Integrative Structure Validation Report

July 22, 2024 - 03:31 PM PDT

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

Python-IHM Version 1.3
MolProbity Version 4.5.2
ATSAS Version 3.2.1 (r14885)
Integrative Modeling Validation Version 1.2

PDB ID	8ZZ4
PDB-Dev ID	PDBDEV_00000004
Structure Title	Structure of K63-linked Diubiquitin
Structure Authors	Liu Z; Gong Z; Cao Y; Ding YH; Dong MQ; Lu YB; Zhang WP; Tang C

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

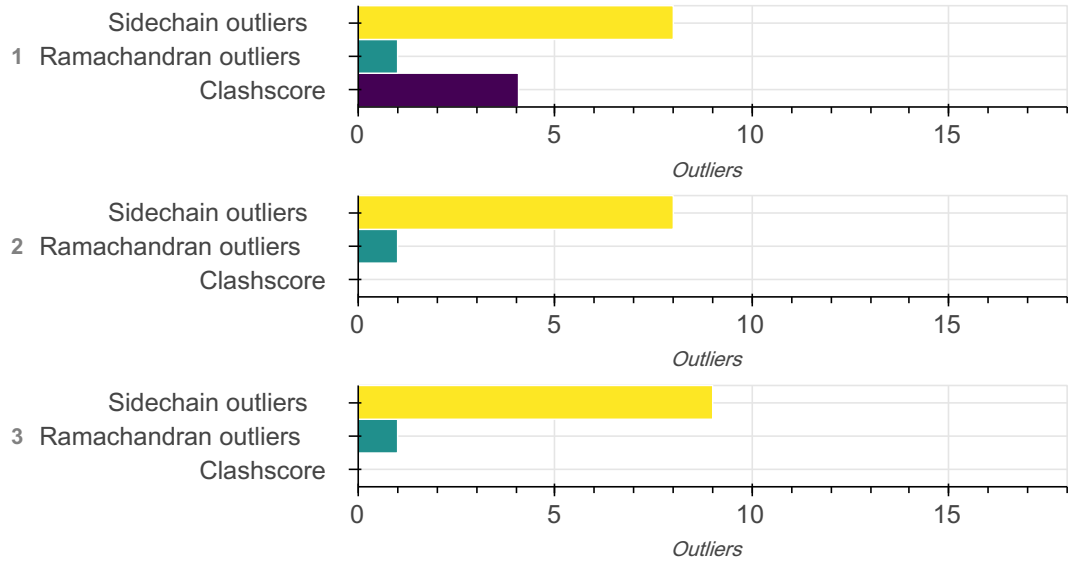
A user guide is available at https://pdb-dev.wwpdb.org/validation_help.html with specific help available everywhere you see the ? symbol.

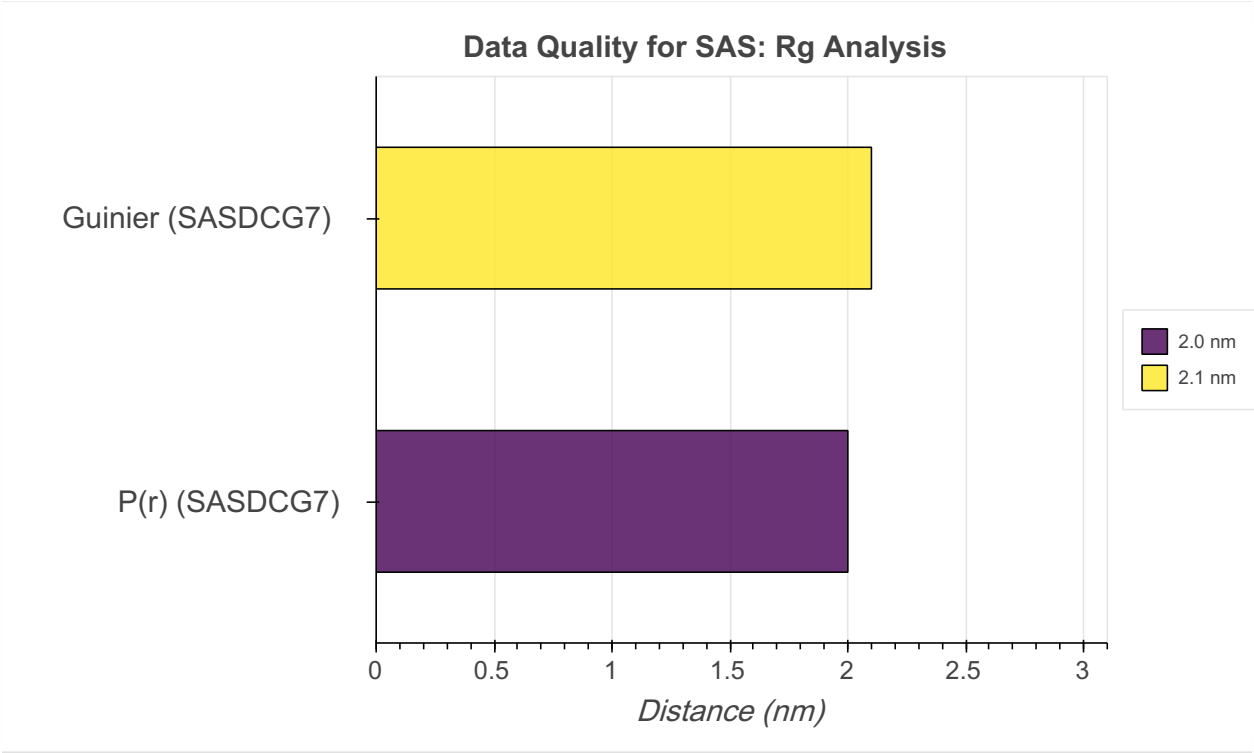
List of references used to build this report is available here.

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. 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 3 unique models, with 2 subunits in each model. A total of 5 datasets or restraints were used to build this entry. Each model is represented by 0 rigid bodies and 2 flexible or non-rigid units.

Entry composition

There are 3 unique types of models in this entry.These models are titled None/Model 1, None/Model 2, None/Model 3 respectively.

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Chain ID [auth]	Total residues
1	1	1	Ubiquitin	A	A	76
1	2	1	Ubiquitin	B	B	76
2	1	1	Ubiquitin	A	A	76
2	2	1	Ubiquitin	B	B	76
3	1	1	Ubiquitin	A	A	76
3	2	1	Ubiquitin	B	B	76

Datasets used for modeling

There are 5 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	SAS data	SASBDB	SASDCG7
2	Experimental model	PDB	1UBQ
3	Experimental model	PDB	2N2K
4	Crosslinking-MS data	File	10.5281/zenodo.1006721
5	Single molecule FRET data	File	10.5281/zenodo.1006721

Representation

This entry has only one representation and includes 0 rigid bodies and 2 flexible units

Chain ID	Rigid bodies	Non-rigid segments
A	-	1-76

Chain ID	Rigid bodies	Non-rigid segments
B	-	1-76

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	True	False

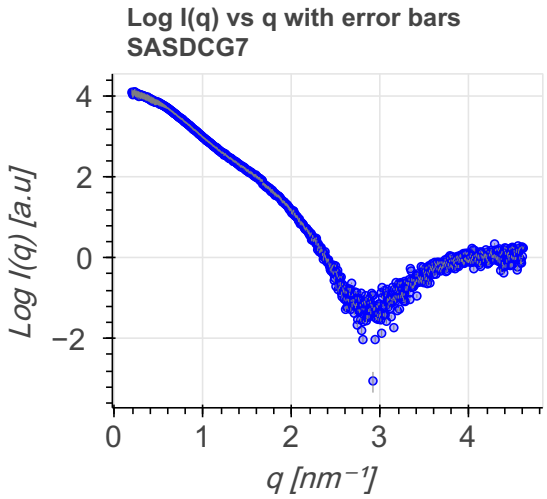
Software packages used for modeling were either not reported or not used.

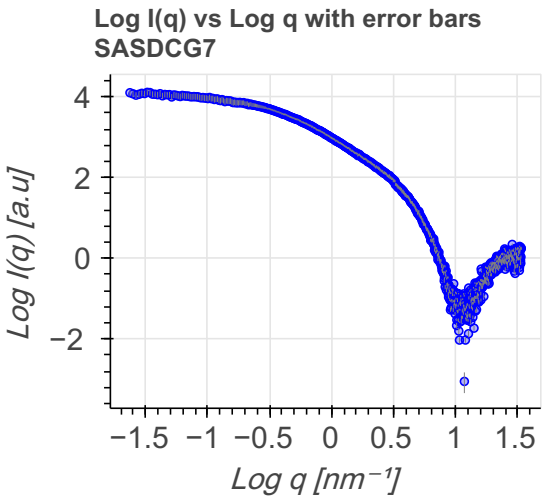
Data quality

Scattering profile

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

Scattering profile for SASDCG7: 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.

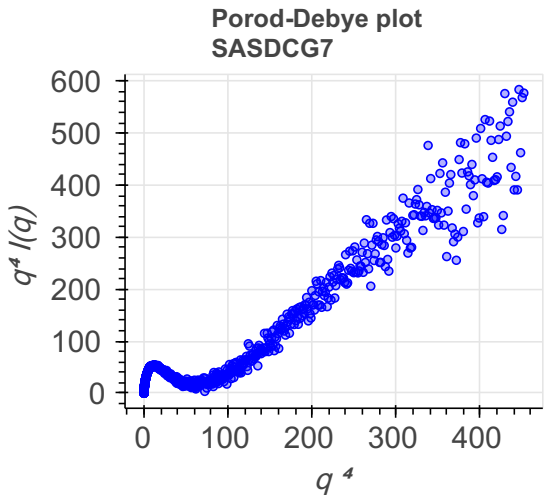
SASDB ID	Chemical composition MW	Standard MW	Porod Volume/MW
SASDCG7	13.0 kDa	N/A	1.69 nm³/kDa

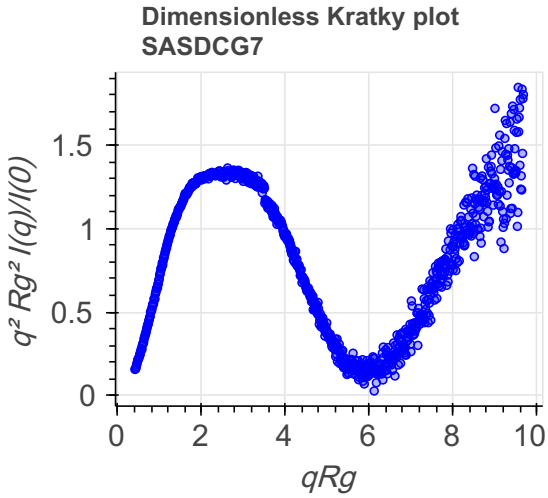
Volume estimates from experiments and analysis: estimated volume can be compared to Porod volume obtained from scattering profiles.

SASDB ID	Estimated Volume	Porod Volume	Specific Volume	Sample Contrast	Sample Concentration
SASDCG7	N/A	22.00 nm³	N/A	N/A	2.60 mg/mL

Flexibility analysis

Flexibility analysis for SASDCG7: 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.



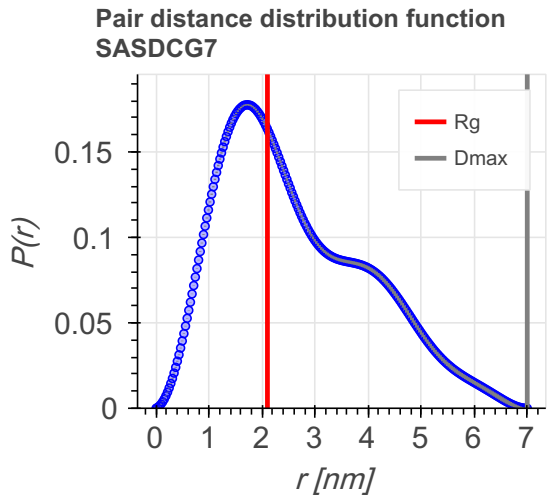


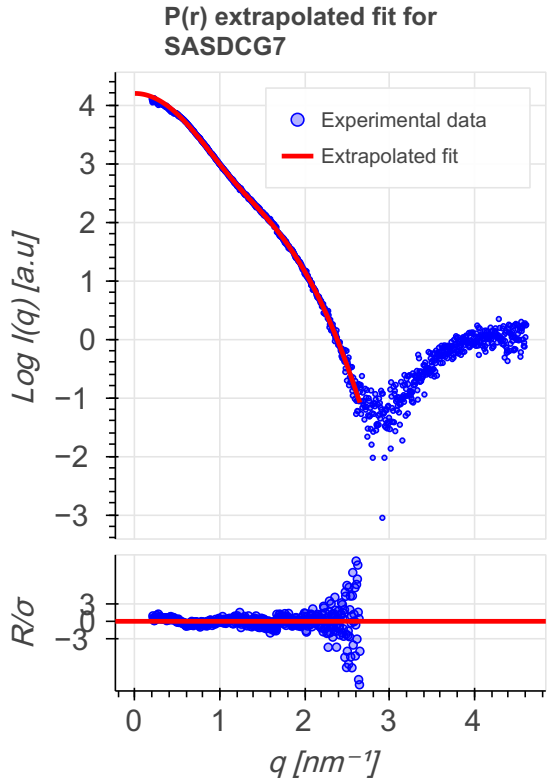
Pair-distance distribution analysis

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	Dmax	Dmax error	Rg	Rg error
SASDCG7	GNOM 5.0	7.000 nm	N/A	2.100 nm	0.010 nm

P(r) for SASDCG7: The value of P(r) should be zero beyond r=D_{max}.





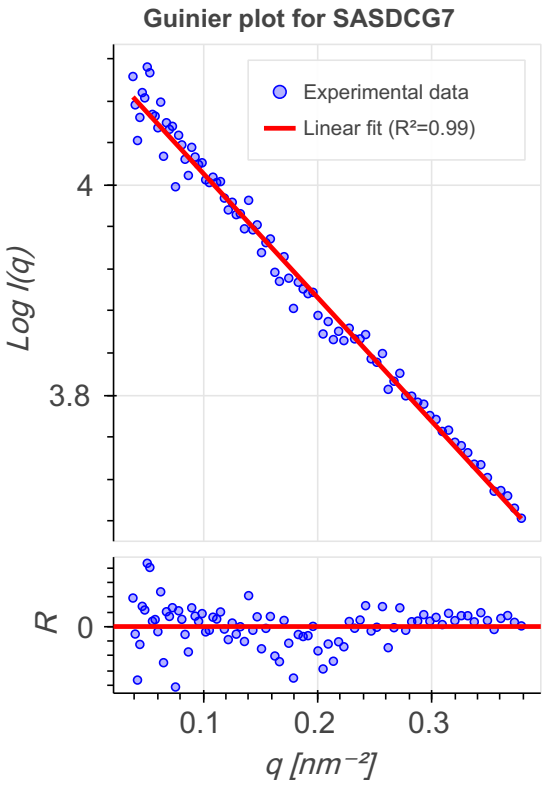
Guinier analysis

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	Rg	Rg error	MW	MW error
SASDCG7	2.00 nm	0.15 nm	N/A	N/A

Guinier analysis for SASDCG7: 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.



Single molecule FRET

Validation for this section is under development.

Crosslinking-MS

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 3774 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
CD--HD3	1.06	0.97	18
ND2--HD21	0.96	0.86	6

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
CD--HD3	1.07	0.97	6
NE2--HE21	0.96	0.86	11
CB--HB3	1.07	0.97	14
CB--HB2	1.07	0.97	19
CG--HG2	1.07	0.97	3
CB--HB3	1.08	0.97	297
CB--HB2	1.08	0.97	287
CA--HA	1.08	0.97	409
CG--HG2	1.08	0.97	159
NE2--HE21	0.97	0.86	7
CA--HA2	1.08	0.97	36
CA--HA3	1.08	0.97	36
CD--HD3	1.08	0.97	60
CG--HG3	1.08	0.97	162
CD--HD2	1.08	0.97	60
CE--HE2	1.08	0.97	48
CG--HG	1.08	0.97	54
CD2--HD21	1.08	0.97	54
CD1--HD12	1.08	0.97	96
CG2--HG21	1.08	0.97	108
CG1--HG12	1.08	0.97	66
CG2--HG23	1.08	0.97	108
CD2--HD23	1.08	0.97	54

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
CG1--HG13	1.08	0.97	66
CG2--HG22	1.08	0.97	108
CD1--HD13	1.08	0.97	96
CE--HE3	1.08	0.97	48
CD1--HD11	1.08	0.97	96
CD2--HD22	1.08	0.97	54
CG1--HG11	1.08	0.97	24
CB--HB1	1.08	0.97	12
CB--HB	1.08	0.97	102
CE--HE1	1.08	0.97	6
CB--HB	1.09	0.97	6
CA--HA	1.09	0.97	11
CB--HB3	1.09	0.97	1
NH1--HH11	0.98	0.86	6
CB--HB2	1.09	0.97	6
N--H	0.98	0.86	432
OH--HH	0.96	0.84	6
NE2--HE2	0.98	0.86	6
NE2--HE21	0.98	0.86	18
ND2--HD22	0.98	0.86	6
OG1--HG1	0.96	0.84	42
ND2--HD21	0.98	0.86	6
OG--HG	0.96	0.84	18

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
ND1--HD1	0.98	0.86	6
NE2--HE22	0.98	0.86	24
CD--HD2	1.09	0.97	6
NH2--HH21	0.98	0.86	1
NE2--HE22	0.99	0.86	10
NH2--HH21	0.99	0.86	5
CD--HD2	1.10	0.97	18
NH1--HH11	0.99	0.86	6
NE2--HE22	1.00	0.86	2
ND2--HD22	1.00	0.86	6
NH1--HH11	1.00	0.86	12
NH2--HH22	1.00	0.86	18
NH2--HH21	1.00	0.86	18
NH1--HH12	1.00	0.86	18
NZ--HZ1	1.04	0.89	42
NZ--HZ3	1.04	0.89	39
NH2--HH22	1.01	0.86	6
NE--HE	1.01	0.86	24
N--H3	1.04	0.89	6
CD1--HD1	1.08	0.93	18
CZ--HZ	1.08	0.93	12
CD2--HD2	1.08	0.93	24
CE2--HE2	1.08	0.93	18

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
N--H1	1.04	0.89	6
CE1--HE1	1.08	0.93	24
NZ--HZ2	1.04	0.89	39
N--H2	1.04	0.89	6
NH1--HH12	1.02	0.86	6

Standard geometry: angle outliers

There are 104 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
CD-NE-CZ	124.40	136.01	1
CD-NE-CZ	124.40	135.97	2
CD-NE-CZ	124.40	135.95	1
CD-NE-CZ	124.40	135.94	1
CD-NE-CZ	124.40	135.87	1
CA-CB-OG1	109.60	98.60	1
CA-CB-OG1	109.60	98.62	2
CA-CB-OG1	109.60	98.64	2
CA-CB-OG1	109.60	98.67	1
CA-CB-CG	112.60	105.31	1
CA-CB-CG	112.60	105.35	1
CA-CB-CG	112.60	105.36	1
CA-CB-CG	112.60	105.38	2
CA-CB-CG	112.60	105.41	1
CA-CB-CG	113.80	107.40	3

Angle type	Observed angle (°)	Ideal angle (°)	Number of outliers
CA-CB-CG	113.80	107.41	1
CA-CB-CG	113.80	107.47	1
CA-CB-CG	113.80	107.48	1
CD-NE-CZ	124.40	116.33	1
CD-NE-CZ	124.40	116.38	1
CD-NE-CZ	124.40	116.40	1
CD-NE-CZ	124.40	116.47	1
CD-NE-CZ	124.40	116.48	1
CD-NE-CZ	124.40	116.49	1
CA-CB-OG1	109.60	101.28	2
CA-CB-OG1	109.60	101.31	1
CA-CB-OG1	109.60	101.35	1
CA-CB-OG1	109.60	101.36	1
CA-CB-OG1	109.60	101.40	1
CA-CB-OG1	109.60	102.15	2
CA-CB-OG1	109.60	102.16	2
CA-CB-OG1	109.60	102.21	1
CA-CB-OG1	109.60	102.22	4
CA-CB-OG1	109.60	102.27	1
CG-CD-OE1	118.40	129.63	1
CA-CB-OG1	109.60	102.28	1
CG-CD-OE1	118.40	129.61	1
CA-CB-OG1	109.60	102.30	1

Angle type	Observed angle (°)	Ideal angle (°)	Number of outliers
CG-CD-OE1	118.40	129.60	2
CG-CD-OE1	118.40	129.59	1
CG-CD-OE1	118.40	129.54	1
CG-CD-OE1	118.40	107.49	1
CG-CD-OE1	118.40	107.53	3
CG-CD-OE1	118.40	107.54	1
O-C-N	123.00	130.54	1
CG-CD-OE1	118.40	107.56	1
O-C-N	123.00	130.51	1
O-C-N	123.00	130.47	3
O-C-N	123.00	130.45	1
CA-CB-CG	112.60	117.03	1
CA-C-O	120.80	113.30	1
CA-C-O	120.80	113.32	2
CA-C-O	120.80	113.33	1
CA-C-O	120.80	113.37	1
CA-C-O	120.80	113.38	3
CA-CB-CG	112.60	116.96	1
NE-CZ-NH1	121.50	125.85	1
CA-C-O	120.80	113.40	1
CA-C-O	120.80	113.42	3
CA-CB-CG	112.60	116.91	2
NE-CZ-NH1	121.50	125.80	1

Angle type	Observed angle (°)	Ideal angle (°)	Number of outliers
CA-CB-CG	112.60	116.90	1
NE-CZ-NH1	121.50	125.77	3
CA-CB-CG	112.60	116.87	1
CA-CB-CG	112.60	108.34	1
CA-CB-CG	112.60	108.35	1
NE-CZ-NH1	121.50	125.75	1
CA-CB-CG	112.60	108.38	1
CA-CB-CG	112.60	108.40	2
CA-CB-CG	112.60	108.47	1
CA-CB-CG	114.10	122.20	2
O-C-N	123.00	129.48	1
O-C-N	123.00	129.46	1
CA-CB-CG	114.10	122.15	4

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
1	4.07	10
2	0.00	0
3	0.00	0

All 10 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:42:ARG:HE	A:49:GLN:NE2	0.602
1	B:42:ARG:HE	B:49:GLN:NE2	0.587

Model ID	Atom-1	Atom-2	Clash overlap (Å)
1	A:42:ARG:HE	A:49:GLN:HE21	0.529
1	B:42:ARG:HE	B:49:GLN:HE21	0.515
1	A:23:ILE:HB	A:52:ASP:HA	0.463
1	B:23:ILE:HB	B:52:ASP:HA	0.456
1	B:62:GLN:HB3	B:62:GLN:HE21	0.427
1	B:71:LEU:HA	B:71:LEU:HD12	0.415
1	B:72:ARG:O	B:73:LEU:O	0.404
1	B:26:VAL:HG21	B:56:LEU:HD21	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	Analyzed	Favored	Allowed	Outliers
1	148	144	3	1
2	148	146	1	1
3	148	145	2	1

Detailed list of outliers are tabulated below.

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	136	116	12	8
2	136	116	12	8
3	136	114	13	9

Detailed list of outliers are tabulated below.

Model ID	Chain	Residue ID	Residue type
1	A	13	ILE

Model ID	Chain	Residue ID	Residue type
1	A	15	LEU
1	A	39	ASP
1	A	71	LEU
1	B	13	ILE
1	B	15	LEU
1	B	39	ASP
1	B	71	LEU
2	A	13	ILE
2	A	15	LEU
2	A	39	ASP
2	A	71	LEU
2	B	13	ILE
2	B	15	LEU
2	B	39	ASP
2	B	71	LEU
3	A	13	ILE
3	A	15	LEU
3	A	39	ASP
3	A	71	LEU
3	A	74	ARG
3	B	13	ILE
3	B	15	LEU
3	B	39	ASP

Model ID	Chain	Residue ID	Residue type
3	B	71	LEU

Fit of model to data used for modeling

Fit of model(s) to SAS data

χ^2 goodness of fit and cormap analysis

Model(s) and/or fit for this entry have not been deposited.

Single molecule FRET

Validation for this section is under development.

Crosslinking-MS

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 labcontributed model validation metrics and software packages.

Implementation of validation methods for SAS data and SAS-based models are funded byRCSB PDB (grant number DBI-1832184). Dr. Stephen Burley, Dr. John Westbrook, and Dr. Jasmine Young from RCSB PDB, Dr. Jill Trewhella, 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.