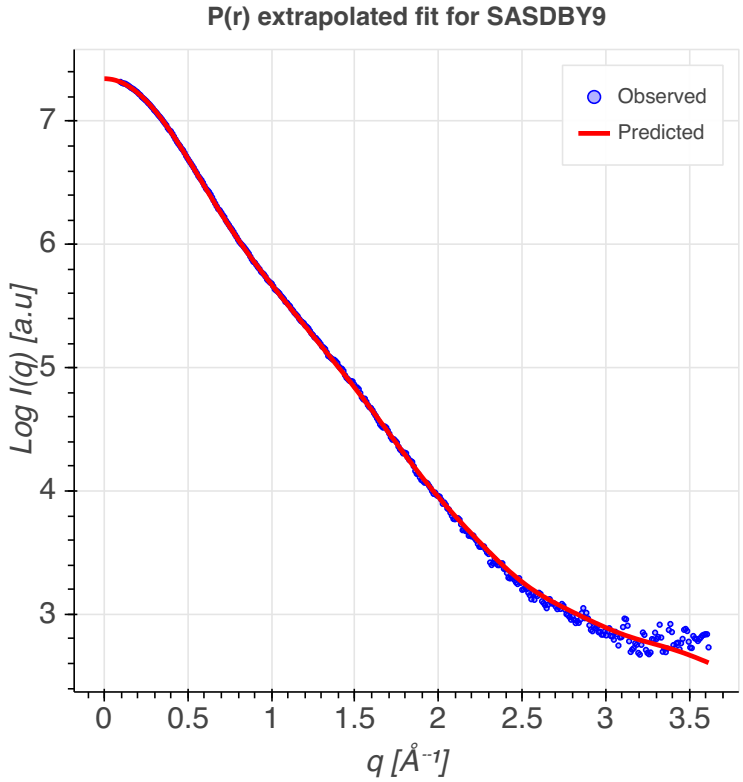
Residuals and error weighted residuals for P(r) analysis for SASDBX9: Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.





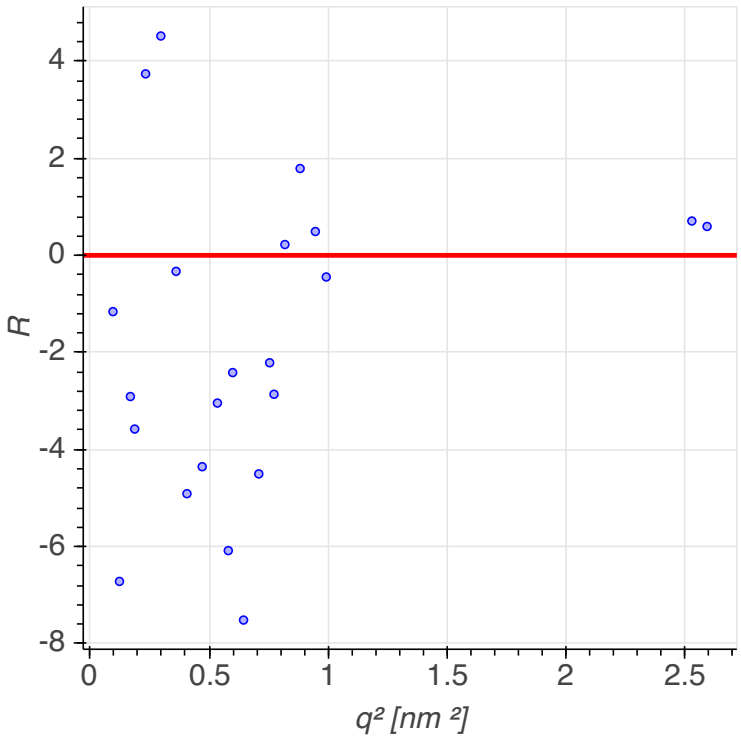
P(r) for SASDBY9: The value of P(r) should be zero beyond r=Dmax.



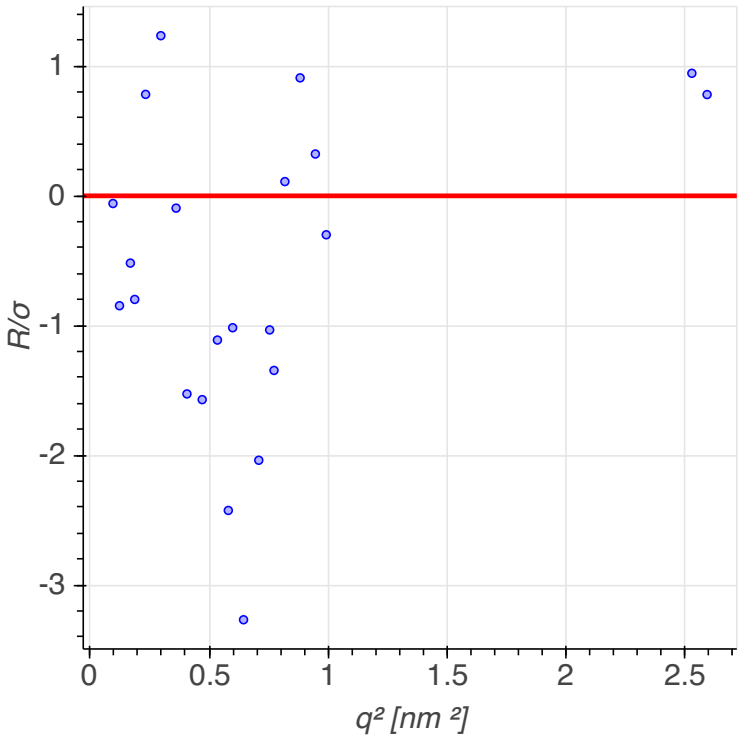


Residuals and error weighted residuals for P(r) analysis for SASDBY9: Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.

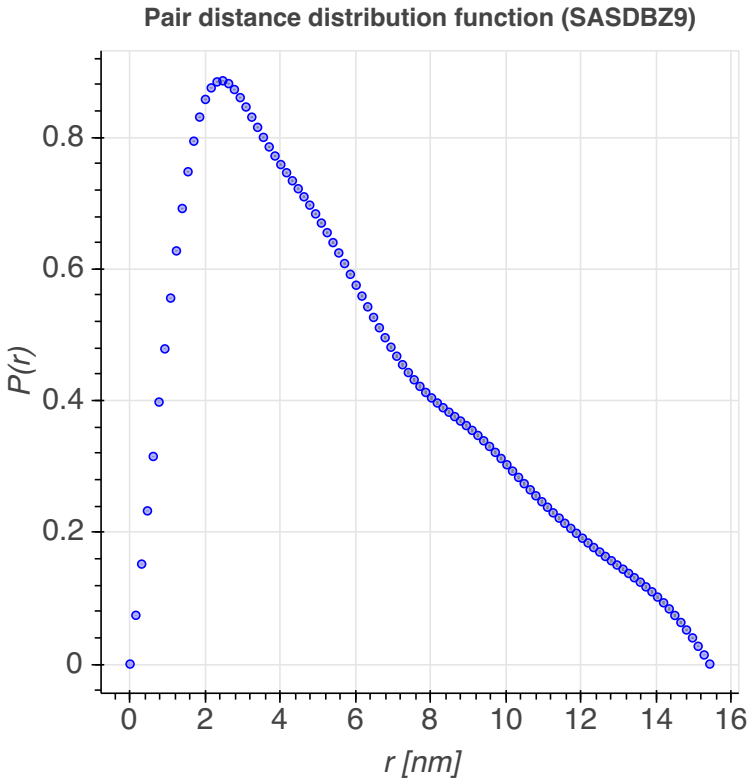
Residuals for P(r) fit (SASDBY9)

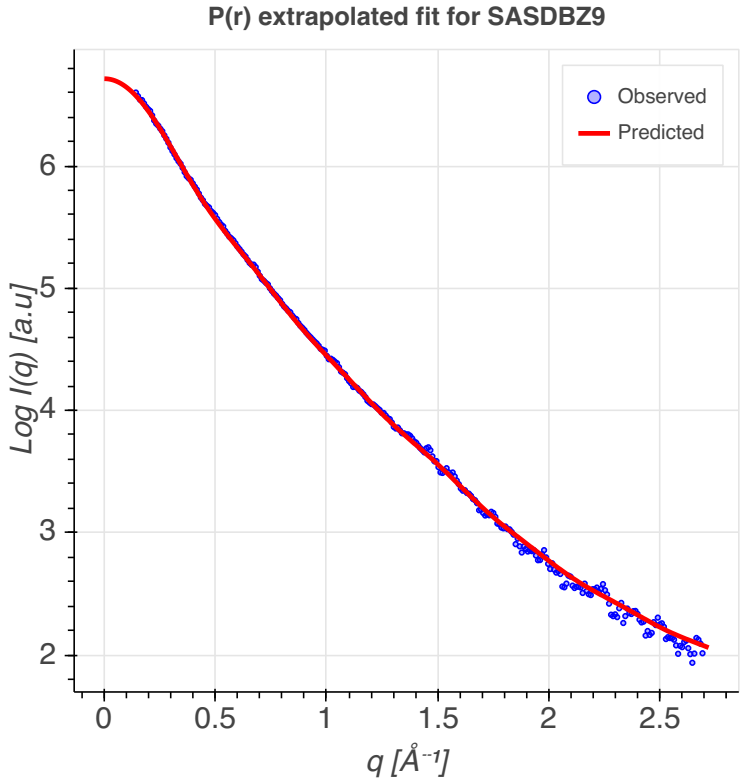


Error weighted residuals for P(r) fit (SASDBY9)

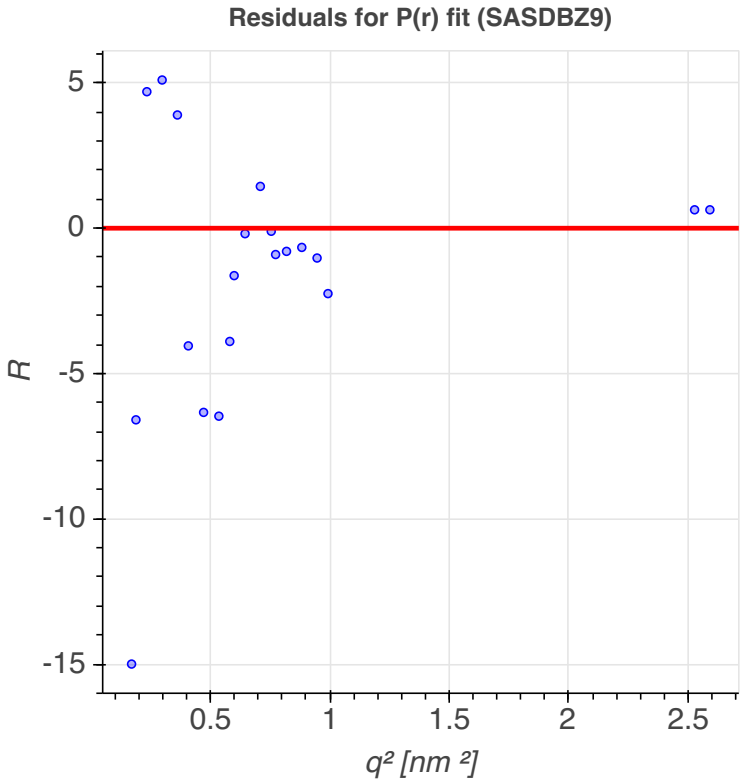


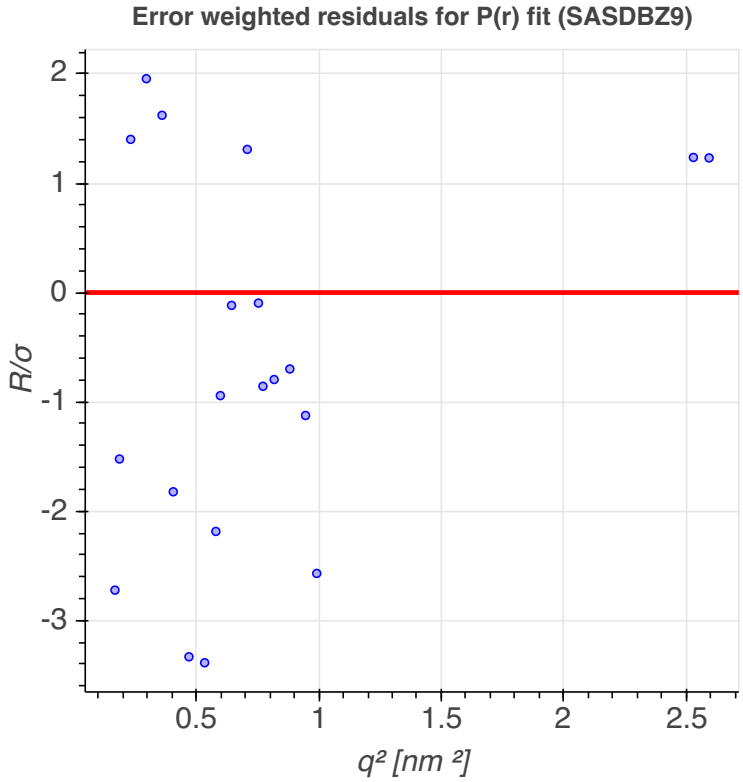
P(r) for SASDBZ9: The value of P(r) should be zero beyond r=Dmax.





Residuals and error weighted residuals for P(r) analysis for SASDBZ9: Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.



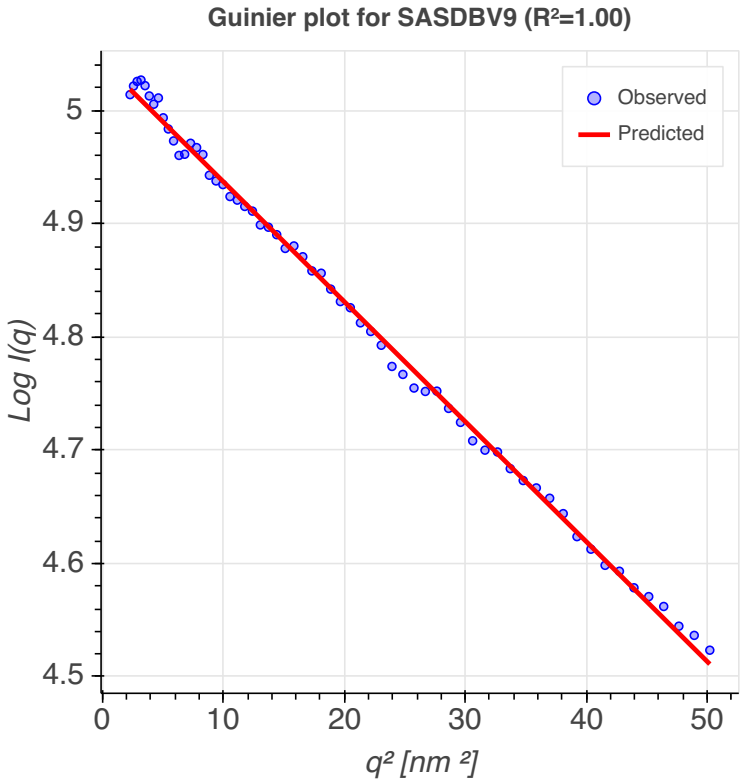


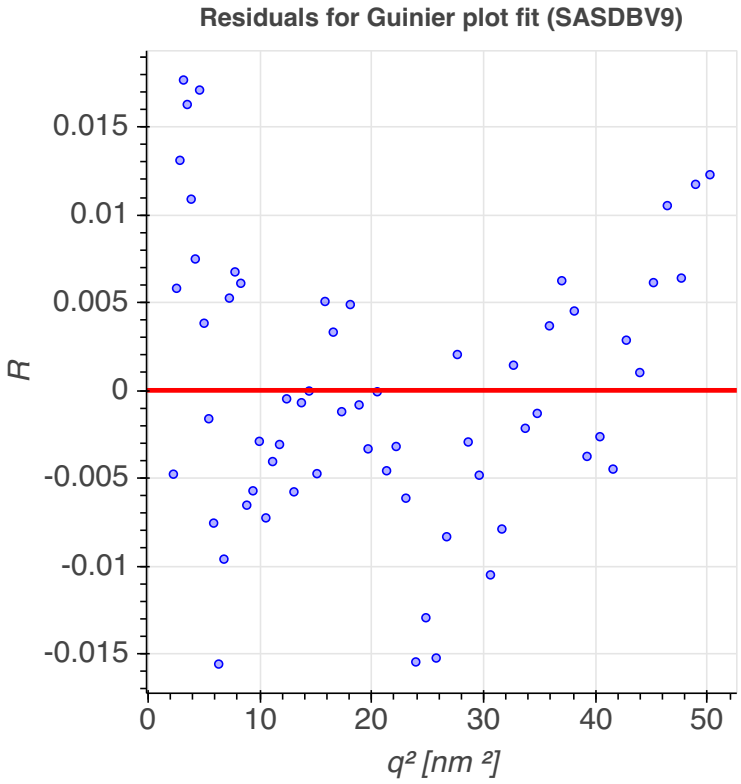
Guinier analysis

Guinier analysis: agreement between the p(r) and Guinier-determined Rg (table below) is a good measure of the self-consistency of the SAS profile. Molecular weight estimates can also be compared to Porod and sample molecular weights for consistency.

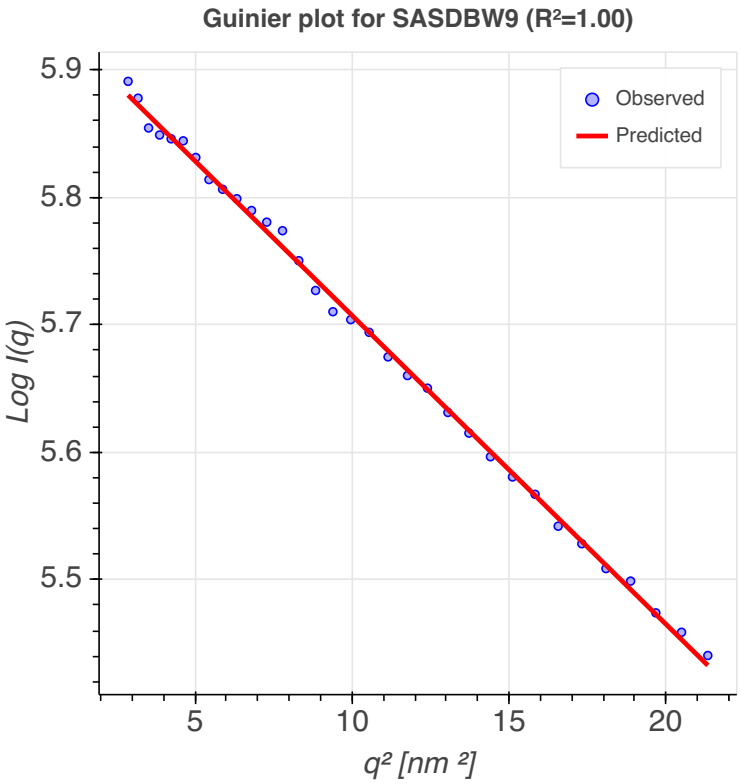
SASDB ID	Rg	Rg error	MW	MW error
SASDBV9	1.77 nm	0.048 nm	12.200 nm	N/A
SASDBW9	2.71 nm	0.063 nm	25.200 nm	N/A
SASDBX9	2.78 nm	0.180 nm	14.700 nm	N/A
SASDBY9	2.95 nm	0.110 nm	25.200 nm	N/A
SASDBZ9	4.34 nm	0.170 nm	48.300 nm	N/A

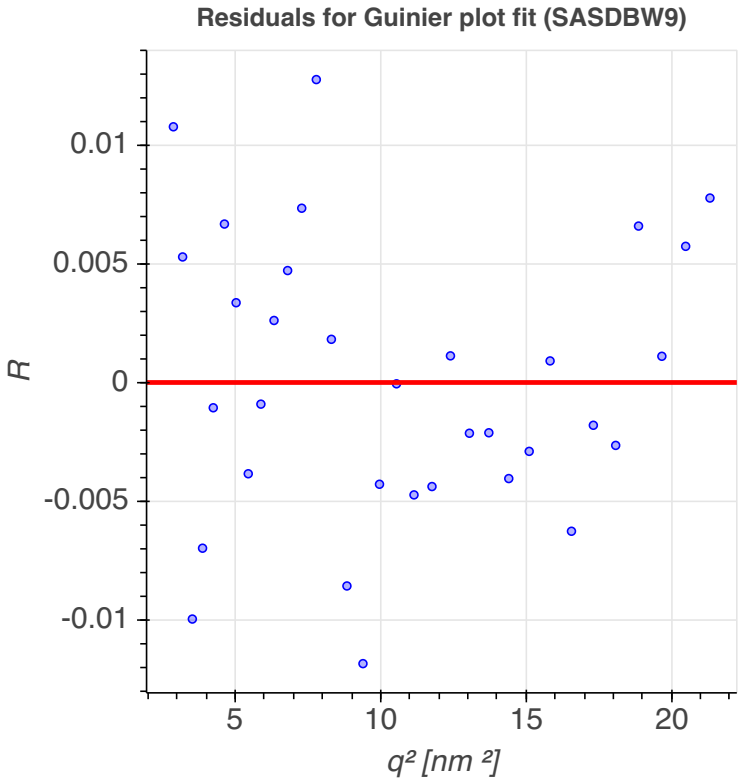
Guinier analysis for SASDBV9: the linearity of the Guinier plot is a sensitive indicator of the quality of the experimental SAS data; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the samesize. Deviations from linearity usually point to strong interference effects, polydispersity of the samples or improper background subtraction. Residual value plot and coefficient of determination (R^2) are measures to assess linear fit to the data. A perfect fit has an R^2 value of 1. Residual values should be equally and randomly spaced around the horizontal axis.



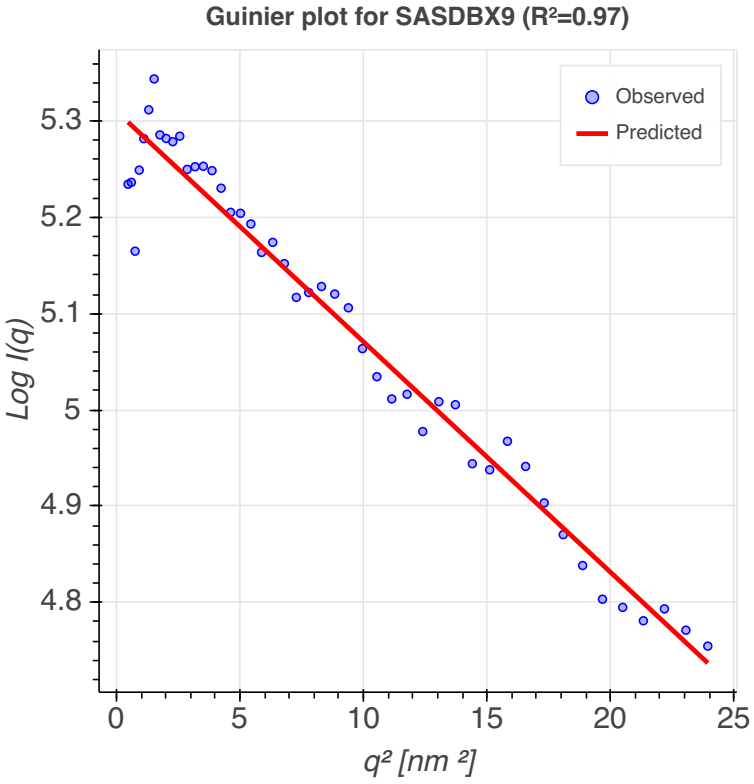


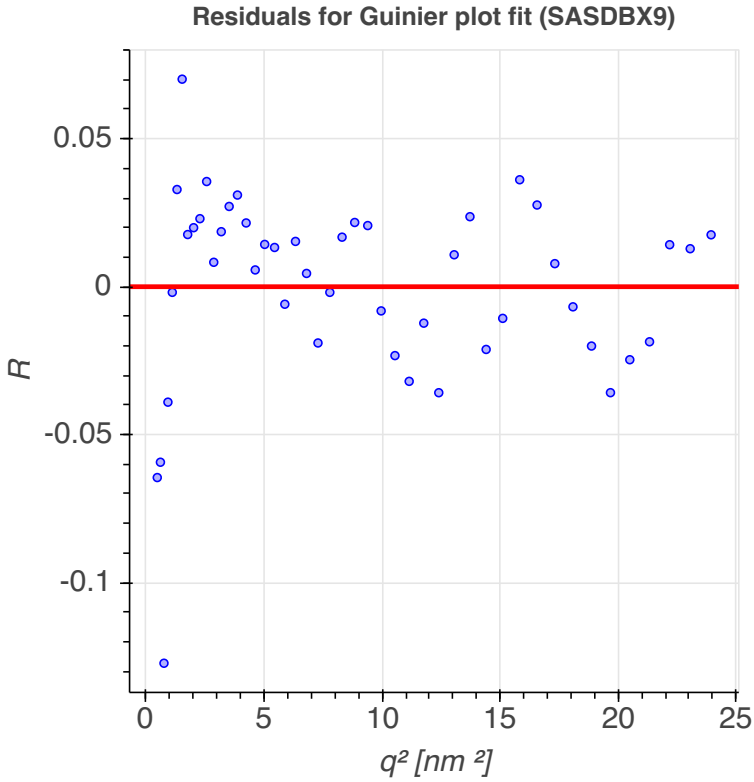
Guinier analysis for SASDBW9: the linearity of the Guinier plot is a sensitive indicator of the quality of the experimental SAS data; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the samesize. Deviations from linearity usually point to strong interference effects, polydispersity of the samples or improper background subtraction. Residual value plot and coefficient of determination (R^2) are measures to assess linear fit to the data. A perfect fit has an R^2 value of 1. Residual values should be equally and randomly spaced around the horizontal axis.



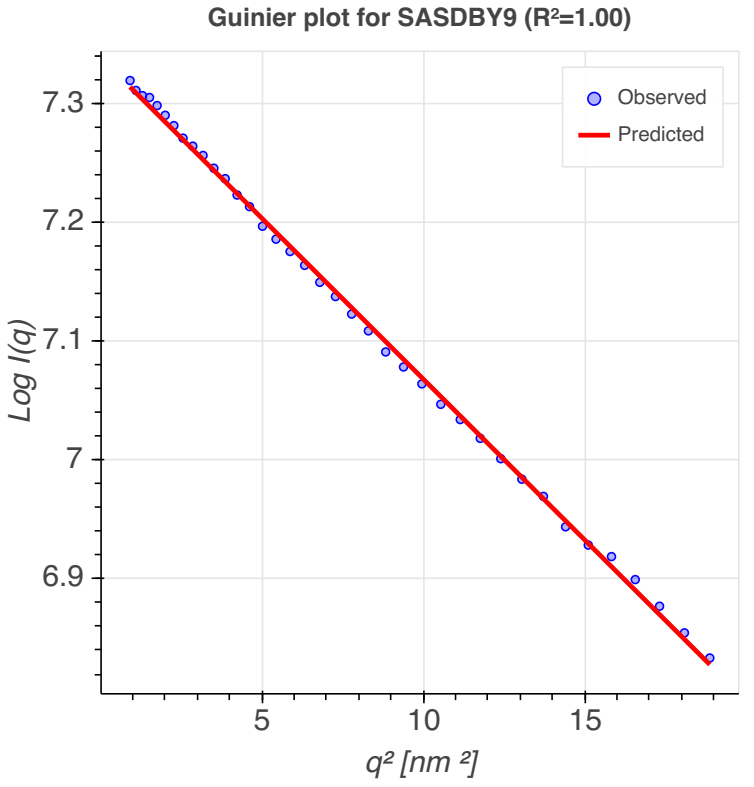


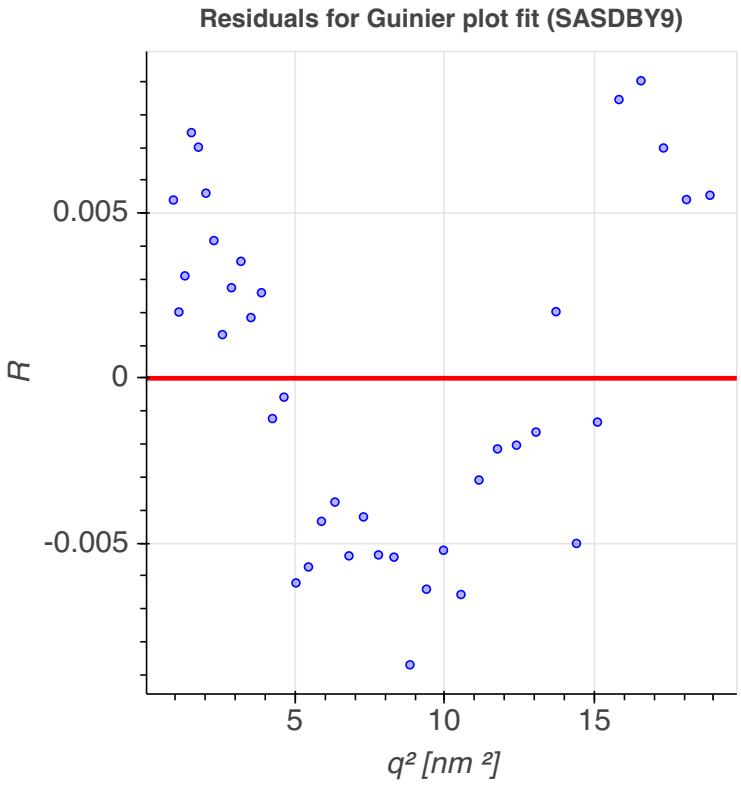
Guinier analysis for SASDBX9: the linearity of the Guinier plot is a sensitive indicator of the quality of the experimental SAS data; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the samesize. Deviations from linearity usually point to strong interference effects, polydispersity of the samples or improper background subtraction. Residual value plot and coefficient of determination (R^2) are measures to assess linear fit to the data. A perfect fit has an R^2 value of 1. Residual values should be equally and randomly spaced around the horizontal axis.



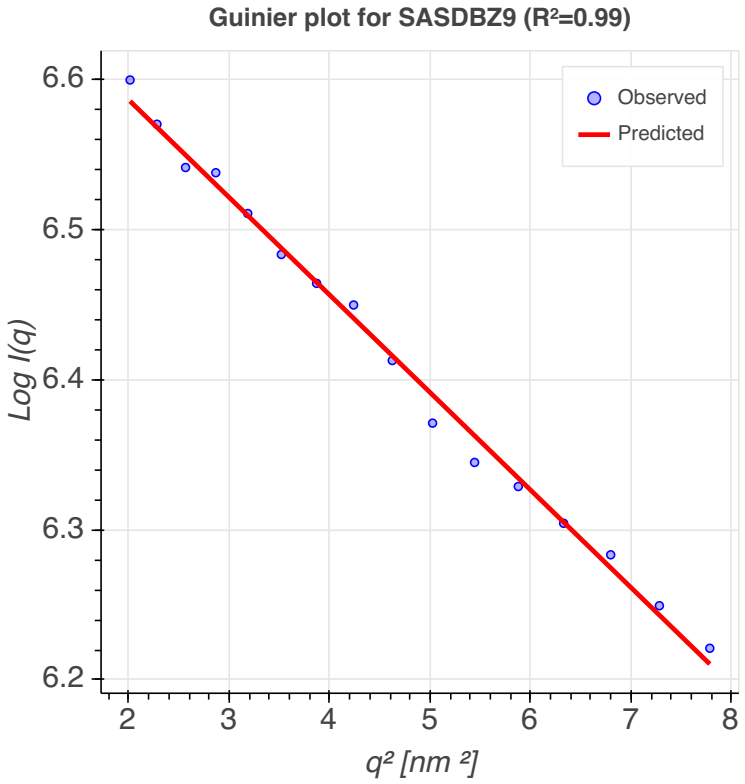


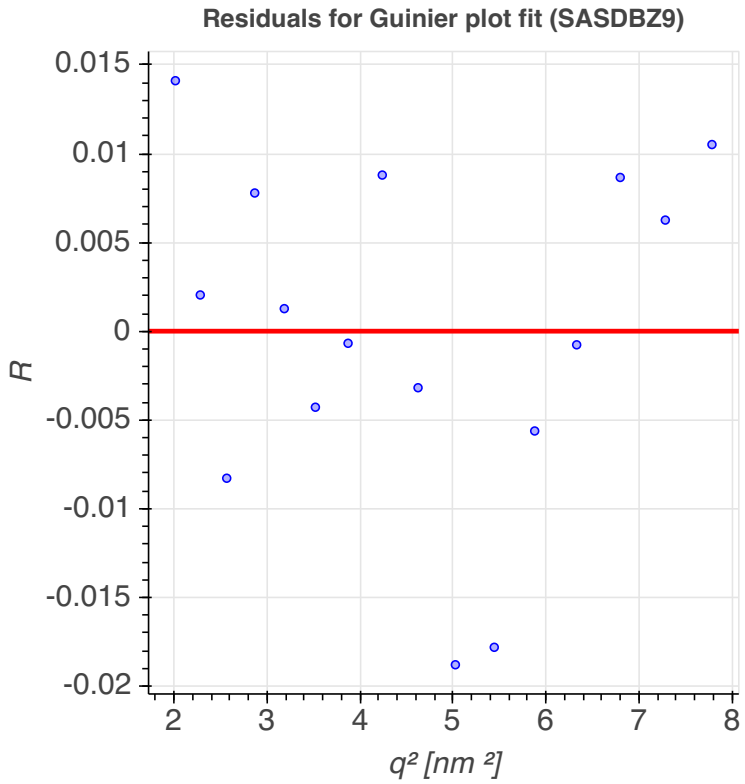
Guinier analysis for SASDBY9: the linearity of the Guinier plot is a sensitive indicator of the quality of the experimental SAS data; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the samesize. Deviations from linearity usually point to strong interference effects, polydispersity of the samples or improper background subtraction. Residual value plot and coefficient of determination (R^2) are measures to assess linear fit to the data. A perfect fit has an R^2 value of 1. Residual values should be equally and randomly spaced around the horizontal axis.





Guinier analysis for SASDBZ9: the linearity of the Guinier plot is a sensitive indicator of the quality of the experimental SAS data; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the samesize. Deviations from linearity usually point to strong interference effects, polydispersity of the samples or improper background subtraction. Residual value plot and coefficient of determination (R^2) are measures to assess linear fit to the data. A perfect fit has an R^2 value of 1. Residual values should be equally and randomly spaced around the horizontal axis.





Model quality

Excluded volume satisfaction

Excluded volume satisfaction for the models in the entry are listed below.

Models	Excluded Volume Satisfaction	Number of violations
1	99.46	2089.0

Fit of model to data used for modeling

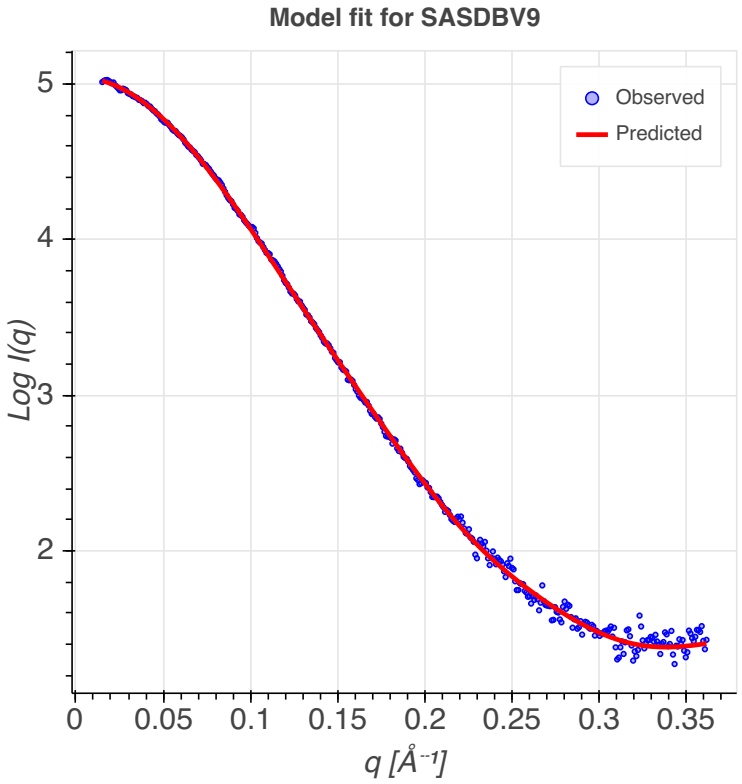
Fit of model(s) to SAS data

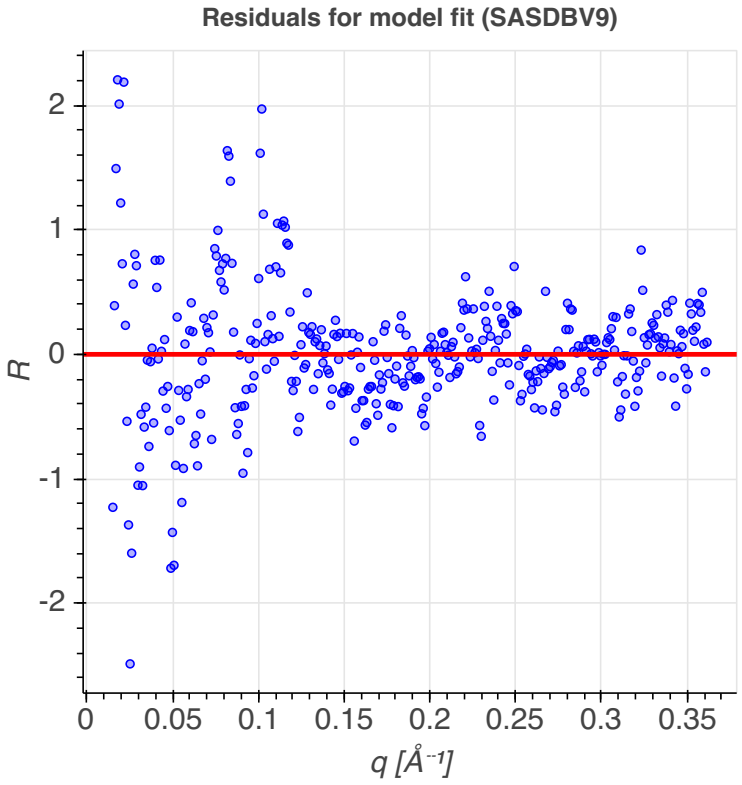
χ^2 goodness of fit analysis

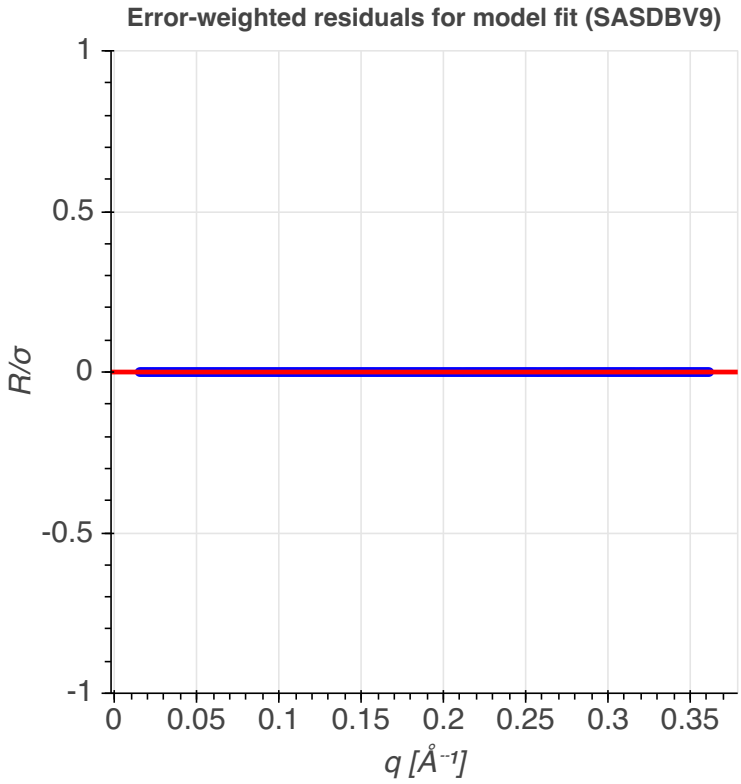
Model and fits displayed below were obtained from SASBDB. χ^2 values are a measure of fit of the model to data. A perfect fit has a χ^2 value of zero.

SASDB ID	Model	χ^2
SASDBV9	1	1.13
SASDBV9	2	1.1
SASDBW9	1	1.97
SASDBX9	1	2.86
SASDBY9	1	2.02
SASDBZ9	1	1.94

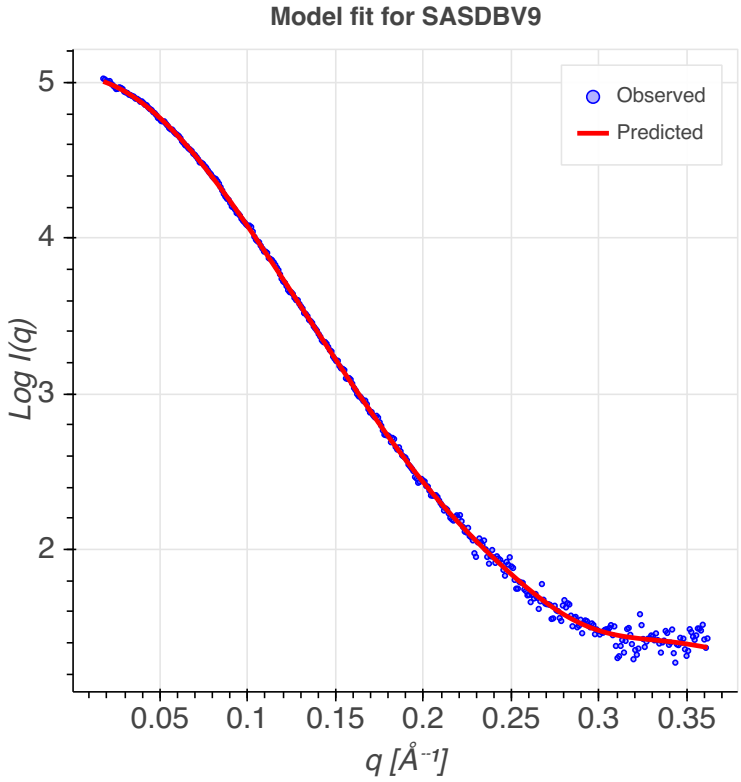
Model fit for SASDBV9 (fit/model number 1): Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.

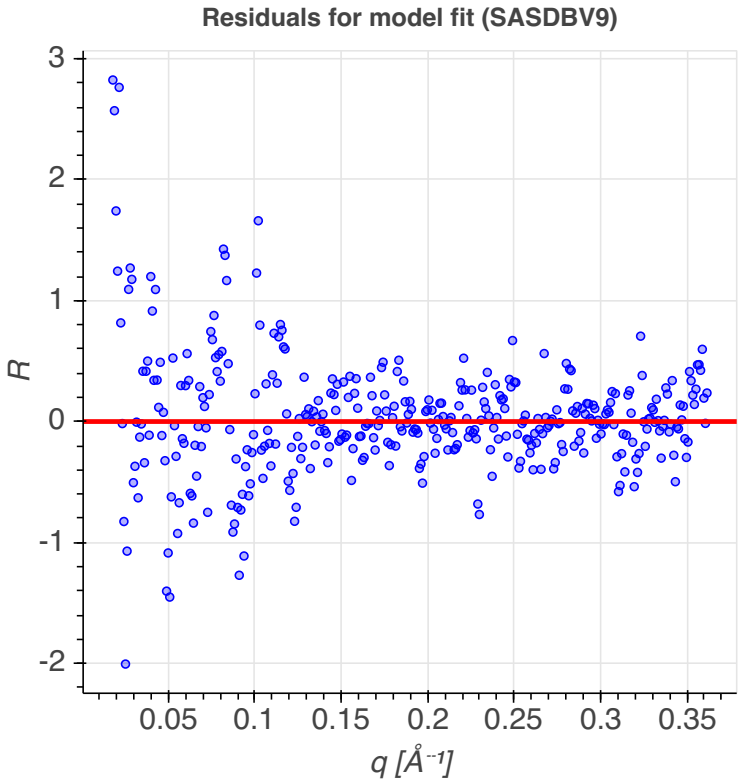

Model fit for SASDBV9
Observed
Predicted
Log I(q)
q [Å⁻¹]
5
4
3
2
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35

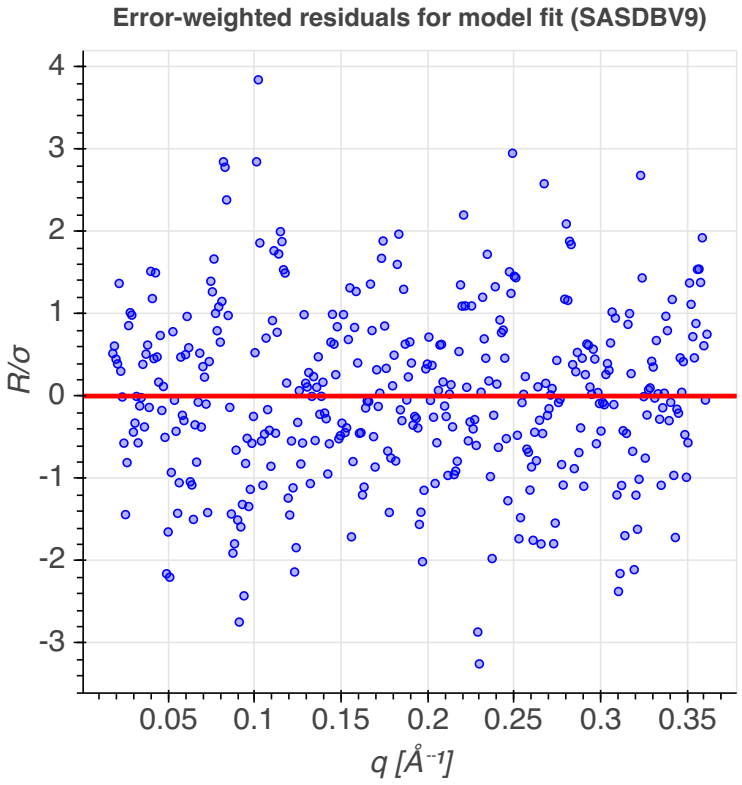

Residuals for model fit (SASDBV9)
R
q [Å⁻¹]
2
1
0
-1
-2
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35



Model fit for SASDBV9 (fit/model number 2): Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.

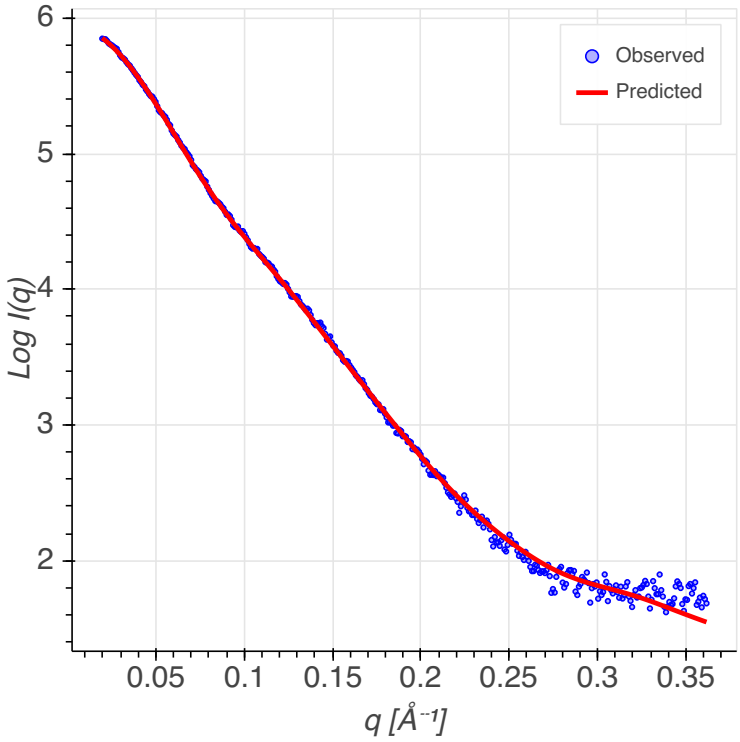




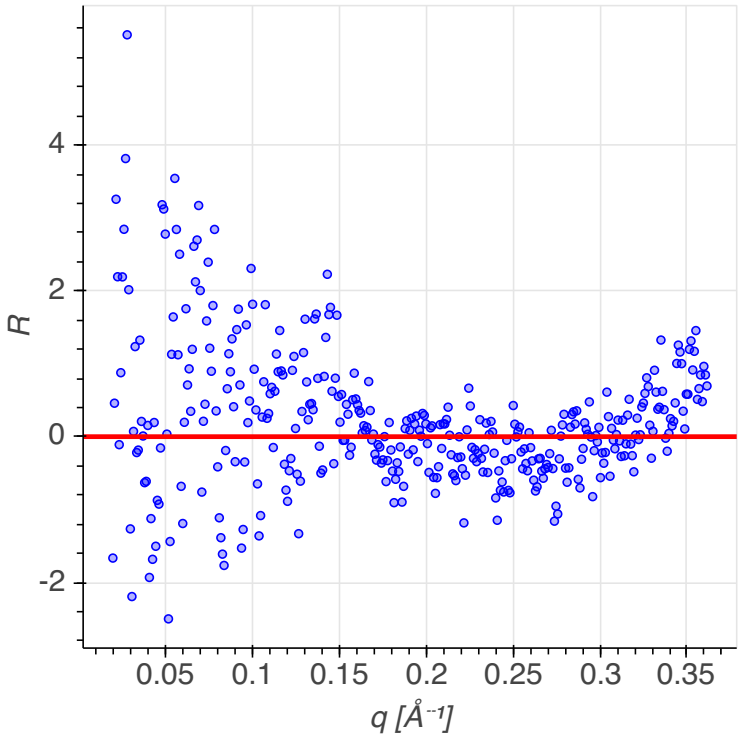


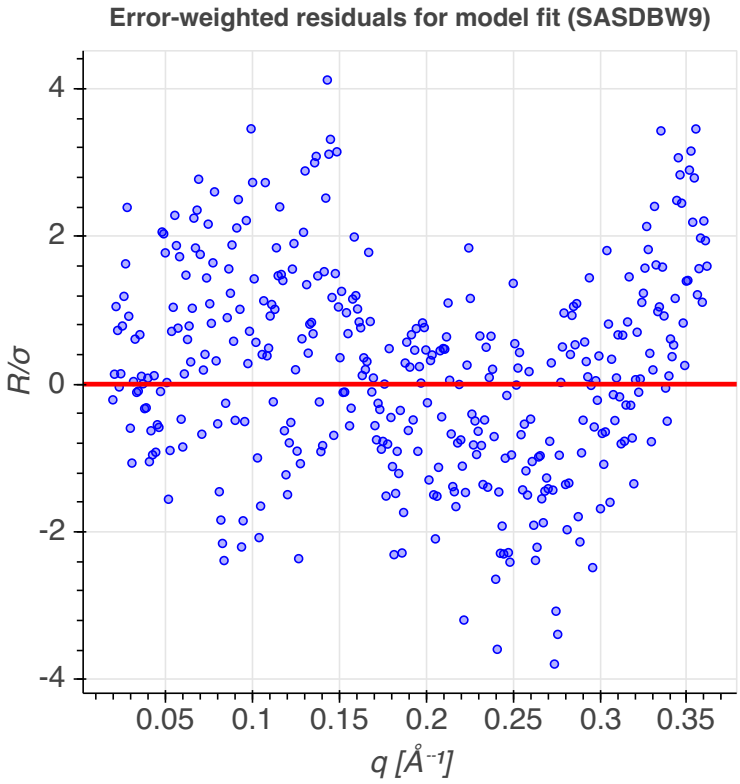
Model fit for SASDBW9 (fit/model number 1): Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.

Model fit for SASDBW9

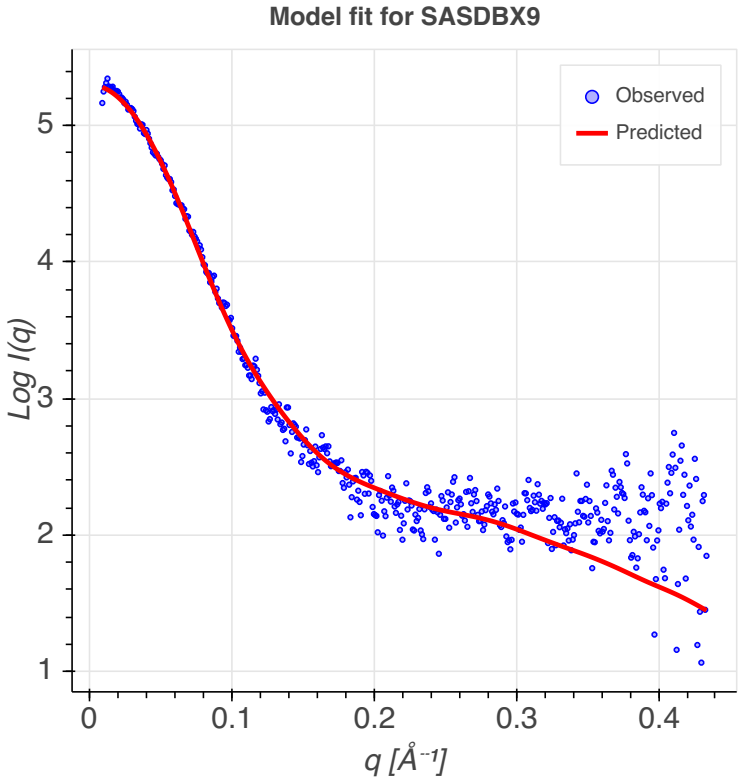


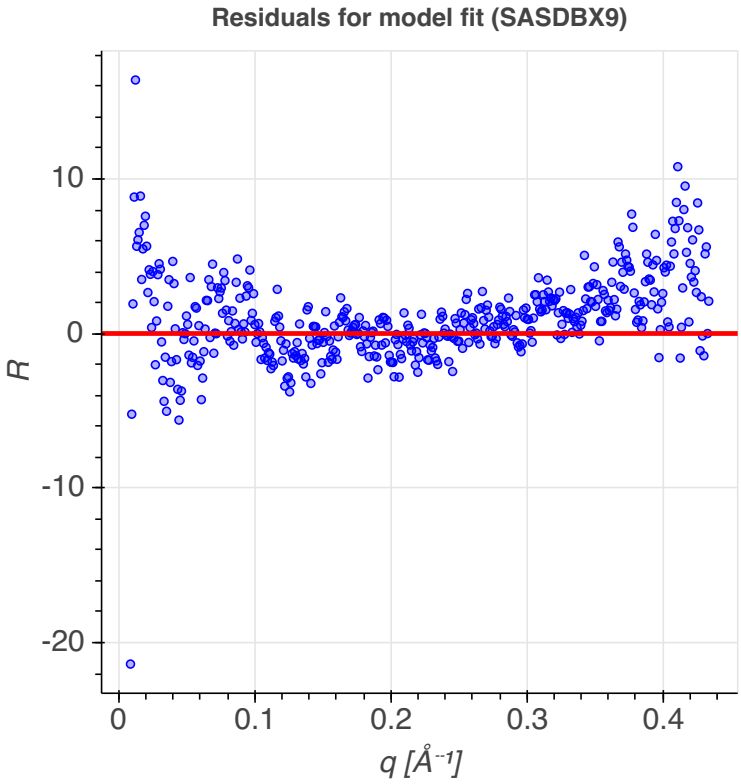
Residuals for model fit (SASDBW9)

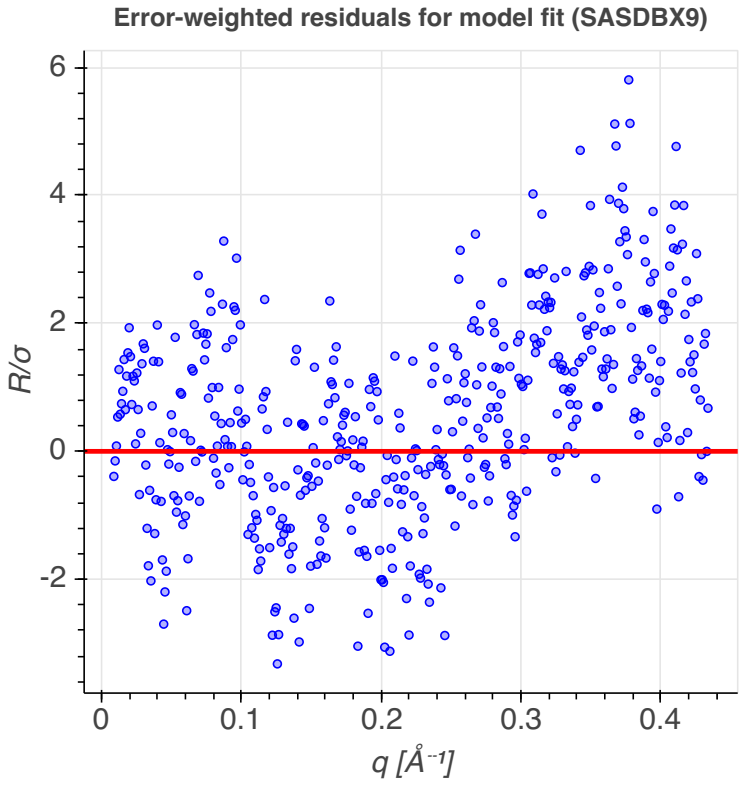




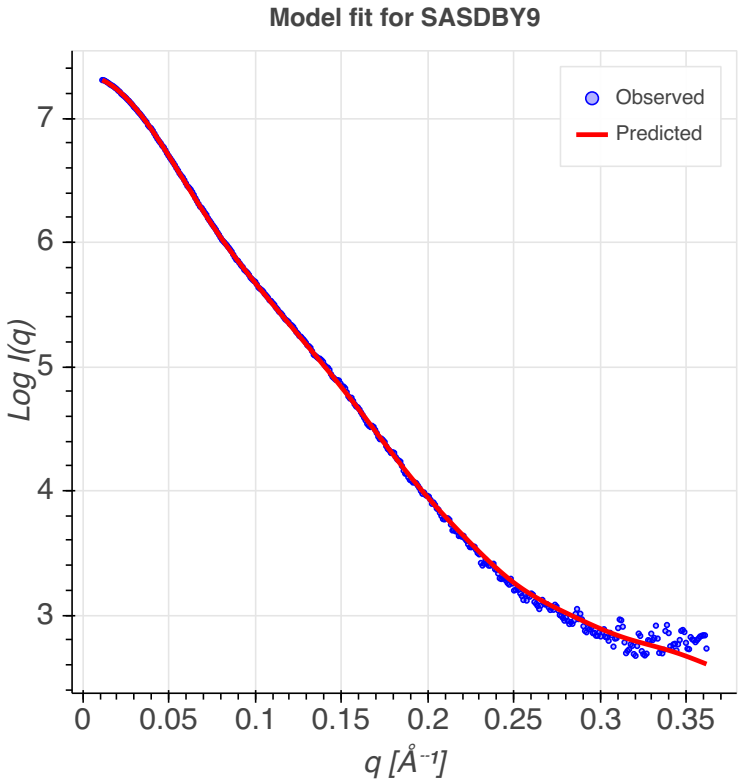
Model fit for SASDBX9 (fit/model number 1): Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.

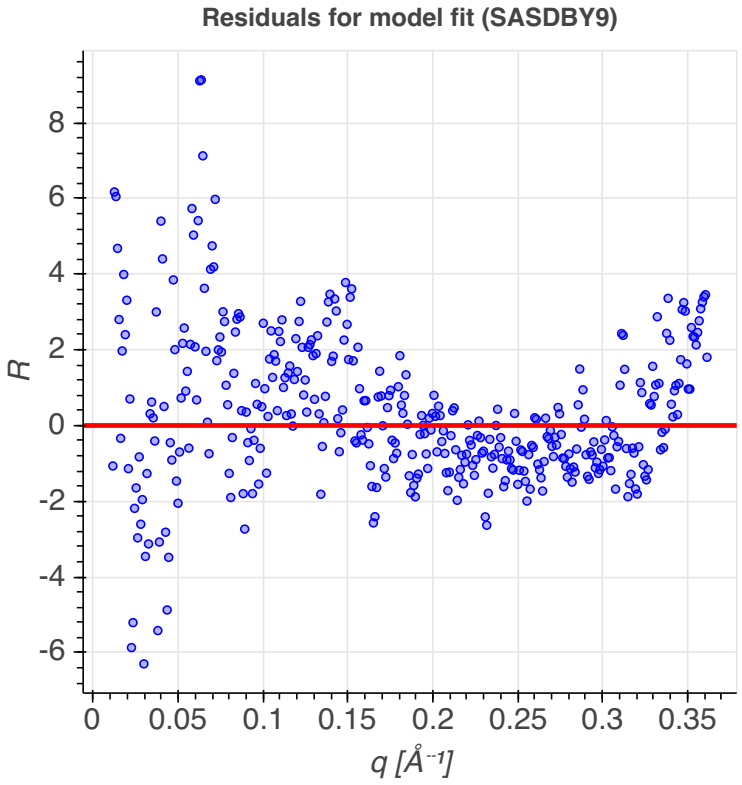


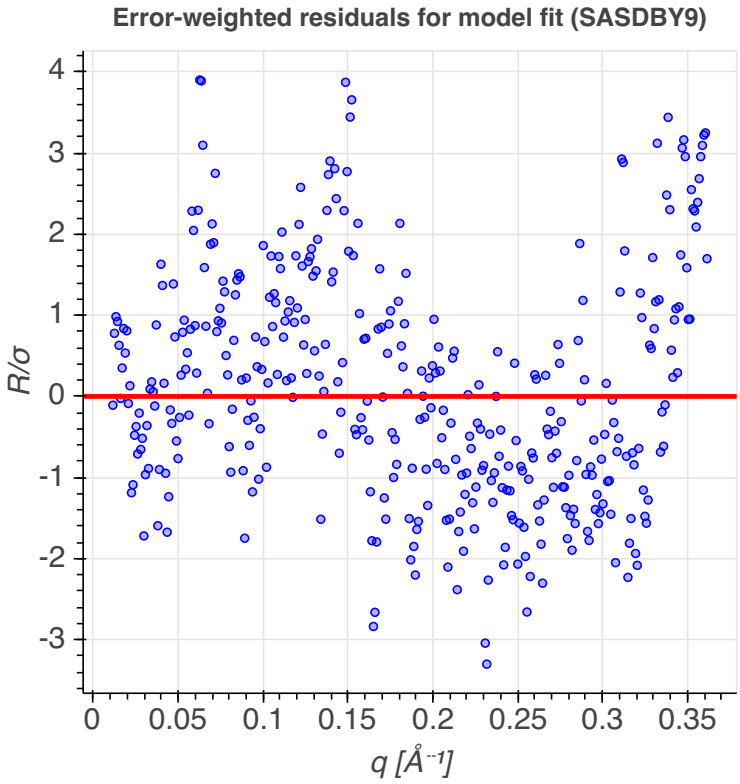




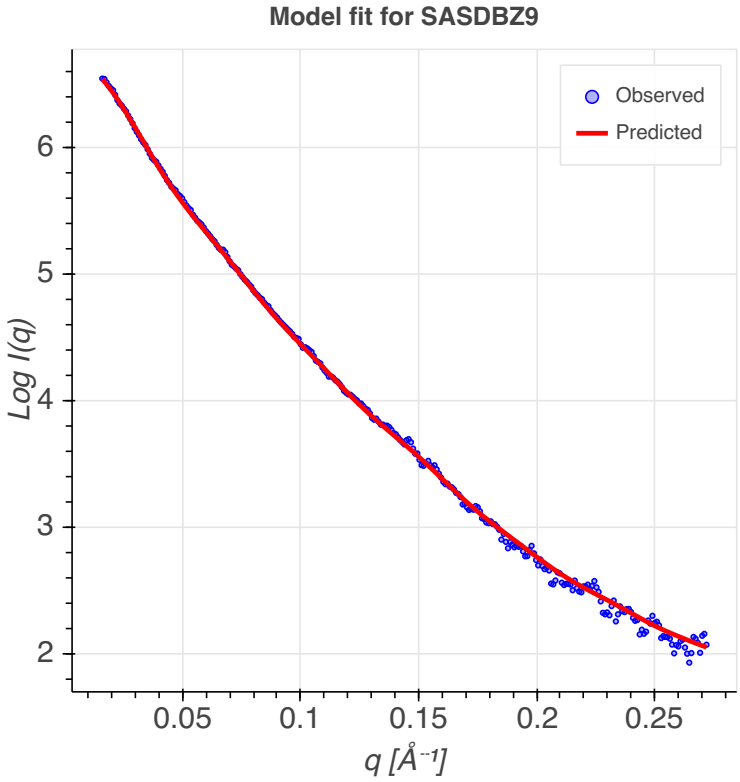
Model fit for SASDBY9 (fit/model number 1): Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.

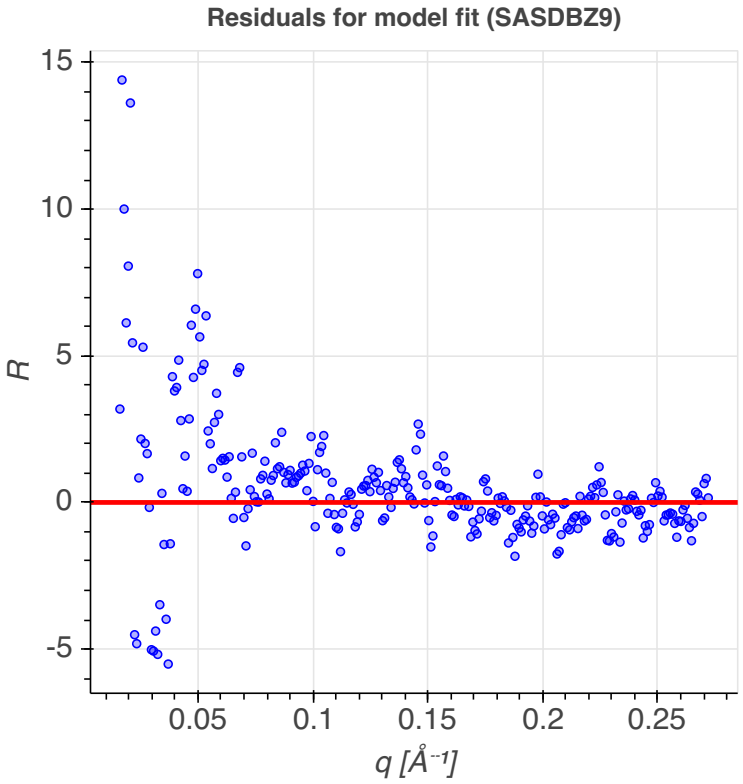

Model fit for SASDBY9
Observed
Predicted
Log I(q)
7
6
5
4
3
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
q [Å⁻¹]

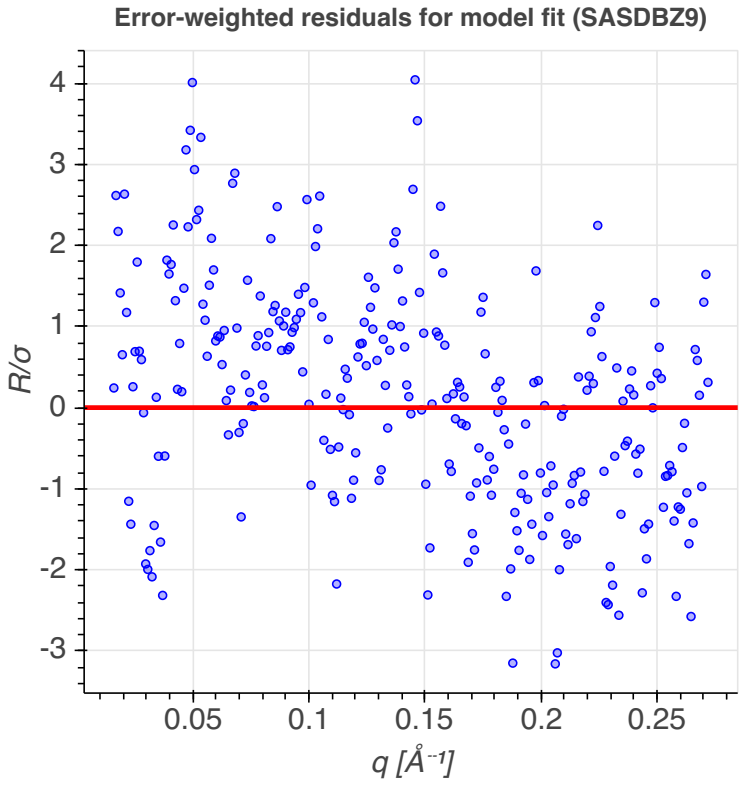

Residuals for model fit (SASDBY9)
R
8
6
4
2
0
-2
-4
-6
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
q [Å⁻¹]



Model fit for SASDBZ9 (fit/model number 1): Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.







Cormap p-value analysis of fits

ATSAS datcmp was used for hypothesis testing.All data sets are similar (i.e. the fit and the data collected) is the null hypothesis. p-value is a measure of evidence against the null

hypothesis, smaller the value, the stronger the evidence that you should reject the null hypothesis.

SASDB ID	Model	p-value
SASDBV9	1	2.22E-02
SASDBV9	2	5.53E-03
SASDBW9	1	1.00E-05
SASDBX9	1	0.00E+00
SASDBY9	1	3.00E-06
SASDBZ9	1	0.00E+00

Fit of model to data not used for modeling

Uncertainty of data and model