

# Integrative Structure Validation Report ?

November 11, 2021 - 09:47 PM

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

*Molprobability* Version 4.4

*ATSAS* Version 3.0.3

*Integrative Modeling Validation* Version 1.0

*Integrative Modeling Platform* Version 2.15.0

PDB ID	PDBDEV_00000009
Structure Name	Structure of the human Rev7 dimer
Publication Title	Rev7 dimerization is important for assembly and function of the Rev1/Pol $\zeta$ 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

*This is a PDB-Dev IM Structure Validation Report for a publicly released PDB-Dev entry.*

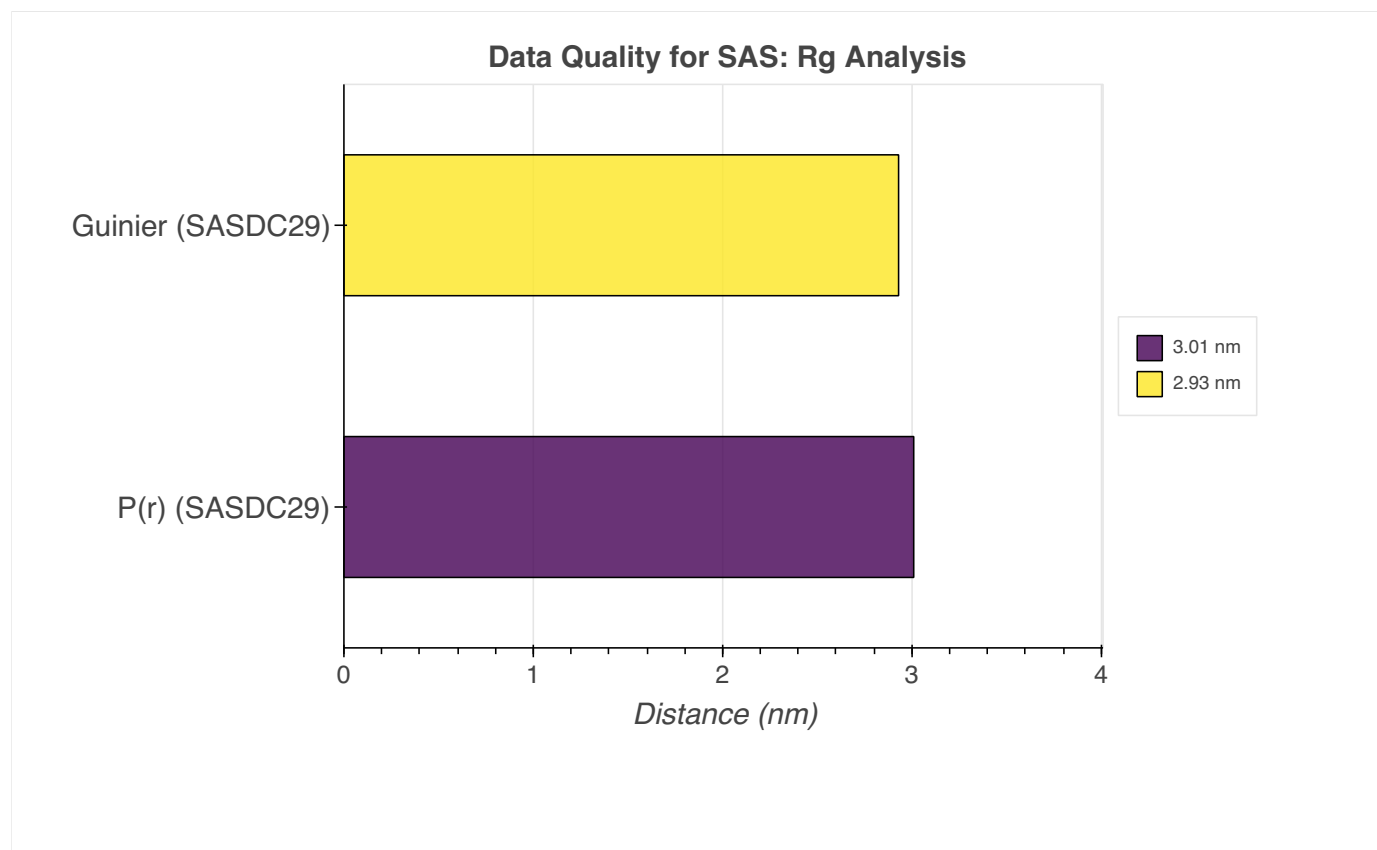
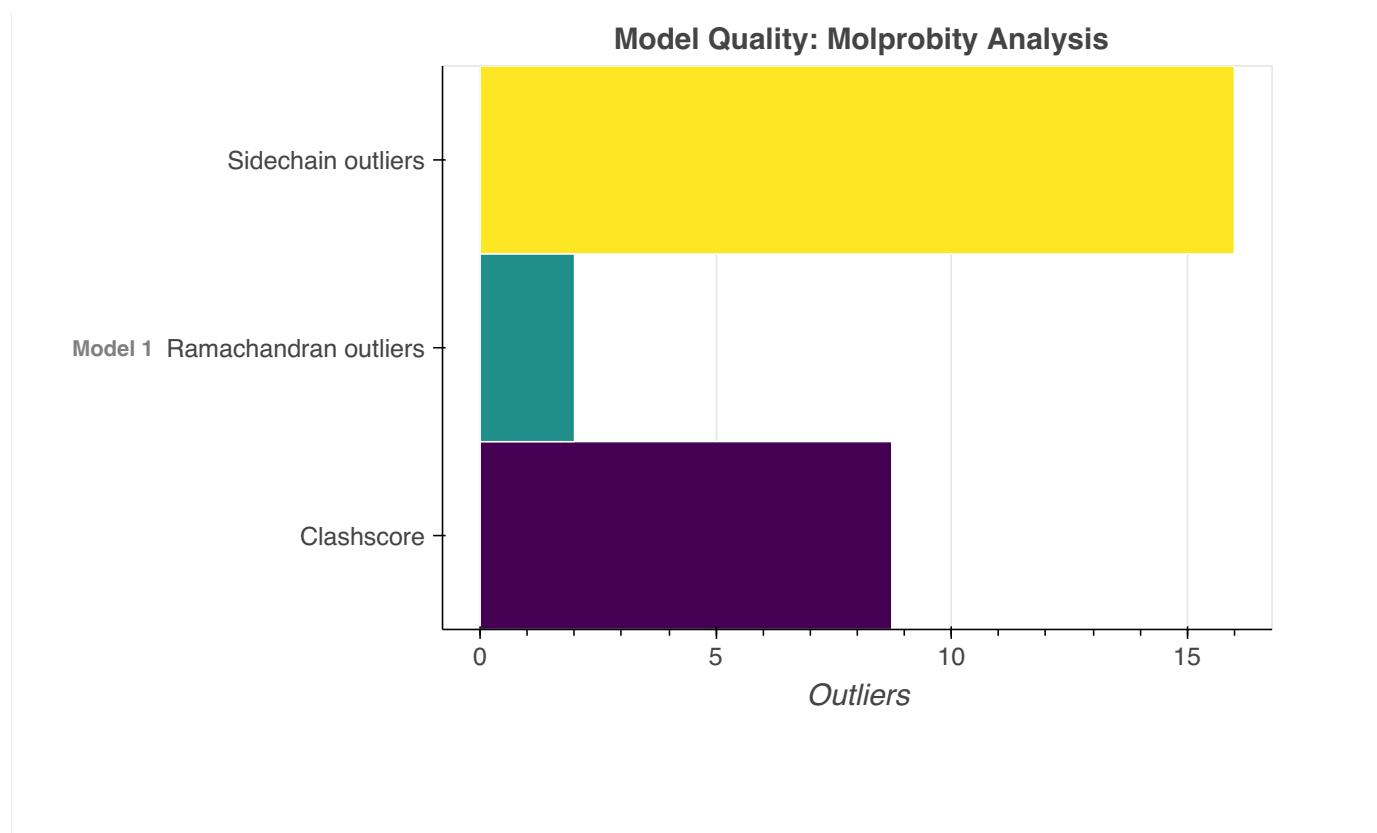
*We welcome your comments at [pdb-dev@mail.wwpdb.org](mailto:pdb-dev@mail.wwpdb.org)*

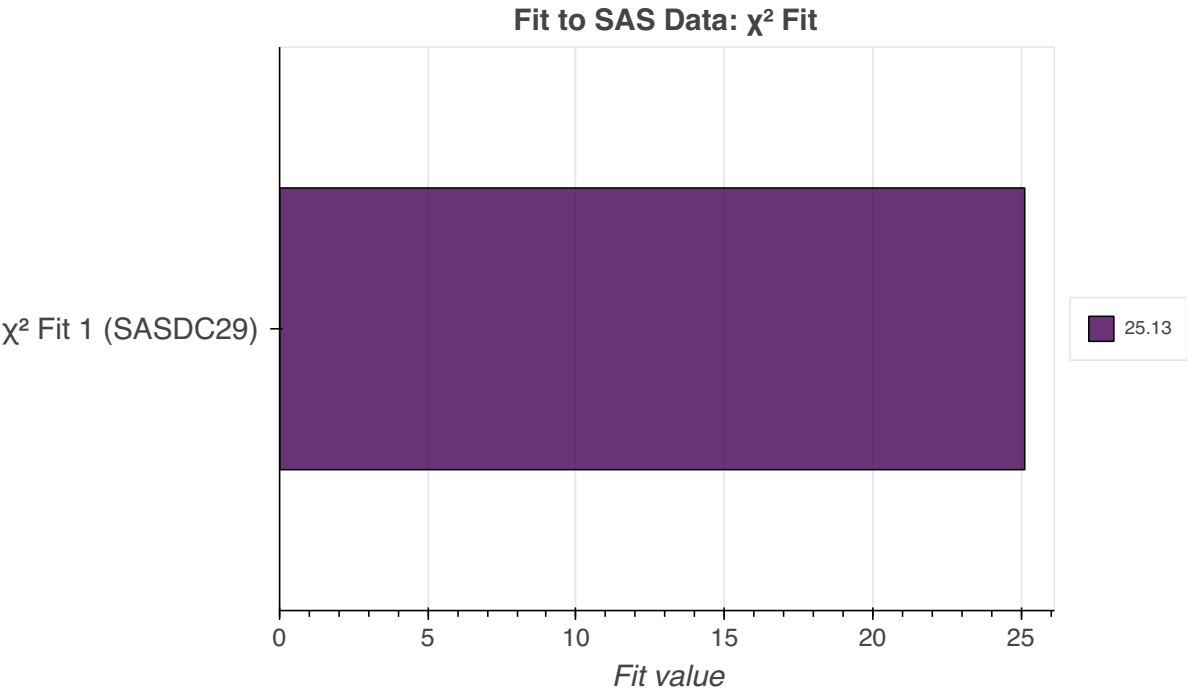
*A user guide is available at [https://pdb-dev-beta.wwpdb.org/validation\\_help.html](https://pdb-dev-beta.wwpdb.org/validation_help.html) with specific help available everywhere you see the ? symbol.*

*List of references used to build this report is available [here](#).*

## Overall quality ?

*This validation report contains model quality assessments for all structures, data quality assessment for SAS datasets and fit to model assessments for SAS datasets. Data quality and fit to model assessments for other datasets and model uncertainty are under development.*





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
1	1	1	Rev7-monomer	A	212
1	2	1	Rev7-monomer	C	212
1	3	2	Rev3-RBM2	B	28

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Total residues
1	4	2	Rev3-RBM2	D	28

### Datasets used for modeling ?

There are 3 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	SAS data	SASBDB	SASDC29
2	Experimental model	PDB	6BC8
3	Mutagenesis data	Not available	None

### Representation ?

This entry has only one representation and includes 0 rigid bodies and 4 flexible units.

Chain ID	Rigid bodies	Non-rigid segments
A	-	1-212.
B	-	1-28.
C	-	1-212.
D	-	1-28.

### 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

There is 1 software package reported in this entry.

ID	Software name	Software version	Software classification	Software location
----	---------------	------------------	-------------------------	-------------------

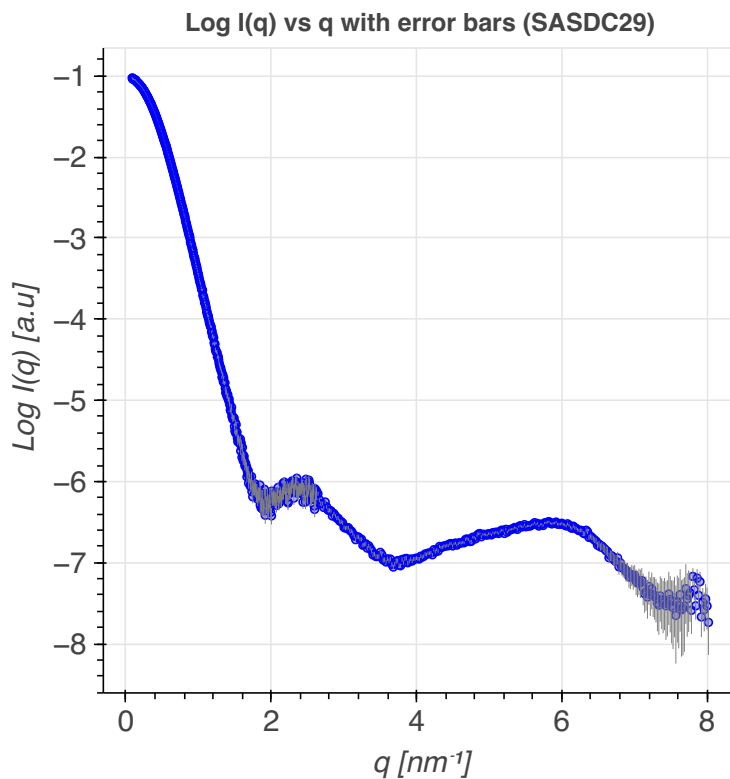
ID	Software name	Software version	Software classification	Software location
1	HADDOCK	Not available	model building	<a href="http://haddock.science.uu.nl/services/HADDOCK/">http://haddock.science.uu.nl/services/HADDOCK/</a>

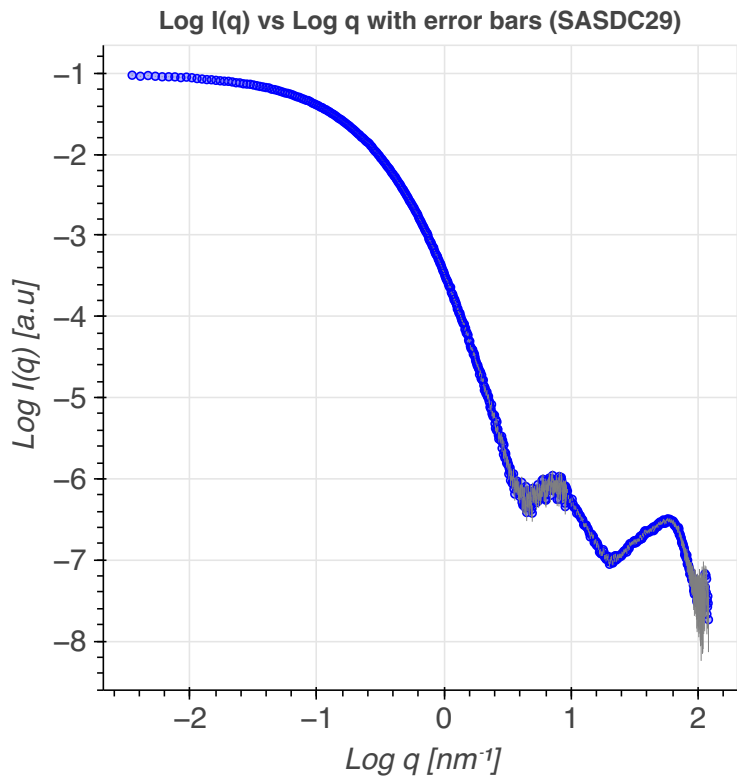
### Data quality ?

### 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.





Key experimental estimates ?

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

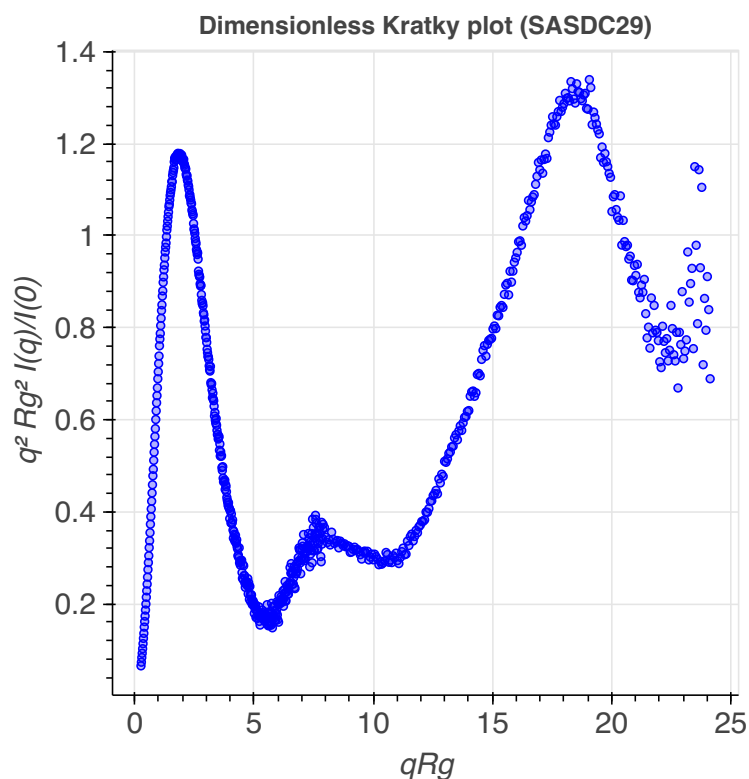
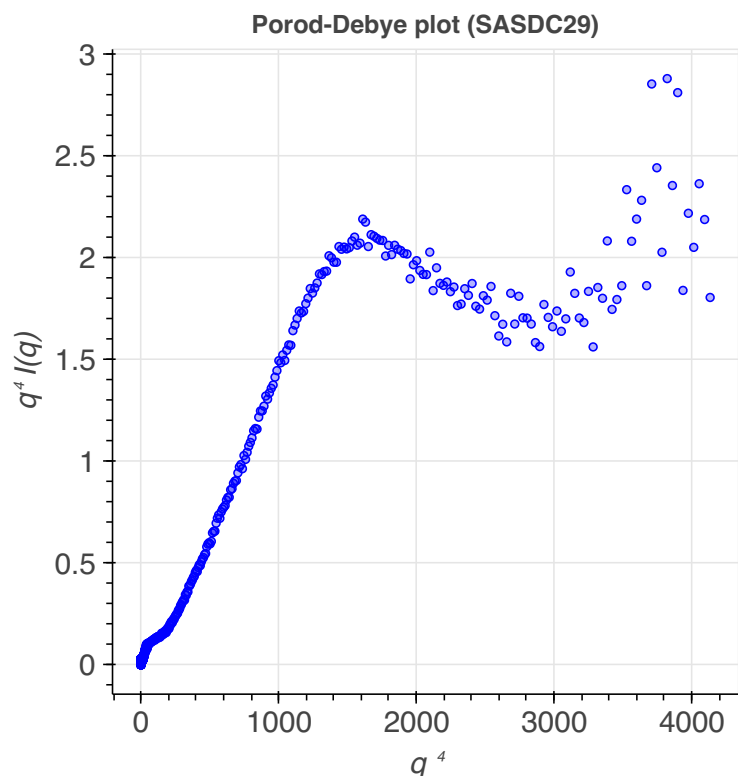
SASDB ID	Chemical composition MW	Standard MW	Porod Volume/MW
SASDC29	53.4 kDa	53.4 kDa	1.7 nm <sup>3</sup> /kDa

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

SASDB ID	Estimated Volume	Porod Volume	Specific Volume	Sample Contrast	Sample Concentration
SASDC29	N/A	108.4 nm <sup>3</sup>	N/A	N/A	10.6 mg/ml

Flexibility analysis ?

Flexibility analysis for SASDC29: 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.



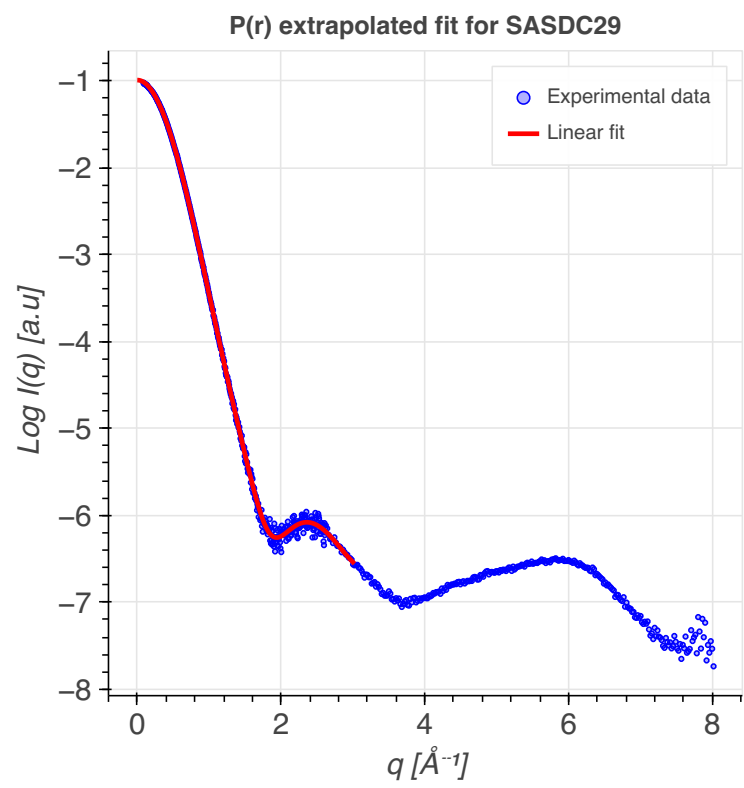
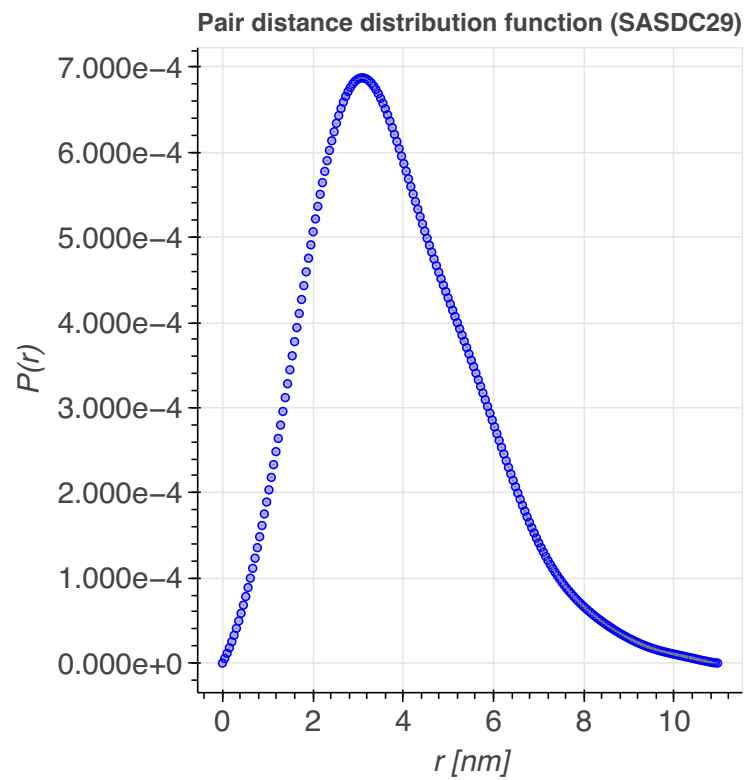
### Pair-distance distribution analysis

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  $D_{\text{max}}$  and represents the maximum size of the particle.

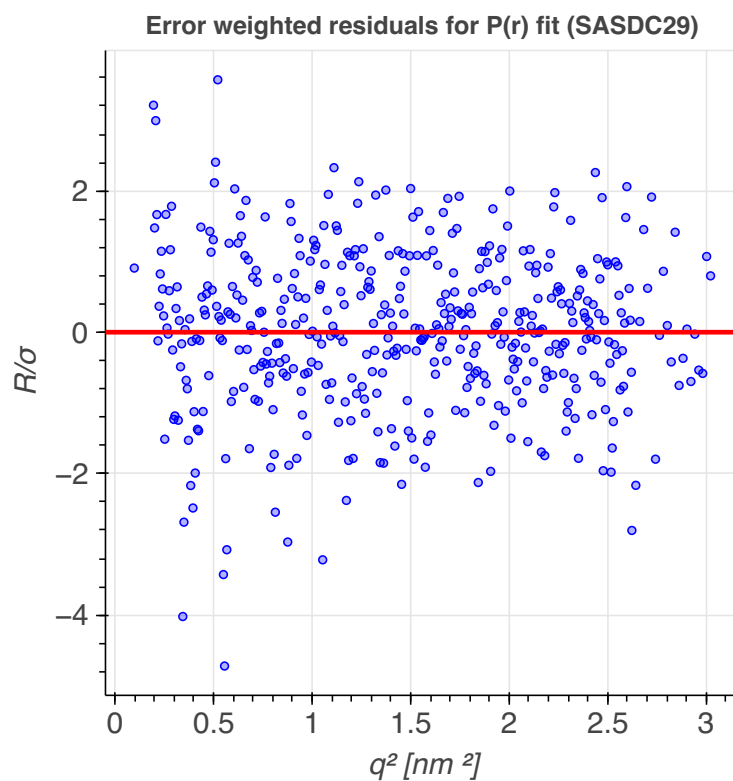
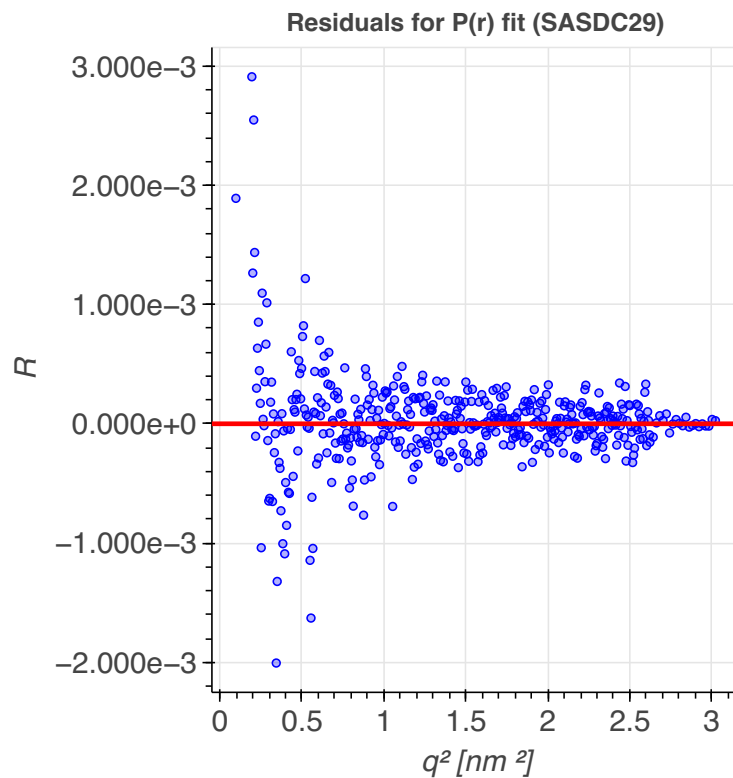
SASDB ID	Software used	Dmax	Dmax error	Rg	Rg error
SASDC29	ATSAS GNOM	11.0 nm	None nm	3.01 nm	0.004 nm

P(r) for SASDC29: The value of P(r) should be zero beyond  $r=D_{\text{max}}$ .





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



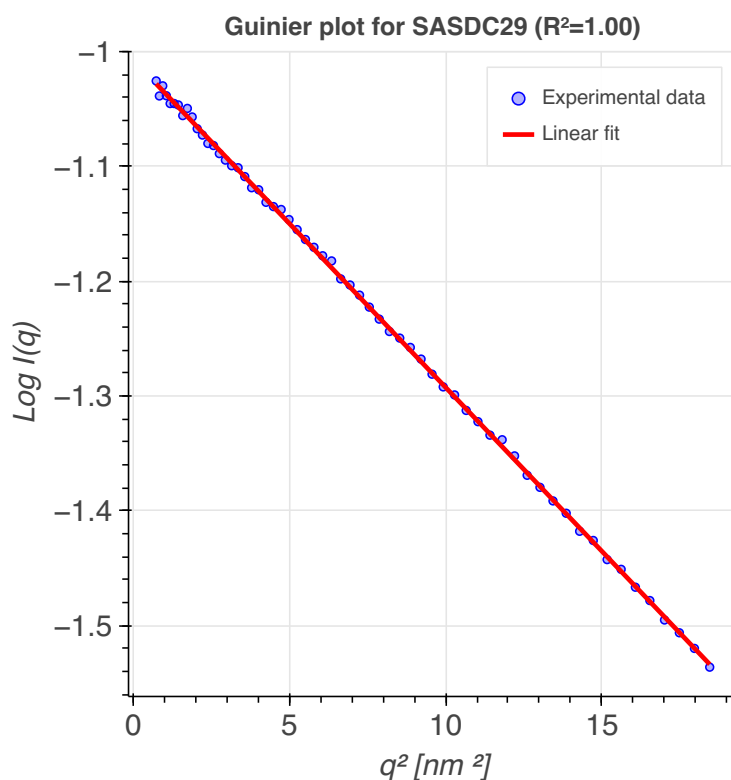
### Guinier analysis ?

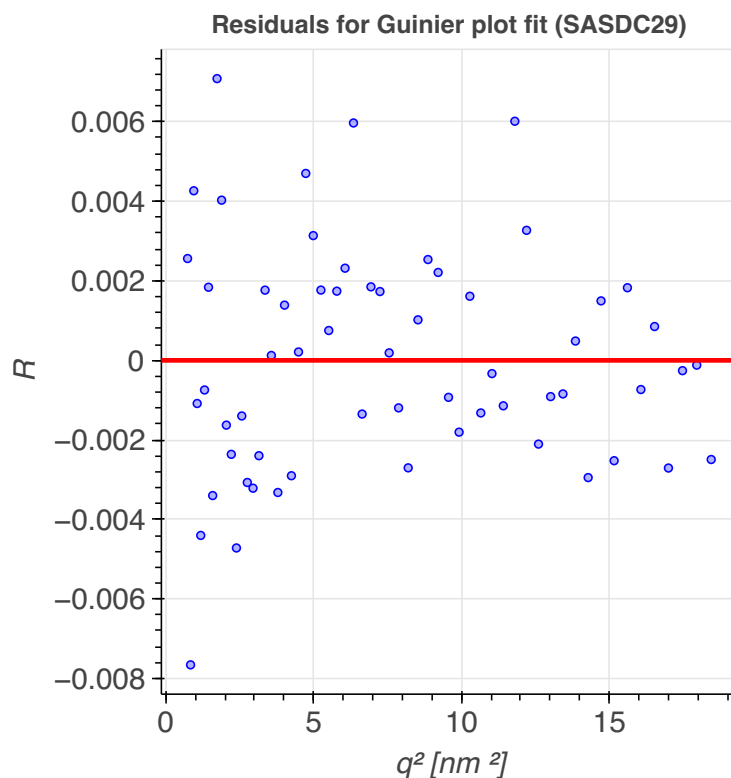
Guinier analysis: agreement between the P(r) and Guinier-determined  $R_g$  (table below) is a good measure of the self-

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	0.0 nm	53.4 nm	0.04 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 same size. 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.





### Mutagenesis

Validation for this section is under development.

### Model quality ?

#### Standard geometry: bond outliers ?

There are 123 bond outliers in this entry. A summary is provided below, and a detailed list of outliers can be found [here](#).

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
N--H	1.08	0.86	14
NE2--HE2	1.03	0.86	2
CD2--HD2	1.09	0.93	2
CE1--HE1	1.08	0.93	2
CB--HB	1.11	0.97	2
CE--HE3	1.11	0.97	2
CB--HB2	1.11	0.97	12

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
CB--HB3	1.11	0.97	12
CE--HE2	1.11	0.97	2
CD--HD2	1.11	0.97	2
CG--HG2	1.11	0.97	6
NZ--HZ1	1.03	0.89	2
CA--HA2	1.11	0.97	2
CD--HD3	1.11	0.97	2
CA--HA	1.11	0.97	6
CG2--HG23	1.11	0.97	2
CG--HG3	1.11	0.97	2
CG2--HG21	1.11	0.97	2
CE--HE1	1.11	0.97	1
CG2--HG22	1.11	0.97	1
CA--HA	1.1	0.97	8
CA--HA2	1.1	0.97	2
CG2--HG22	1.1	0.97	1
CE--HE1	1.1	0.97	3
NZ--HZ2	1.02	0.89	2
CG--HG3	1.1	0.97	6
CE--HE2	1.1	0.97	4
CE--HE3	1.1	0.97	4
CG--HG2	1.1	0.97	2
N--H1	1.08	0.96	4
OG--HG	0.96	0.84	2

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
OG1--HG1	0.96	0.84	2
NZ--HZ3	1.01	0.89	2
CA--HA3	1.05	0.97	3

### Standard geometry: angle outliers ?

There are 48 angle outliers in this entry. A summary is provided below, and a detailed list of outliers can be found [here](#).

Angle type	Observed angle (°)	Ideal angle (°)	Number of outliers
ND1-CE1-NE2	108.4	118.08	1
ND1-CE1-NE2	108.4	118.05	1
C-N-CA	106.25	15.45	1
C-N-CA	106.26	15.44	1
C-N-CA	121.7	106.28	1
C-N-CA	121.7	106.29	1
C-N-CA	121.7	106.3	2
C-N-CA	121.7	106.31	2
C-N-CA	121.7	106.32	1
C-N-CA	121.7	106.33	2
C-N-CA	121.7	106.35	1
C-N-CA	121.7	106.36	1
C-N-CA	121.7	106.37	1
C-N-CA	121.7	109.48	1
C-N-CA	109.5	12.2	1
C-N-CA	109.51	12.19	1
C-N-CA	121.7	109.51	1

Angle type	Observed angle (°)	Ideal angle (°)	Number of outliers
CB-CG-CD2	131.2	122.65	1
CB-CG-CD2	131.2	122.73	1
CD2-NE2-CE1	109.0	102.59	1
CD2-NE2-CE1	109.0	102.61	1
CG-ND1-CE1	109.3	99.85	1
CG-ND1-CE1	109.3	99.88	1
ND1-CG-CD2	106.1	111.47	2
CG-CD-NE	112.0	102.05	1
CG-CD-NE	112.0	102.09	1
CA-CB-CG	114.1	122.52	1
CA-CB-CG	114.1	122.5	1
CA-CB-CG	114.1	105.85	1
CA-CB-CG	114.1	105.87	1
CA-N-H	126.91	114.0	1
CA-N-H	126.87	114.0	2
CA-N-H	126.85	114.0	2
CA-N-H	126.84	114.0	1
CA-N-H	126.83	114.0	1
CA-N-H	126.81	114.0	3
CA-N-H	126.8	114.0	1
CA-N-H	126.78	114.0	2
CA-N-H	126.77	114.0	1

### 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.

Model ID	Atom-1	Atom-2	Clash overlap (Å)
1	A:46:LYS:NZ	C:139:HIS:NE2	1.297
1	D2009:TRP:CZ2	D2014:GLU:OE1	1.230
1	B2010:LEU:O	D2013:LYS:HE3	1.223
1	B2009:TRP:CZ2	B2014:GLU:OE1	1.217
1	A:46:LYS:NZ	C:139:HIS:CE1	1.193
1	A:46:LYS:HZ1	C:139:HIS:CE1	1.178
1	A:132:VAL:HB	C:132:VAL:HG12	1.071
1	A:32:TYR:OH	C:39:VAL:HB	1.048
1	A:46:LYS:HZ1	C:139:HIS:CD2	1.026
1	B2009:TRP:CE2	B2014:GLU:OE1	0.979
1	D2009:TRP:CE2	D2014:GLU:OE1	0.963
1	B2013:LYS:HA	D2013:LYS:HE2	0.950
1	A:132:VAL:HG11	C:132:VAL:HA	0.949
1	B2010:LEU:O	D2013:LYS:CE	0.936
1	A:132:VAL:CB	C:132:VAL:HG12	0.870
1	D2009:TRP:CZ2	D2014:GLU:CD	0.867
1	A:132:VAL:CG1	C:132:VAL:HA	0.858
1	B2009:TRP:CZ2	B2014:GLU:CD	0.854
1	A:46:LYS:CE	C:139:HIS:CE1	0.826
1	B2013:LYS:HE3	D2012:ALA:O	0.798

Model ID	Atom-1	Atom-2	Clash overlap (Å)
1	A:35:GLU:OE2	C:44:LYS:CB	0.789
1	B2013:LYS:HE3	D2012:ALA:C	0.775
1	A:35:GLU:OE2	C:44:LYS:HB2	0.765
1	A:46:LYS:HE2	C:139:HIS:CE1	0.737
1	A:46:LYS:HZ3	C:139:HIS:CE1	0.726
1	A:132:VAL:HB	C:132:VAL:CG1	0.720
1	A:132:VAL:CG2	C:132:VAL:HG12	0.719
1	A:44:LYS:HE2	C:35:GLU:HB3	0.679
1	A:35:GLU:OE2	C:44:LYS:HB3	0.670
1	A:44:LYS:HE2	C:35:GLU:CB	0.637
1	A:43:GLN:CD	A:56:BCYS:SG	0.571
1	B2013:LYS:HD3	D2013:LYS:HE2	0.571
1	C:43:GLN:CD	C:56:BCYS:SG	0.570
1	A:48:TYR:HA	A:120:GLU:HG3	0.556
1	C:48:TYR:HA	C:120:GLU:HG3	0.548
1	D2009:TRP:CZ2	D2014:GLU:OE2	0.547
1	A:132:VAL:CB	C:132:VAL:HA	0.547
1	B2009:TRP:CZ2	B2014:GLU:OE2	0.544
1	B2009:TRP:CH2	B2014:GLU:OE1	0.539
1	D2009:TRP:CH2	D2014:GLU:OE1	0.528
1	B2009:TRP:HZ2	B2014:GLU:OE2	0.511
1	A:32:TYR:CE2	C:32:TYR:OH	0.510
1	D2009:TRP:HZ2	D2014:GLU:OE2	0.492
1	A:132:VAL:HG12	C:131:SER:O	0.491



Model ID	Atom-1	Atom-2	Clash overlap (Å)
1	A:32:TYR:HE2	C:32:TYR:OH	0.489
1	A:35:GLU:CD	C:44:LYS:HB3	0.479
1	A:115:LEU:HD23	A:196:ILE:HD11	0.466
1	A:134:ASP:O	C:51:PRO:HG2	0.458
1	C:115:LEU:HD23	C:196:ILE:HD11	0.454
1	A:23:LEU:O	A:27:VAL:HG23	0.453
1	C:23:LEU:O	C:27:VAL:HG23	0.450
1	A:192:MET:HB3	A:192:MET:HE2	0.439
1	C:97:LYS:O	C:203:VAL:HA	0.434
1	A:97:LYS:O	A:203:VAL:HA	0.433
1	D2009:TRP:HZ2	D2014:GLU:CD	0.431
1	C:157:THR:H	C:160:MET:HE2	0.422
1	A:157:THR:H	A:160:MET:HE2	0.417
1	B2013:LYS:HD3	D2013:LYS:CE	0.415
1	C:78:ASN:O	C:153:ARG:HD3	0.415
1	A:167:LYS:O	A:168:ASP:HB2	0.414
1	A:32:TYR:OH	C:32:TYR:HE1	0.413
1	A:78:ASN:O	A:153:ARG:HD3	0.413
1	D1988:GLU:O	D1989:ASP:OD1	0.409
1	A:132:VAL:HG23	C:132:VAL:HG12	0.408
1	B1988:GLU:O	B1989:ASP:OD1	0.408
1	B2009:TRP:HZ2	B2014:GLU:CD	0.407
1	A:32:TYR:HH	C:39:VAL:HB	0.407
1	C:167:LYS:O	C:168:ASP:HB2	0.406

Model ID	Atom-1	Atom-2	Clash overlap (Å)
1	A:27:VAL:O	A:31:LEU:HG	0.400

### 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	Analysed	Favored	Allowed	Outliers
1	482	472	8	2

Detailed list of outliers are tabulated below.

Model ID	Chain	Residue 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	Analysed	Favored	Allowed	Outliers
1	454	416	22	16

Detailed list of outliers are tabulated below.

Model ID	Chain	Residue ID	Residue type
1	A	3	THR
1	A	110	ILE
1	A	120	GLU
1	A	158	ARG
1	A	168	ASP
1	A	208	HIS
1	B	1990	LYS

Model ID	Chain	Residue ID	Residue type
1	B	2011	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	D	1990	LYS
1	D	2011	GLN

### Fit of model to data used for modeling ?

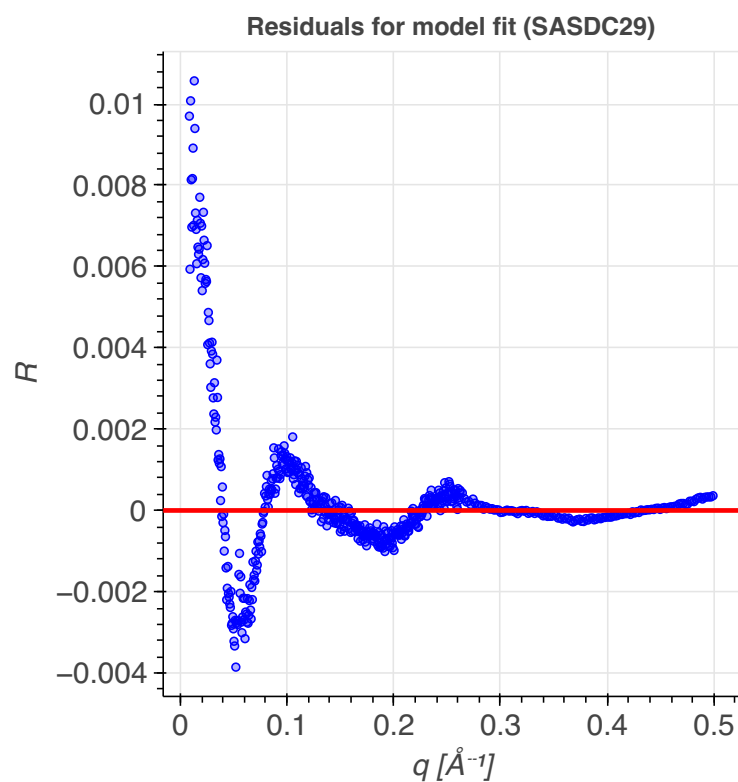
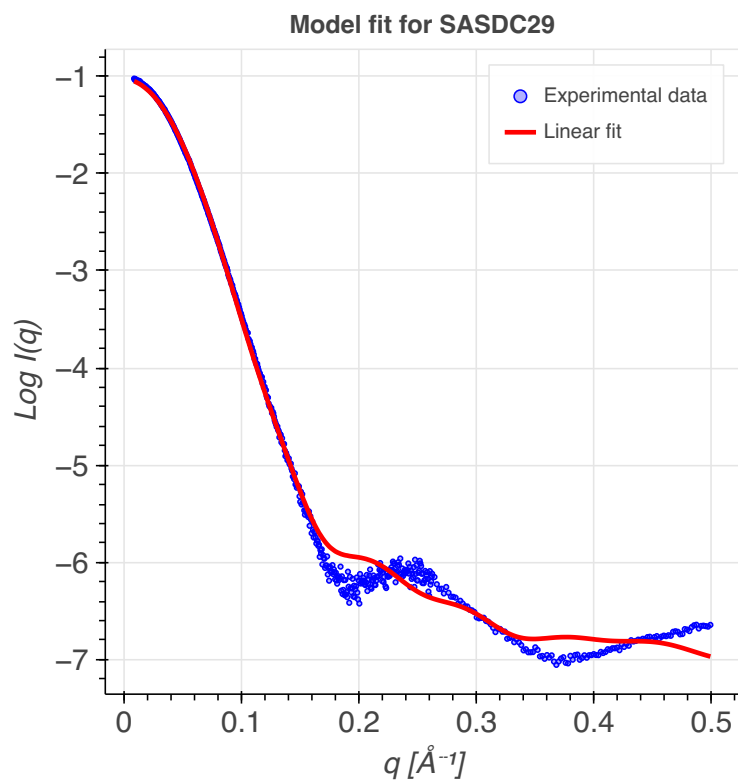
#### Fit of model(s) to SAS data

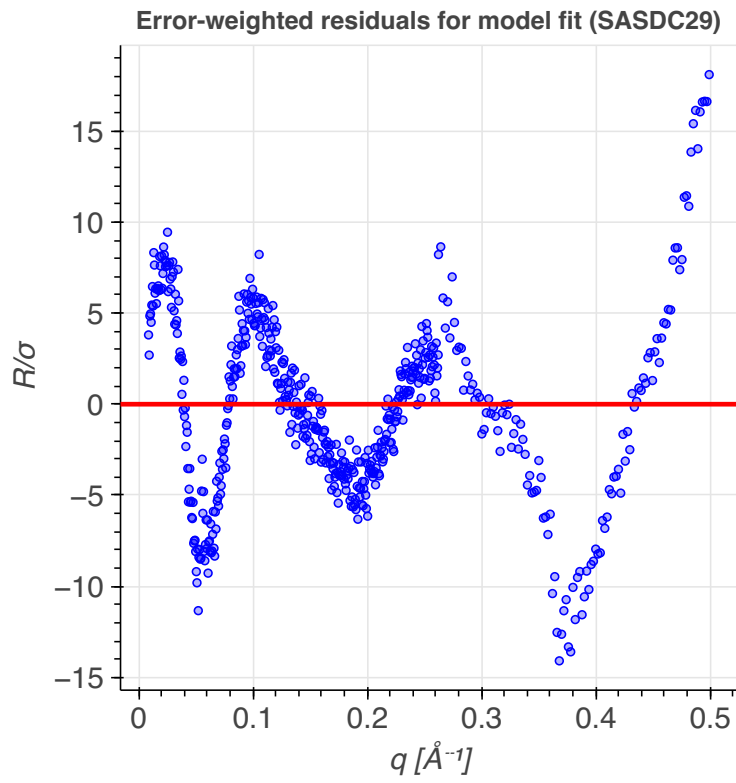
### $\chi^2$ goodness of fit analysis ?

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

SASDB ID	Model	$\chi^2$
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.





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

Mutagenesis

Validation for this section is under development.

Fit of model to data used for validation ?

Validation for this section is under development.

Acknowledgements

Development of integrative model validation metrics, implementation of a model validation pipeline, and creation of a

validation report for integrative structures, are funded by NSF ABI awards (DBI-1756248, DBI-2112966, DBI-2112967, DBI-2112968, and DBI-1756250). The [PDB-Dev team](#) and members of [Sali lab](#) contributed model validation metrics and software packages.

Implementation of validation methods for SAS data and SAS-based models are funded by [RCSB PDB](#) (grant number DBI-1832184). Dr. Stephen Burley, Dr. John Westbrook, and Dr. Jasmine Young from [RCSB PDB](#), Dr. Jill Trehella, Dr. Dina Schneidman, and members of the [SASBDB](#) repository are acknowledged for their advice and support in implementing SAS validation methods.

Members of the [wwPDB Integrative/Hybrid Methods Task Force](#) provided recommendations and community support for the project.