

# Integrative Structure Validation Report

October 09, 2025 - 04:37 PM PDT

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

*IHMValidation Version 3.0*

*Python-IHM Version 2.5*

*MolProbity Version 4.5.2*

*ATSAS Version 3.2.1 (r14885)*

PDB ID	8ZZ9   pdb_00008zz9
PDB-Dev ID	PDBDEV_00000009
Structure Title	Structure of the human Rev7 dimer
Structure Authors	Rizzo, A.A.; Vassel, F.-M.; Chatterjee, N.; D'Souza, S.; Li, Y.; Hao, B.; Hemann, M.T.; Walker, G.C.; Korzhnev, D.M.
Deposited on	2017-11-16

*This is a PDB-IHM Structure Validation Report.*

*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](#).*

## 1. Overview

### 1.1. Summary

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

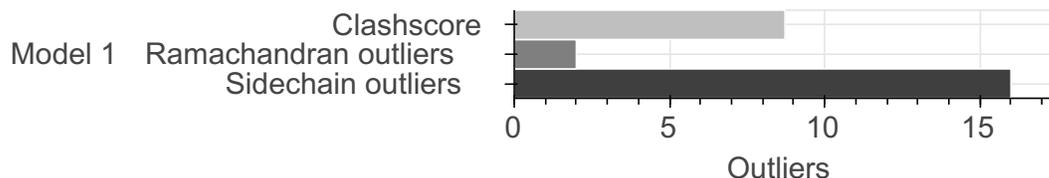
Name	Type	Count
Mutagenesis data	Experimental data	1
SAS data	Experimental data	1

Name	Type	Count
Experimental model	Starting model	1

## 1.2. Overall quality ?

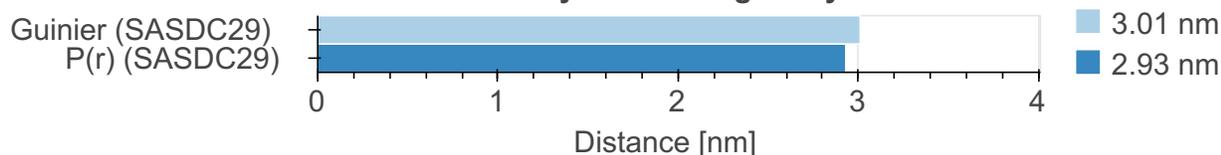
This validation report contains model quality assessments for all structures, data quality and fit to model assessments for SAS and crosslinking-MS datasets. Data quality and fit to model assessments for other datasets and model uncertainty are under development. Number of plots is limited to 256.

### Model Quality: MolProbity Analysis ?



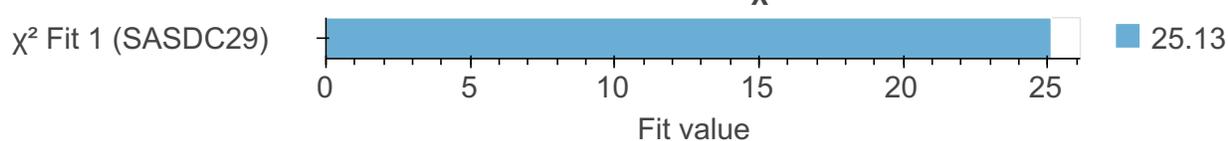
### Data Quality ?

#### Data Quality for SAS: Rg Analysis



### Fit to Data Used for Modeling ?

#### Fit to SAS Data: $\chi^2$ Fit



## 2. Model Details ?

### 2.1. Ensemble information ?

This entry consists of 0 distinct ensemble(s).

### 2.2. 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

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

### 2.3. 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="#">pdb_00006bc8</a>
3	Mutagenesis data	Zenodo	<a href="#">10.5281/zenodo.1323686</a>

### 2.4. 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	Not available	Not available	Not available	Not available	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>

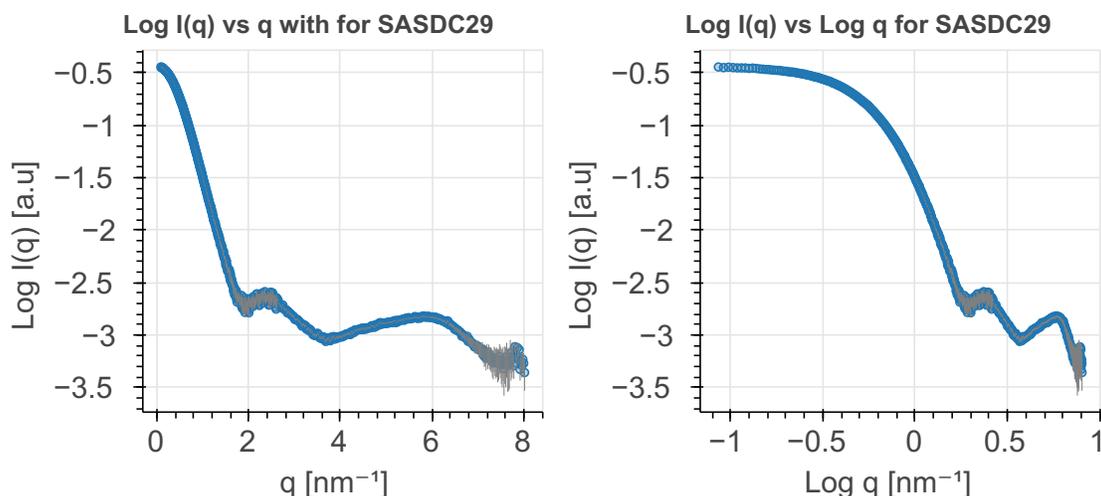
## 3. Data quality ?

### 3.1. SAS ?

#### 3.1.1. 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.



### 3.1.2. Key experimental estimates ?

Molecular weight (MW) estimates from experiments and analysis: Theoretical MW can be compared to SAS-derived values using the forward scatter ( $I(0)$ ) and the known concentration and partial specific volume of the scattering particle, or as estimated from the Porod volume and partial specific volume (Trehella et al., 2017, Trehella et al., 2023).

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

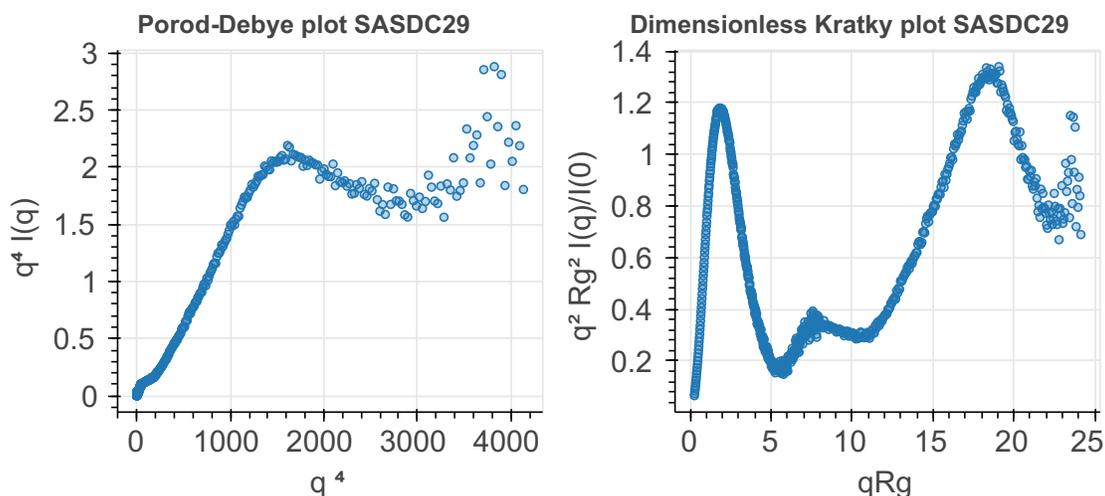
Volume estimates from experiments and analysis: estimated volume can be compared to Porod volume obtained from scattering profiles.

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

### 3.1.3. Flexibility analysis ?

In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas flexible-modular, fully unfolded domains or extended/stiff rodshaped domains lack a discernible plateau (Rambo and Tainer 2013). A bell-shaped Kratky plot ( $q^2I(q)$  vs.  $q$ ) with a well-defined maximum is observed for compact/folded structures. For partially flexible/modular or extended structures the Kratky plot can show multiple maxima and/or an increase in intensity at higher  $q$ -values depending on the degree of flexibility and extension. Fully intrinsically disordered structures yield a Kratky plot that systematically increases with increasing  $q$  values and will be near linear for highly extended molecules. The dimensionless Kratky plot ( $(qR_g)^2I(q)$  vs.  $qR_g$ ) is useful for quantifying differences in shape and foldedness among scattering objects of different sizes (Trehella et al., 2023).

Flexibility analysis for SASDC29

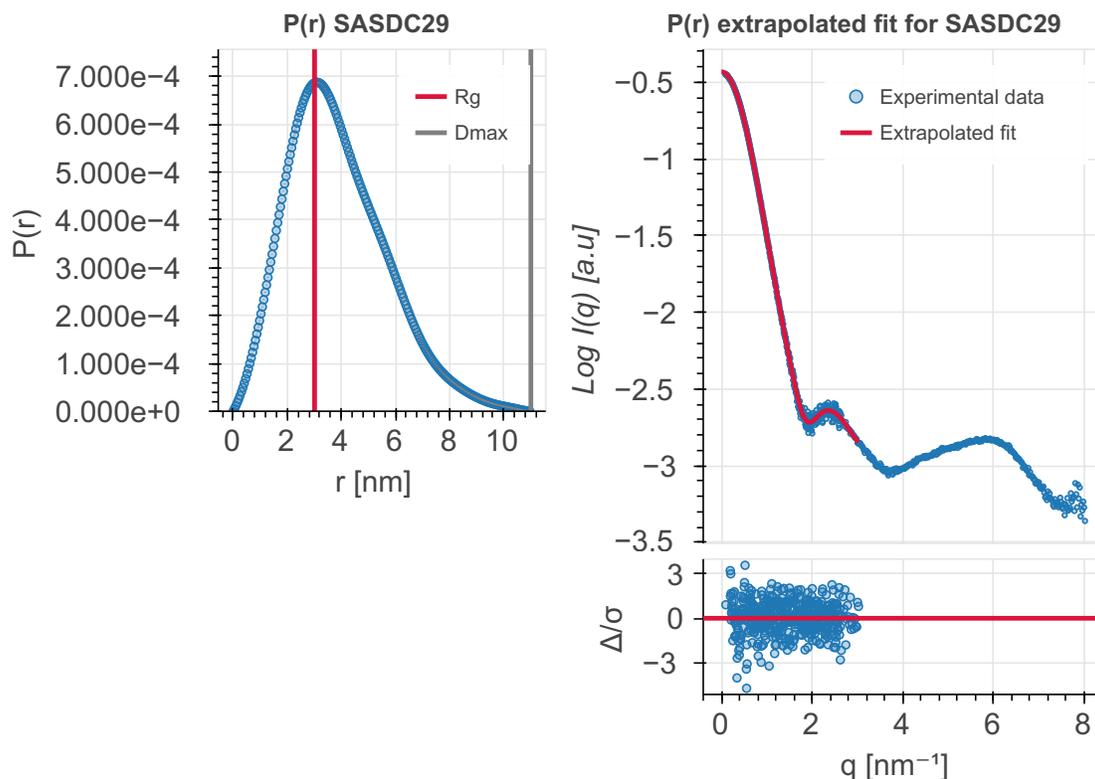


### 3.1.4. Pair-distance distribution analysis ?

The atom-pair distance distribution function (PDDF) or  $P(r)$  represents the distribution of distances between all pairs of atoms within the particle weighted by the respective scattering contrasts (Moore, 1980). The second moment of  $P(r)$  yields the radius of gyration ( $R_g$ ), which is a measure of the overall size and shape of a macromolecule (i.e. the spatial distribution of volume elements). 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.

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

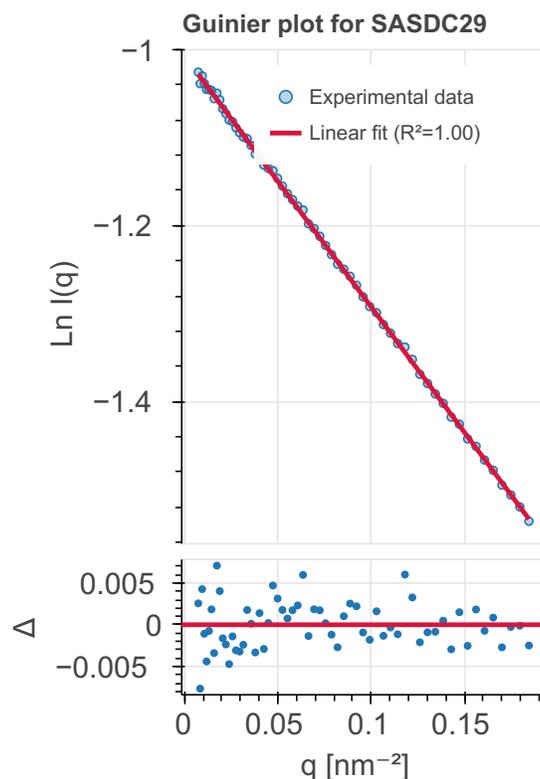
$P(r)$  for SASDC29: The value of  $P(r)$  should be zero beyond  $r=D_{\max}$ .



### 3.1.5. Guinier analysis ?

The linearity of the Guinier plot ( $\ln(q)$  vs.  $q^2$ ) at very-low angle ( $qR_g < 1.3$ ) is a sensitive indicator of the quality of the sample in relation to its homogeneity; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the same size. Deviations from linearity can point to strong interference effects from particle attraction or repulsion, polydispersity of the samples, or improper background subtraction (Feigin et al., 2013). Residual difference plots and Pearson correlation coefficient determination ( $R^2$ ) are measures to assess quality of the linear fit to the Guinier region. A perfect fit has an  $R^2$  value of 1. Residual values should be equally and randomly spaced around the horizontal axis with no evident systematic upward or downward curvature.

SASDB ID	$R_g$	$R_g$ error	MW	MW error
SASDC29	2.93 nm	0.00 nm	53.4 kDa	0.0 kDa



### 3.4. Mutagenesis ?

Validation for this section is under development.

## 4. 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.

### 4.1b. MolProbity Analysis ?

*Excluded volume satisfaction for the models in the entry are listed below. The Analysed column shows the number of particle-particle or particle-atom pairs for which excluded volume was analysed.*

Standard geometry: bond outliers ?

There are 22 bond length outliers in this entry (0.55% of 3980 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
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 (0.63% of 5408 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

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
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
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	8.72	69

There are 69 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
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
C:44:GLN:CD	C:57:CYS:SG	0.62	1	1
A:44:GLN:CD	A:57:CYS:SG	0.62	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

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
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
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	454	416	22	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

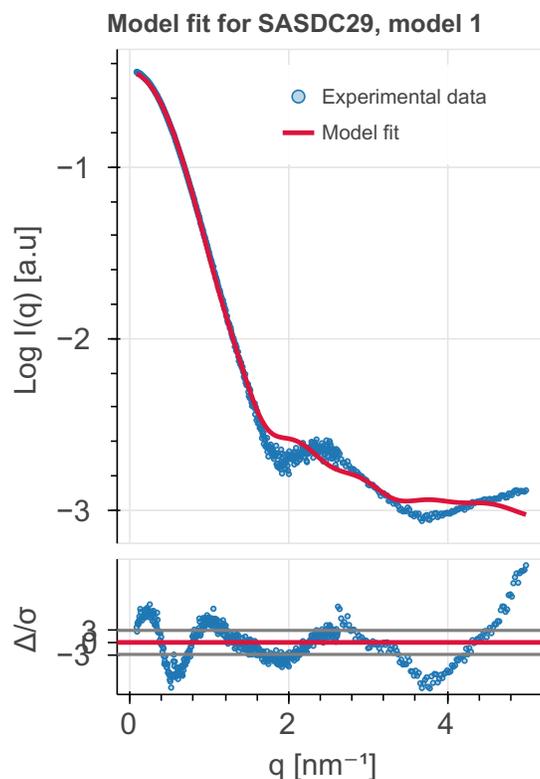
## 5. Fit to Data Used for Modeling Assessment ?

## 5.1. SAS ?

Model and fits displayed below were obtained from SASBDB.

### 5.1.1 Model versus Experimental Scattering Profiles ?

Experimental (blue) and model (red) scattering profiles are presented as  $\log I(q)$  vs.  $q$  together with error weighted residual difference plot between the experimental and model  $I(q)$  vs  $q$ . The  $I(q)$  is the intensity (preferably in absolute units of  $\text{cm}^{-1}$  or arbitrary units) and  $q$  is the modulus of the scattering vector. For a good fit, residual values should be equally and randomly spaced around the horizontal axis largely within  $\pm$  three standard deviations.



### 5.1.2 $\chi^2$ goodness of fit and 5.1.3 cormap analysis ?

$\chi^2$  values are a measure of the overall fit of the model to the 1D scattering profile. A model that fits the data within its error estimates will have a  $\chi^2$  value close to one, provided that the dominant errors are the random statistical errors (i.e. no systematic errors) from the SAS measurement that are correctly propagated. Correlation Map (CorMap) test is a variance-covariance analysis on the scattering intensities comparing two (or more) scattering profiles (e.g. model versus experiment or multiple measures from the same sample). The CorMap test complements  $\chi^2$  and importantly is independent of the reported errors. The method assigns a probability (P-value based on a 1-tailed Schilling test) for finding the longest string of experimental data points that lie systematically above (+1) or below (-1) the model profile. The P-value lies between 0 – 1 and a significance threshold is chosen below which the model fit is judged to show systematic deviation from experiment. A typical range statisticians use to indicate significant deviation is 0.01 - 0.05. As implemented in the ATSAS suite, the reported CorMap P-value is green (model fit is good) for  $P > 0.05$ , yellow for  $0.01 < P < 0.05$ , and red (model deviates significantly) for  $P < 0.01$ .

SASDB ID	Model	$\chi^2$	P-value
SASDC29	1	25.13	0.00

## 5.4. Mutagenesis ?

Validation for this section is under development.

---

## 6. Fit to Data Used for Validation Assessment

Validation for this section is under development.

---

### *Acknowledgments*

*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 Trehwella, 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.*