WORLDWIDE
wwPDB
PROTEIN DATA BANK

Full wwPDB Integrative Structure Validation Report

June 29, 2020 -- 12:29 PM

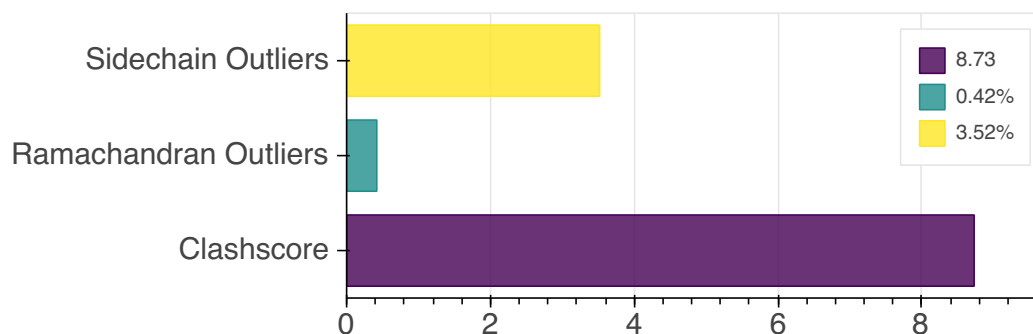**The following software were used in the production of this report:**
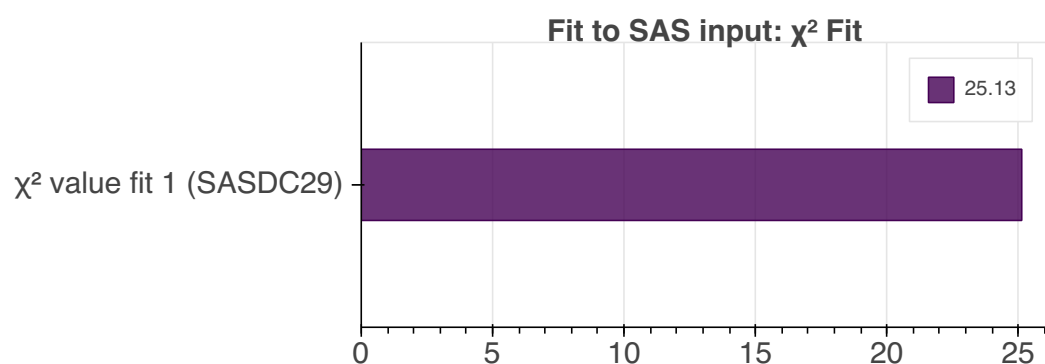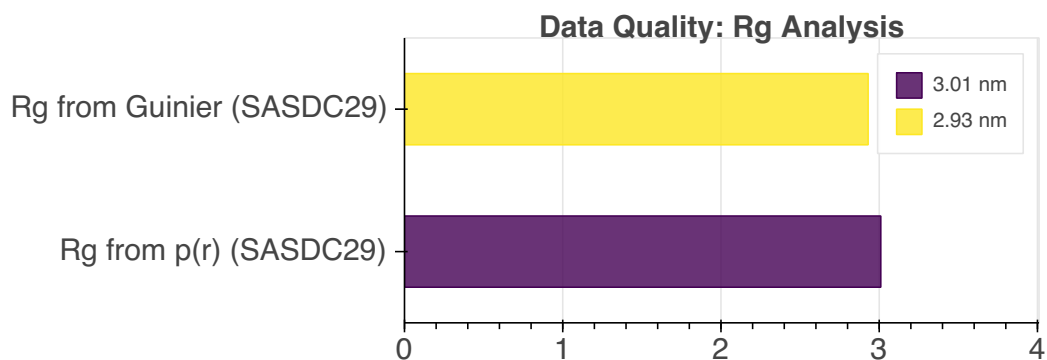
Molprobity : Version 4.4
Integrative Modeling Validation Package : Version 1.0

PDB ID	PDBDEV_00000009
Molecule Name	Structure of the human Rev7 dimer
Title	Rev7 dimerization is important for assembly and function of the Rev1/Pol translesion synthesis complex
Authors	Alessandro A. Rizzo;Faye-Marie Vassel;Nimrat Chatterjee;Sanjay D'Souza;Yunfeng Li;Bing Hao;Michael T. Hemann;Graham C. Walker;Dmitry M. Korzhnev

Overall quality

Model Quality: Molprobity Analysis

Legend:
- 8.73
- 0.42%
- 3.52%

Data Quality: Rg Analysis

| | 3.01 nm |
| | 2.93 nm |

Rg from Guinier (SASDC29)

Rg from p(r) (SASDC29)

0 1 2 3 4

Fit to SAS input: χ² Fit

| | 25.13 |

χ² value fit 1 (SASDC29)

0 5 10 15 20 25

Ensemble information

This entry consists of 0 distinct ensemble.

Summary

This entry consists of 1 unique models, with 4 subunits in each model. A total of 3 datasets or restraints was used to build this entry. Each model is represented by 0 rigid bodies and 4 flexible or non-rigid units.

Entry composition

There is 1 unique type of model in this entry.This model is titled None/Best scoring model, N/C-termini built respectively.

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Total residues

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Total residues
1	1	1	Rev7-monomer	A	212
1	2	1	Rev7-monomer	C	212
1	3	2	Rev3-RBM2	B	28
1	4	2	Rev3-RBM2	D	28

Methodology and software

Step number	Protocol ID	Method name	Method type	Number of computed models	Multi state modeling	Multi scale modeling
1	1	None	None	None	False	False

Data quality

Scattering profile

SAS data used in this integrative model was obtained from 1 deposited SASBDB entry (entries).

Scattering profile for SASDC29: 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.

Log I(q) vs q with error bars (SASDC29)

Log I(q) vs Log q with error bars (SASDC29)
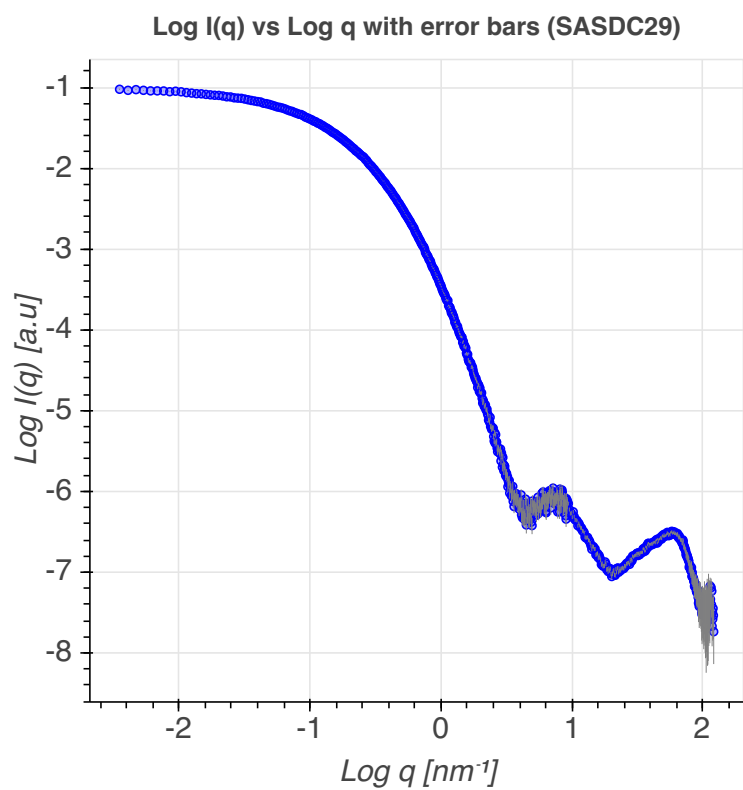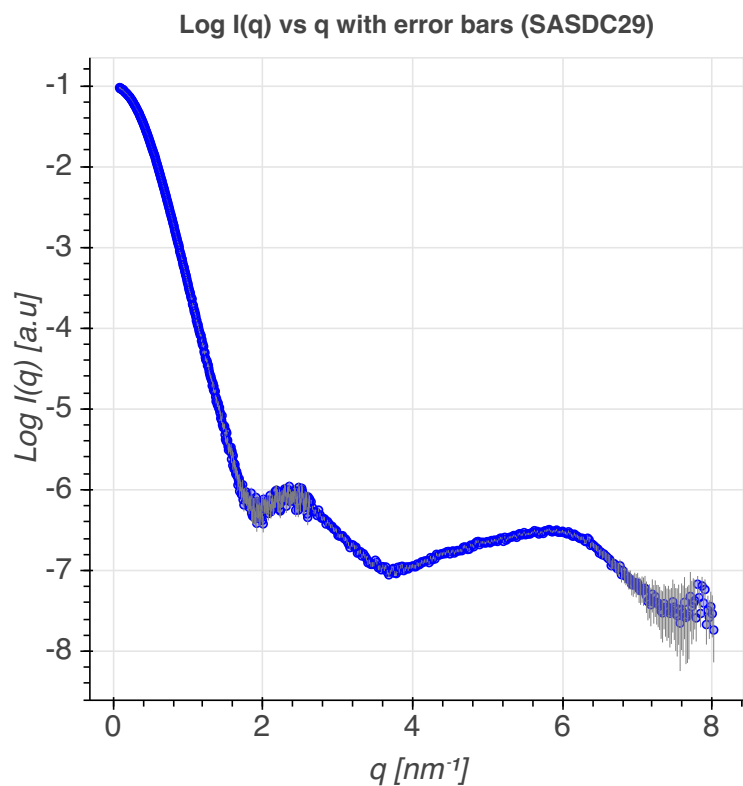
Key experimental estimates

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
SASDC29	53.4 kDa	53.4 kDa	63.8 kDa

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

SASDB ID	Estimated volume	Estimated volume method	Porod volume
SASDC29	None	None	108.40 nm³

Flexibility analysis

<u>Flexibility analysis for SASDC29</u>: 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 (SASDC29)

Dimensionless Kratky plot (SASDC29)
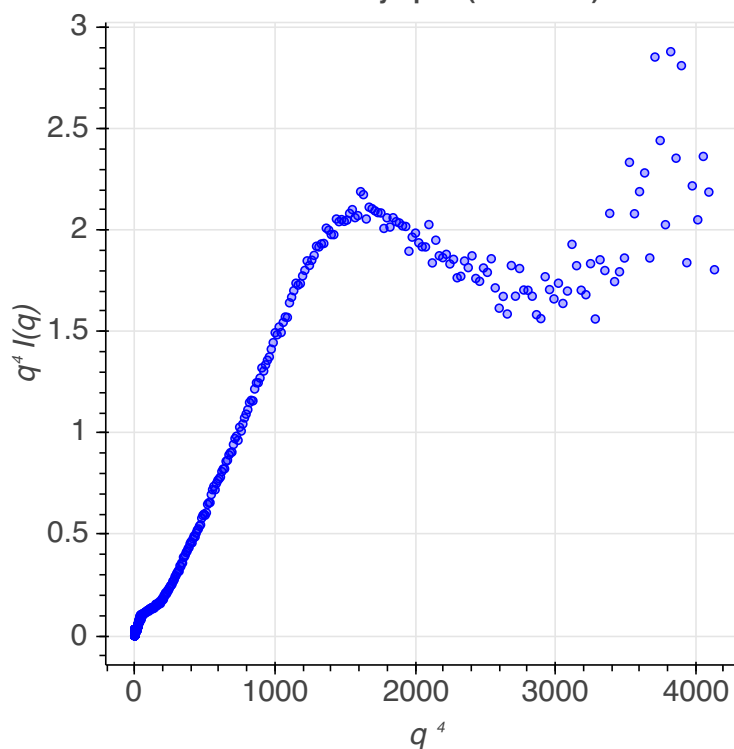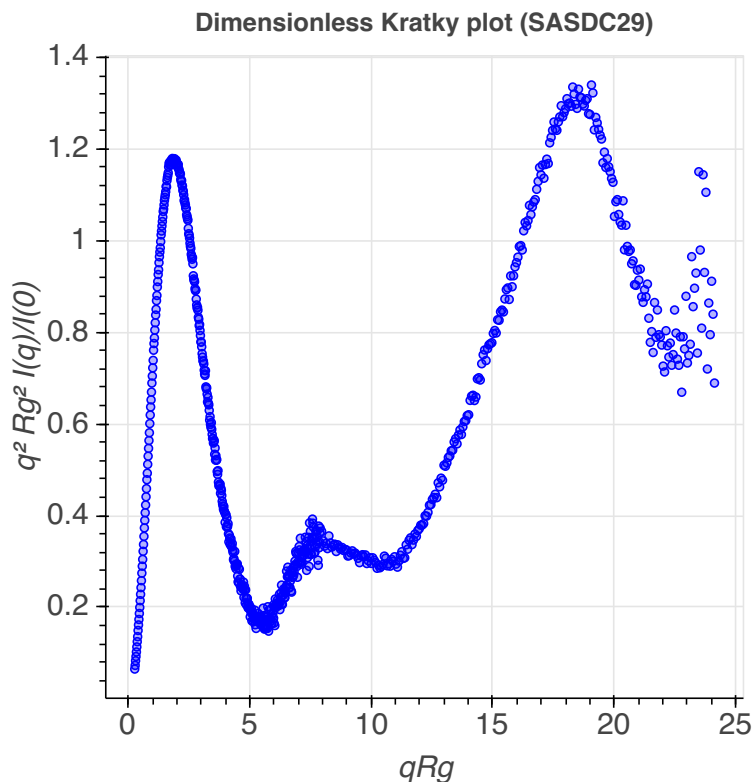

Pair-distance distribution analysis

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
SASDC29	ATSAS GNOM	11.00 nm	N/A	3.010 nm	0.004 nm

P(r) for SASDC29: The value of P(r) should be zero beyond r=Dmax.

Pair distance distribution function (SASDC29)

P(r) extrapolated fit for SASDC29
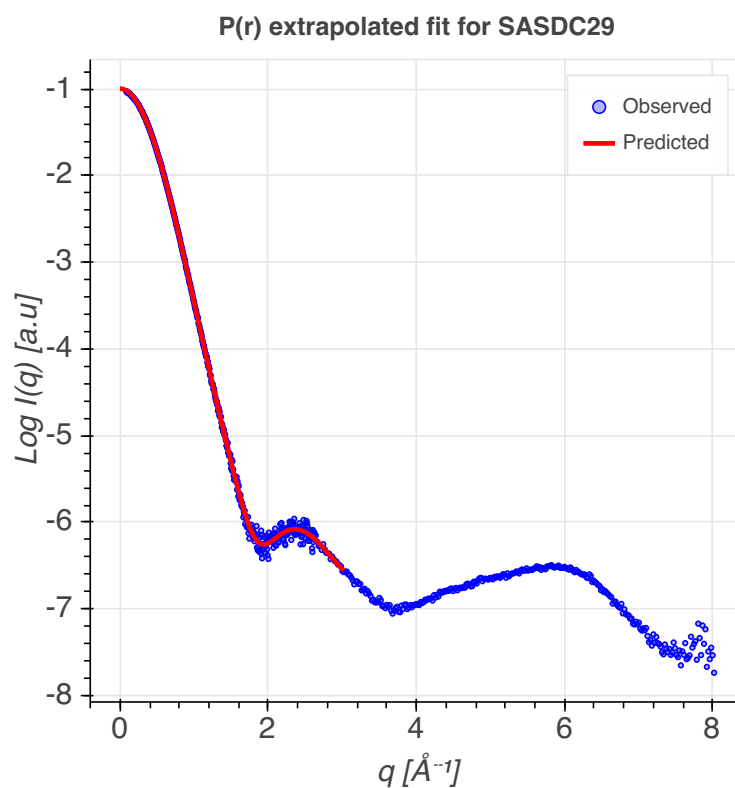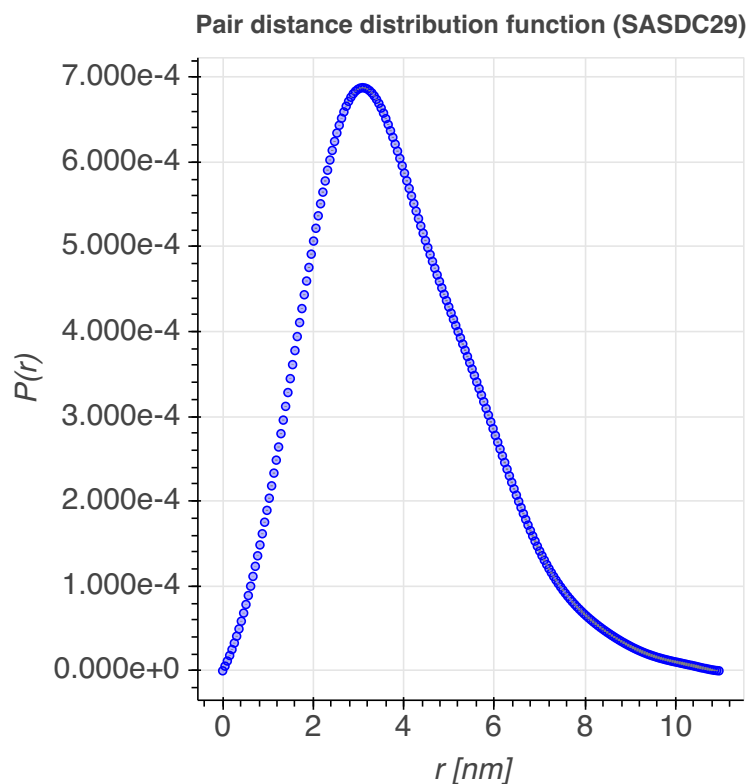
Residuals and error weighted residuals for P(r) analysis for SASDC29: Residual value plot is a measure to assess fit to the data. sResidual values should be equally and randomly spaced around the horizontal axis.

Residuals for P(r) fit (SASDC29)


Error weighted residuals for P(r) fit (SASDC29)
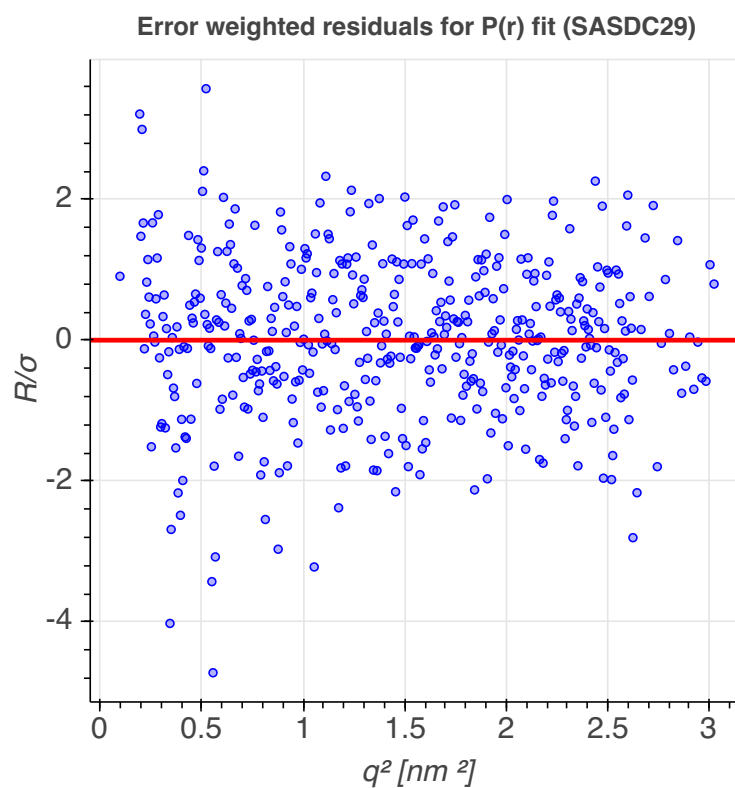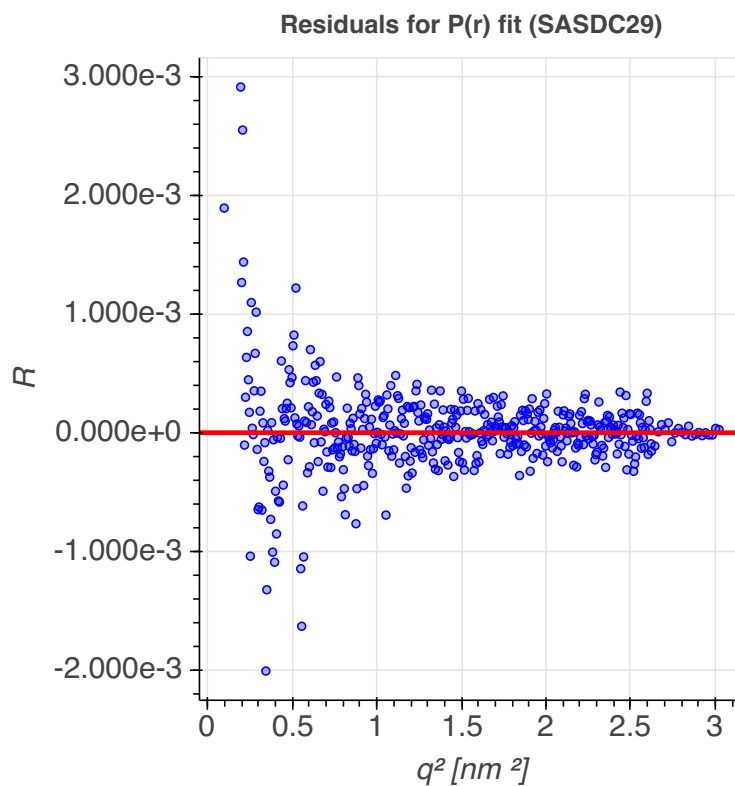
Guinier analysis

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
SASDC29	2.93 nm	0.003 nm	53.400 nm	0.040 nm

Guinier analysis for SASDC29: 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 plot for SASDC29 (R^2=1.00)

Residuals for Guinier plot fit (SASDC29)
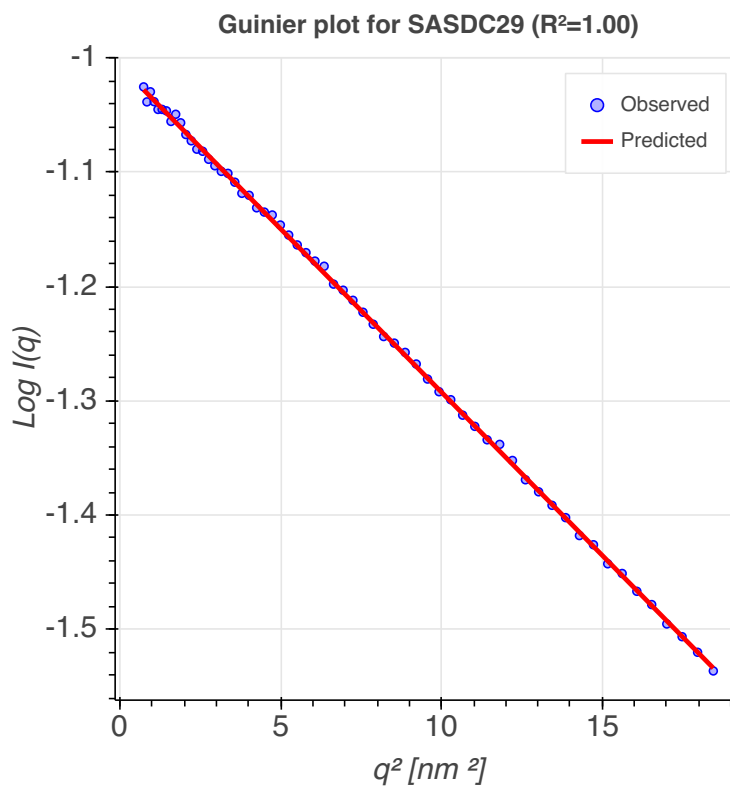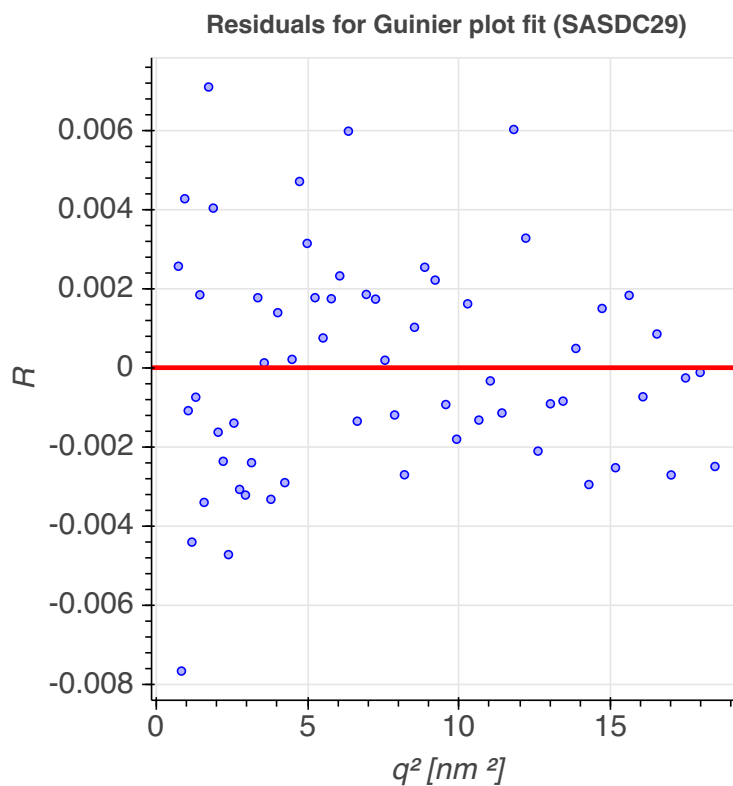
Model quality

Standard geometry

There are 123 bond outliers in this entry.

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
N--H	1.079	0.86	13
NE2--HE2	1.03	0.86	1
CD2--HD2	1.086	0.93	1
CE1--HE1	1.082	0.93	1
CB--HB	1.111	0.97	1
CE--HE3	1.102	0.97	5
CB--HB2	1.106	0.97	11
CB--HB3	1.106	0.97	11
CE--HE2	1.104	0.97	5

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
CD--HD2	1.109	0.97	1
CG--HG2	1.102	0.97	7
NZ--HZ1	1.027	0.89	1
CA--HA2	1.105	0.97	3
CD--HD3	1.106	0.97	1
CA--HA	1.102	0.97	13
CG2--HG23	1.106	0.97	1
CG--HG3	1.102	0.97	7
CG2--HG21	1.106	0.97	1
CE--HE1	1.105	0.97	3
CG2--HG22	1.105	0.97	1
NZ--HZ2	1.025	0.89	1
N--H1	1.08	0.96	3
OG--HG	0.961	0.84	1
OG1--HG1	0.96	0.84	1
NZ--HZ3	1.009	0.89	1
CA--HA3	1.05	0.97	2

There are 14 angle outliers in this entry.

Angle type	Observed angle (°)	Ideal angle (°)	Number of outliers
CA-N-H	126.774	114.0	13

Too-close contacts

The following all-atom clashscore is based on a MolProbity analysis. All-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The table below contains clashscores for all the models in this entry.

Model ID	Clash score	Number of clashes
Model 1	8.73	69

All 69 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Torsion angles: Protein backbone

In the following table, Ramachandran outliers are listed. The Analysed column shows the number of residues for which the backbone conformation was analysed.

Model ID	Analyzed	Favored	Allowed	Outliers
1	482	472	8	2

Detailed list of outliers are tabulated below.

Model ID	Chain and res ID	Residue type
1	A:110	ILE
1	C:110	ILE

Torsion angles: Protein sidechains

In the following table, sidechain outliers are listed. The Analysed column shows the number of residues for which the sidechain conformation was analysed.

Model ID	Analyzed	Favored	Allowed	Outliers
1	454	416	22	16

Detailed list of outliers are tabulated below.

Model ID	Chain and res ID	Residue type
1	A:3	THR
1	A:110	ILE
1	A:120	GLU
1	A:158	ARG
1	A:168	ASP

Model ID	Chain and res ID	Residue type
1	A:208	HIS
1	B1990	LYS
1	B2011	GLN
1	C:3	THR
1	C:110	ILE
1	C:120	GLU
1	C:158	ARG
1	C:168	ASP
1	C:208	HIS
1	D1990	LYS
1	D2011	GLN

# 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
SASDC29	1	25.13

Model fit for SASDC29 (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 SASDC29

Residuals for model fit (SASDC29)

Error-weighted residuals for model fit (SASDC29)
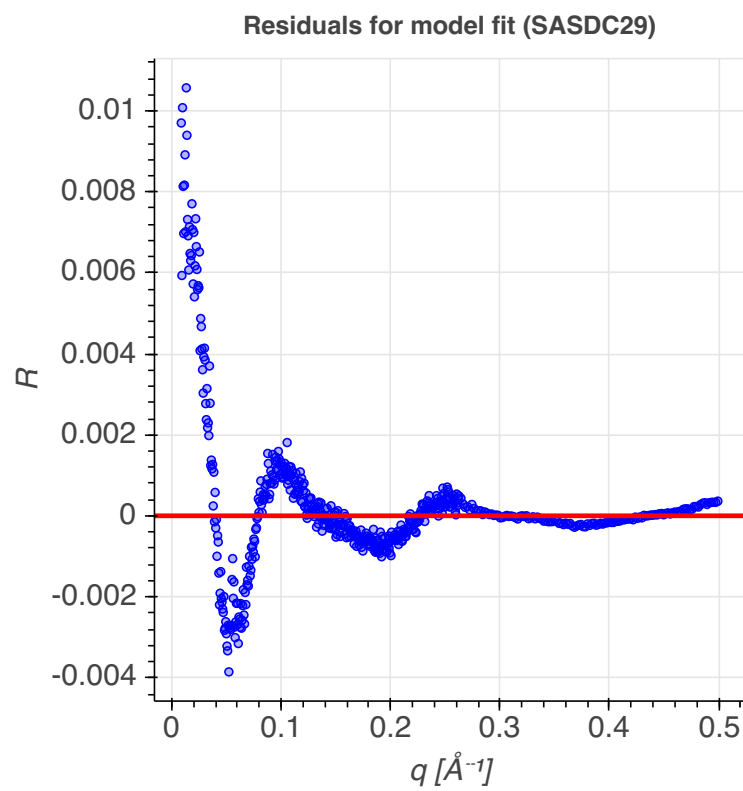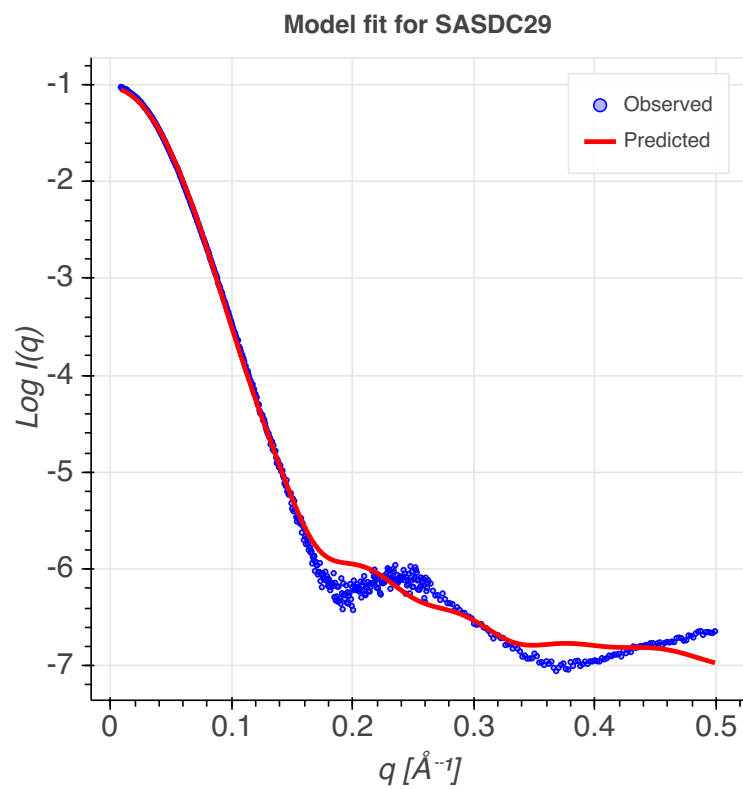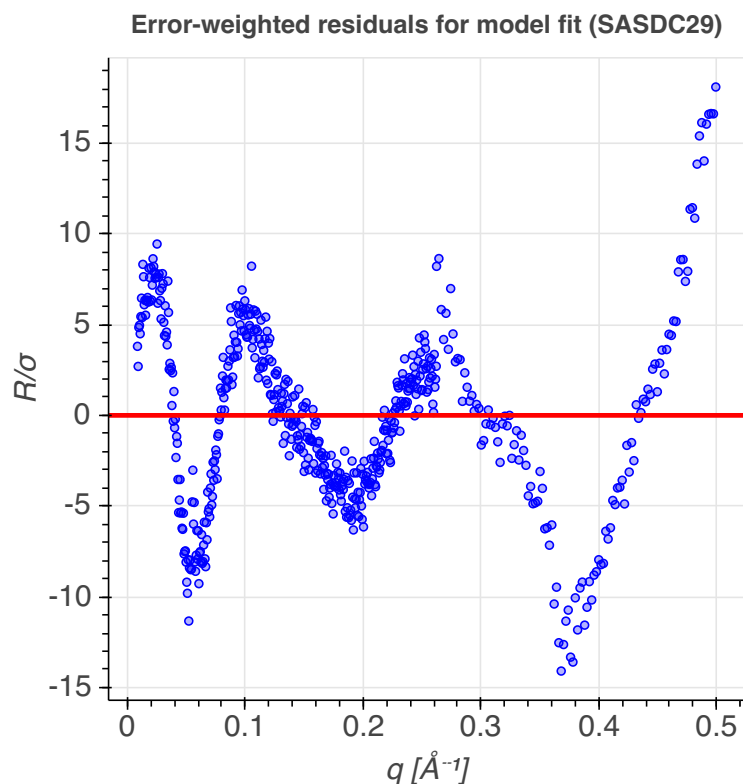
Cormap p-value analysis of fits

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
SASDC29	1	0.00E+00

Fit of model to data not used for modeling

Uncertainty of data and model