

# Integrative Structure Validation Report ?

March 27, 2025 - 09:56 AM PDT

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

*Integrative Modeling Validation Version 2.0*

*Python-IHM Version 1.8*

*MolProbity Version 4.5.2*

*ATSAS Version 3.2.1 (r14885)*

PDB ID	8ZZ9
PDB-Dev ID	PDBDEV_00000009
Structure Title	Structure of the human Rev7 dimer
Structure 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
Deposited on	2017-11-16

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

*We welcome your comments at [helpdesk@pdb-ihm.org](mailto:helpdesk@pdb-ihm.org)*

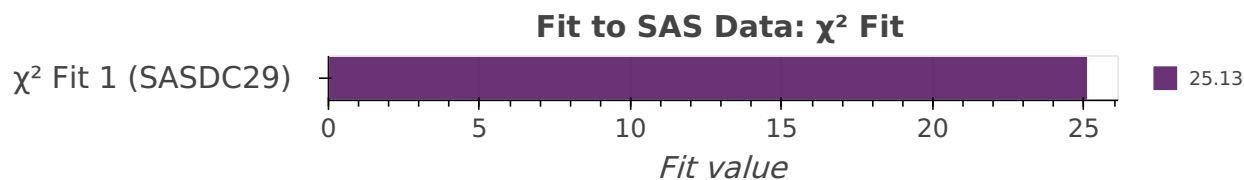
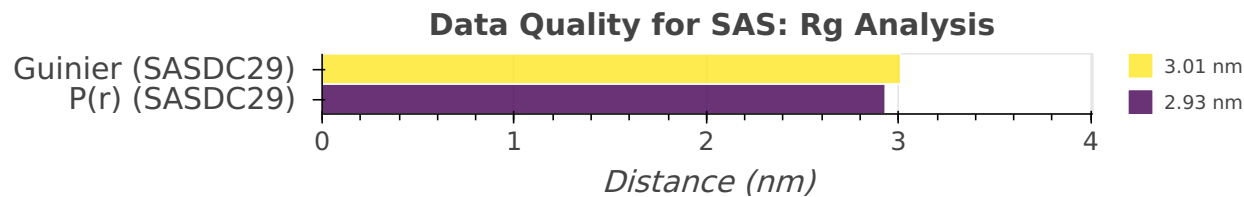
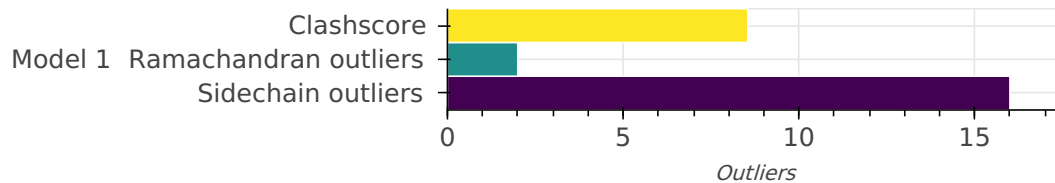
*A user guide is available at [https://pdb-ihm.org/validation\\_help.html](https://pdb-ihm.org/validation_help.html) with specific help available everywhere you see the ? symbol.*

*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 and fit to model assessments for SAS and crosslinking-MS datasets. Data quality and fit to model assessments for other datasets and model uncertainty are under development. Number of plots is limited to 256.*

Model Quality: MolProbity Analysis



### Ensemble information ?

*This entry consists of 0 distinct ensemble(s).*

### Summary ?

*This entry consists of 1 model(s). A total of 3 datasets were used to build this entry.*

### Representation ?

*This entry has 1 representation(s).*

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
1	1	1	Rev7-monomer	A	212	-	1-212	100.00 / 100.00	Atomic
				C					
		2	Rev3-RBM2	B	28	-	1-28	100.00 / 100.00	Atomic
				D					

## 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	<a href="#">SASDC29</a>
2	Experimental model	PDB	<a href="#">6BC8</a>
3	Mutagenesis data	Zenodo	<a href="#">10.5281/zenodo.1323686</a>

## Methodology and software ?

*This entry is a result of 1 distinct protocol(s).*

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

*There is 1 software package reported in this entry.*

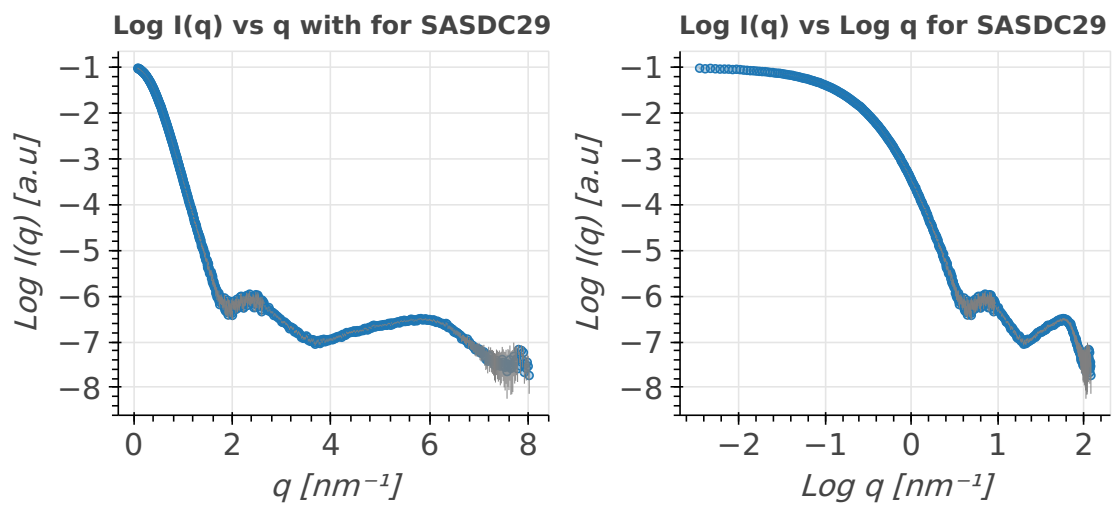
ID	Software name	Software version	Software classification	Software location
1	<a href="#">HADDOCK</a>	Not available	model building	<a href="http://haddock.science.uu.nl/services/HADDOCK/">http://haddock.science.uu.nl/services/HADDOCK/</a>

## 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.

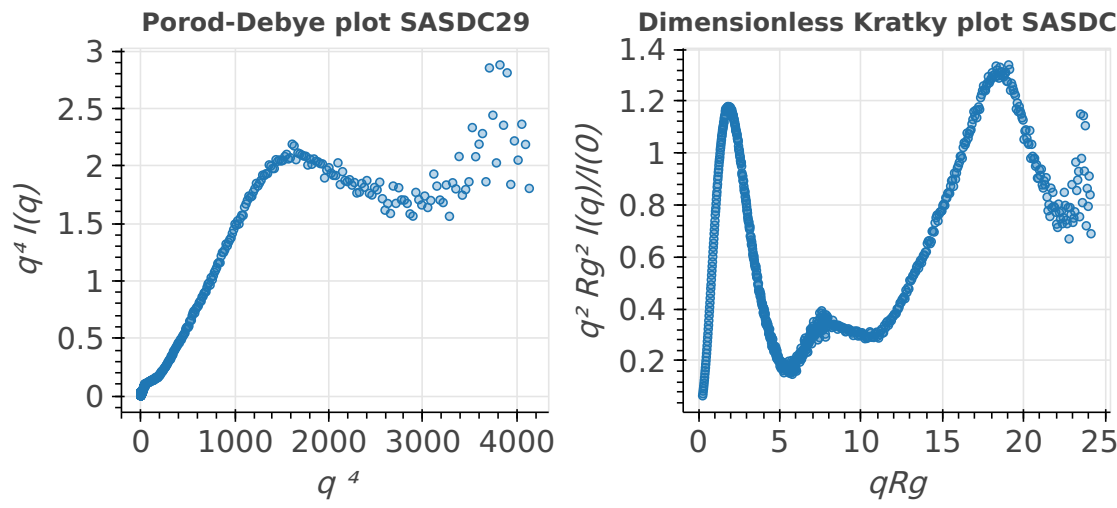
SASDB ID	Chemical composition MW	Standard MW	Porod Volume/MW
SASDC29	53.4 kDa	53.4 kDa	1.70 nm <sup>3</sup> /kDa

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	Not available	108.40 nm <sup>3</sup>	Not available	Not available	10.60 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.

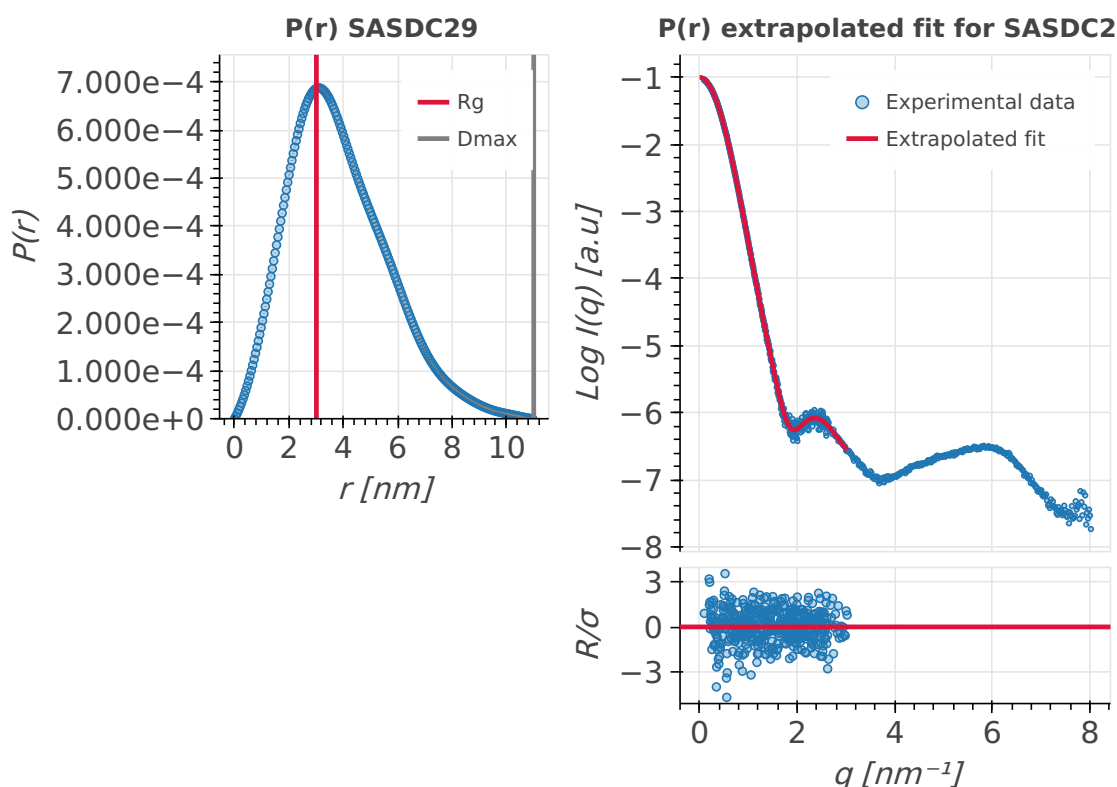


## Pair-distance distribution analysis ?

**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).  $R_g$  can be estimated from integrating the  $P(r)$  function. Agreement between the  $P(r)$  and Guinier-determined  $R_g$  (table below) is a good measure of the self-consistency of the SAS profile.  $R_g$  is a measure for the overall size of a macromolecule; e.g. a protein with a smaller  $R_g$  is more compact than a protein with a larger  $R_g$ , provided both have the same molecular weight (MW). The point where  $P(r)$  is decaying to zero is called  $D_{\max}$  and represents the maximum size of the particle.

SASDB ID	Software used	$D_{\max}$	$D_{\max}$ error	$R_g$	$R_g$ error
SASDC29	GNOM 5.0	11.000 nm	Not available	3.010 nm	0.004 nm

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



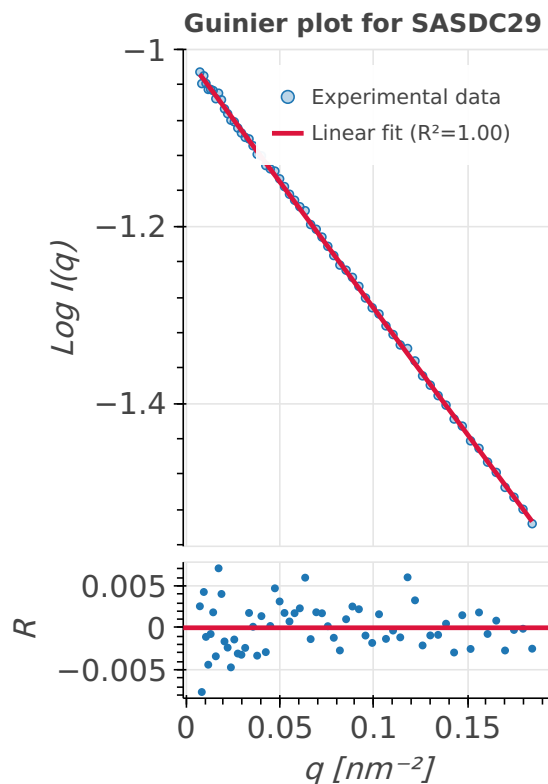
## Guinier analysis ?

**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	$R_g$	$R_g$ error	MW	MW error
SASDC29	2.93 nm	0.00 nm	53.4 kDa	0.0 kDa

**Guinier analysis:** 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 ?

For models with atomic structures, MolProbity analysis is performed. For models with coarse-grained or multi-scale structures, excluded volume analysis is performed.

Standard geometry: bond outliers ?

There are 22 bond length outliers in this entry (0.56% of 3941 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	210	LYS	C-N	10.11	1.47	1.33	1	1
C	210	LYS	C-N	10.10	1.47	1.33	1	1
B	1	MET	C-N	10.10	1.47	1.33	1	1
A	209	HIS	C-N	10.09	1.47	1.33	1	1
C	2	MET	C-N	10.07	1.47	1.33	1	1
C	3	THR	C-N	10.07	1.47	1.33	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	3	THR	C-N	10.07	1.47	1.33	1	1
A	2	MET	C-N	10.06	1.47	1.33	1	1
C	211	GLY	C-N	10.06	1.47	1.33	1	1
A	1	GLY	C-N	10.06	1.47	1.33	1	1
C	209	HIS	C-N	10.06	1.47	1.33	1	1
C	1	GLY	C-N	10.06	1.47	1.33	1	1
D	1	MET	C-N	10.04	1.47	1.33	1	1
B	2	GLU	C-N	10.04	1.47	1.33	1	1
C	208	ALA	C-N	10.04	1.47	1.33	1	1
A	208	ALA	C-N	10.04	1.47	1.33	1	1
A	211	GLY	C-N	10.01	1.47	1.33	1	1
D	2	GLU	C-N	10.01	1.47	1.33	1	1
C	155	GLU	CA-C	4.89	1.42	1.52	1	1
A	155	GLU	CA-C	4.89	1.42	1.52	1	1
C	209	HIS	CE1-NE2	4.14	1.36	1.32	1	1
A	209	HIS	CE1-NE2	4.08	1.36	1.32	1	1

### Standard geometry: angle outliers ?

There are 34 bond angle outliers in this entry (0.64% of 5344 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	209	HIS	ND1-CE1-NE2	9.68	118.08	108.40	1	1
C	209	HIS	ND1-CE1-NE2	9.65	118.05	108.40	1	1
B	1	MET	C-N-CA	8.58	106.25	121.70	1	1
D	1	MET	C-N-CA	8.58	106.26	121.70	1	1
C	211	GLY	C-N-CA	8.57	106.28	121.70	1	1
A	2	MET	C-N-CA	8.56	106.29	121.70	1	1
A	210	LYS	C-N-CA	8.56	106.30	121.70	1	1
C	2	MET	C-N-CA	8.56	106.30	121.70	1	1
C	1	GLY	C-N-CA	8.55	106.31	121.70	1	1
C	208	ALA	C-N-CA	8.55	106.31	121.70	1	1
A	211	GLY	C-N-CA	8.54	106.32	121.70	1	1
A	209	HIS	C-N-CA	8.54	106.33	121.70	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	1	GLY	C-N-CA	8.54	106.33	121.70	1	1
C	210	LYS	C-N-CA	8.53	106.35	121.70	1	1
C	209	HIS	C-N-CA	8.52	106.36	121.70	1	1
A	208	ALA	C-N-CA	8.52	106.37	121.70	1	1
A	3	THR	C-N-CA	6.79	109.48	121.70	1	1
B	2	GLU	C-N-CA	6.78	109.50	121.70	1	1
D	2	GLU	C-N-CA	6.77	109.51	121.70	1	1
C	3	THR	C-N-CA	6.77	109.51	121.70	1	1
C	209	HIS	CB-CG-CD2	6.58	122.65	131.20	1	1
A	209	HIS	CB-CG-CD2	6.52	122.73	131.20	1	1
A	209	HIS	CD2-NE2-CE1	6.41	102.59	109.00	1	1
C	209	HIS	CD2-NE2-CE1	6.39	102.61	109.00	1	1
A	209	HIS	CG-ND1-CE1	5.56	99.85	109.30	1	1
C	209	HIS	CG-ND1-CE1	5.54	99.88	109.30	1	1
C	209	HIS	ND1-CG-CD2	5.37	111.47	106.10	1	1
A	209	HIS	ND1-CG-CD2	5.37	111.47	106.10	1	1
A	7	ARG	CG-CD-NE	4.52	102.05	112.00	1	1
C	7	ARG	CG-CD-NE	4.50	102.09	112.00	1	1
C	97	GLU	CA-CB-CG	4.21	122.52	114.10	1	1
A	97	GLU	CA-CB-CG	4.20	122.50	114.10	1	1
A	155	GLU	CA-CB-CG	4.13	105.85	114.10	1	1
C	155	GLU	CA-CB-CG	4.11	105.87	114.10	1	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 atomic models in this entry.

Model ID	Clash score	Number of clashes
1	8.54	67

There are 67 clashes. The table below contains the detailed list of all clashes based on a MolProbity analysis. Bad clashes are  $\geq 0.4$  Angstrom.

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:47:LYS:NZ	C:140:HIS:NE2	1.30	1	1



Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
D:23:TRP:CZ2	D:28:GLU:OE1	1.23	1	1
B:24:LEU:O	D:27:LYS:HE3	1.22	1	1
B:23:TRP:CZ2	B:28:GLU:OE1	1.22	1	1
A:47:LYS:NZ	C:140:HIS:CE1	1.19	1	1
A:47:LYS:HZ1	C:140:HIS:CE1	1.18	1	1
A:133:VAL:HB	C:133:VAL:HG12	1.07	1	1
A:33:TYR:OH	C:40:VAL:HB	1.05	1	1
A:47:LYS:HZ1	C:140:HIS:CD2	1.03	1	1
B:23:TRP:CE2	B:28:GLU:OE1	0.98	1	1
D:23:TRP:CE2	D:28:GLU:OE1	0.96	1	1
B:27:LYS:HA	D:27:LYS:HE2	0.95	1	1
A:133:VAL:HG11	C:133:VAL:HA	0.95	1	1
B:24:LEU:O	D:27:LYS:CE	0.94	1	1
D:23:TRP:CZ2	D:28:GLU:CD	0.91	1	1
B:23:TRP:CZ2	B:28:GLU:CD	0.90	1	1
A:133:VAL:CB	C:133:VAL:HG12	0.87	1	1
A:133:VAL:CG1	C:133:VAL:HA	0.86	1	1
A:47:LYS:CE	C:140:HIS:CE1	0.83	1	1
B:27:LYS:HE3	D:26:ALA:C	0.82	1	1
B:27:LYS:HE3	D:26:ALA:O	0.80	1	1
A:36:GLU:OE2	C:45:LYS:CB	0.79	1	1
A:36:GLU:OE2	C:45:LYS:HB2	0.77	1	1
A:47:LYS:HE2	C:140:HIS:CE1	0.74	1	1
A:47:LYS:HZ3	C:140:HIS:CE1	0.73	1	1
A:133:VAL:HB	C:133:VAL:CG1	0.72	1	1
A:133:VAL:CG2	C:133:VAL:HG12	0.72	1	1
A:45:LYS:HE2	C:36:GLU:HB3	0.68	1	1
A:36:GLU:OE2	C:45:LYS:HB3	0.67	1	1
A:45:LYS:HE2	C:36:GLU:CB	0.64	1	1
B:27:LYS:HD3	D:27:LYS:HE2	0.57	1	1
A:49:TYR:HA	A:121:GLU:HG3	0.56	1	1
C:49:TYR:HA	C:121:GLU:HG3	0.55	1	1
A:133:VAL:CB	C:133:VAL:HA	0.55	1	1
D:23:TRP:CZ2	D:28:GLU:OE2	0.55	1	1
B:23:TRP:CZ2	B:28:GLU:OE2	0.54	1	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
B:23:TRP:CH2	B:28:GLU:OE1	0.54	1	1
D:23:TRP:CH2	D:28:GLU:OE1	0.53	1	1
A:36:GLU:CD	C:45:LYS:HB3	0.53	1	1
B:23:TRP:HZ2	B:28:GLU:OE2	0.51	1	1
A:33:TYR:CE2	C:33:TYR:OH	0.51	1	1
D:23:TRP:HZ2	D:28:GLU:OE2	0.49	1	1
A:133:VAL:HG12	C:132:SER:O	0.49	1	1
A:33:TYR:HE2	C:33:TYR:OH	0.49	1	1
A:116:LEU:HD23	A:197:ILE:HD11	0.47	1	1
A:33:TYR:HH	C:40:VAL:HB	0.46	1	1
A:135:ASP:O	C:52:PRO:HG2	0.46	1	1
C:116:LEU:HD23	C:197:ILE:HD11	0.45	1	1
A:24:LEU:O	A:28:VAL:HG23	0.45	1	1
C:24:LEU:O	C:28:VAL:HG23	0.45	1	1
B:23:TRP:HZ2	B:28:GLU:CD	0.44	1	1
A:193:MET:HB3	A:193:MET:HE2	0.44	1	1
B:25:GLN:C	B:27:LYS:H	0.44	1	1
C:98:LYS:O	C:204:VAL:HA	0.43	1	1
A:98:LYS:O	A:204:VAL:HA	0.43	1	1
D:25:GLN:C	D:27:LYS:H	0.43	1	1
C:158:THR:H	C:161:MET:HE2	0.42	1	1
A:158:THR:H	A:161:MET:HE2	0.42	1	1
B:27:LYS:HD3	D:27:LYS:CE	0.41	1	1
C:79:ASN:O	C:154:ARG:HD3	0.41	1	1
A:168:LYS:O	A:169:ASP:HB2	0.41	1	1
A:79:ASN:O	A:154:ARG:HD3	0.41	1	1
A:33:TYR:OH	C:33:TYR:HE1	0.41	1	1
D:2:GLU:O	D:3:ASP:OD1	0.41	1	1
A:133:VAL:HG23	C:133:VAL:HG12	0.41	1	1
B:2:GLU:O	B:3:ASP:OD1	0.41	1	1
C:168:LYS:O	C:169:ASP:HB2	0.41	1	1

### 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	472	464	6	2

There are 2 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	111	ILE	1
C	111	ILE	1

### Torsion angles : Protein sidechains ?

In the following table, sidechain rotameric 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	448	412	20	16

There are 16 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	4	THR	1
A	111	ILE	1
A	121	GLU	1
A	159	ARG	1
A	169	ASP	1
A	209	HIS	1
B	4	LYS	1
B	25	GLN	1
C	4	THR	1
C	111	ILE	1
C	121	GLU	1
C	159	ARG	1
C	169	ASP	1
C	209	HIS	1
D	4	LYS	1
D	25	GLN	1

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

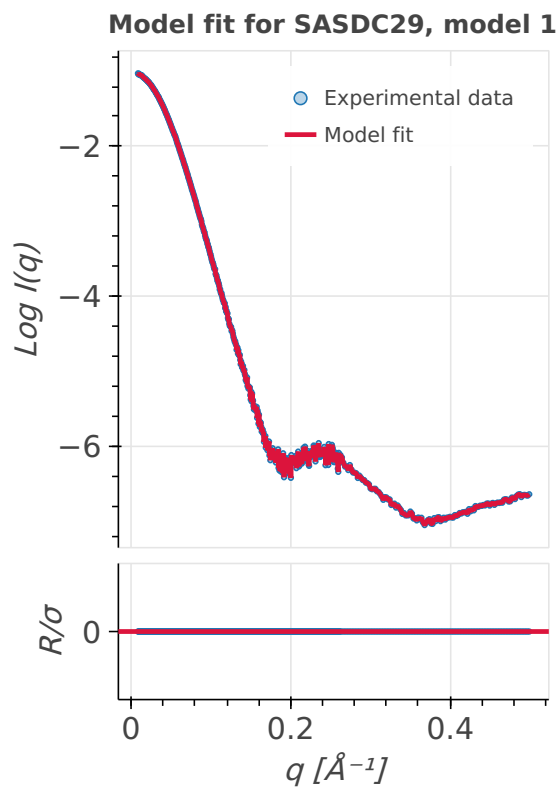
Fit of model(s) to SAS data

$\chi^2$  goodness of fit and cormap 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 1.0. 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	$\chi^2$	p-value
SASDC29	1	25.13	0.00

Model fit(s): Residual value plot is a measure to assess fit to the data. Residual values should be equally and randomly spaced around the horizontal axis.



Mutagenesis

Validation for this section is under development.

Fit of model to data used for validation ?

Validation for this section is under development.

*The 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 awards to the [PDB-IHM team](#) (DBI-1756248, DBI-2112966, DBI-2112967, DBI-2112968, and DBI-1756250) and awards from NSF, NIH, and DOE to the [RCSB PDB](#) (DBI-2321666, R01GM157729, and DE-SC0019749). The PDB-IHM team and members of the [Sali lab](#) contributed model validation metrics and software packages.*

*Dr. Jill Trewhella, Dr. Dina Schneidman, and members of the [SASBDB](#) repository are acknowledged for their advice and support in implementing SAS validation methods. Team members from the labs of Dr. Juri Rappsilber, Dr. Alexander Leitner, Dr. Andrea Graziadei, and members of [PRIDE](#) database are acknowledged for their advice and support in implementing crosslinking-MS validation methods. We are grateful to Dr. Shruthi Viswanath for discussions about uncertainty assessment of integrative structural models.*

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