Integrative Structure Validation Report

February 18, 2025 - 08:42 AM PST

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
pyHMMER Version 0.11.0

PDB ID	9A68
PDB-Dev ID	PDBDEV_00000301
Structure Title	Integrative model of PARC-PARE by crosslinking MS and deep learning
Structure Authors	Kolja Stahl; Oliver Brock; Juri Rappsilber
Deposited on	2024-01-23

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

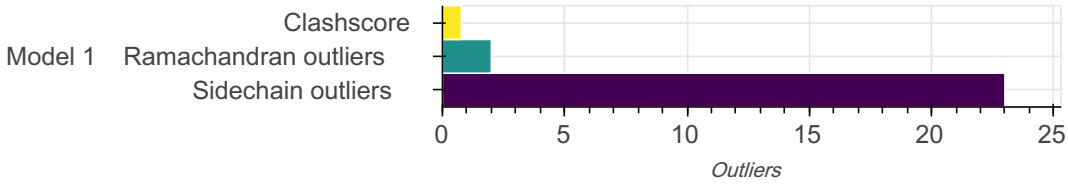
A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the ? symbol.

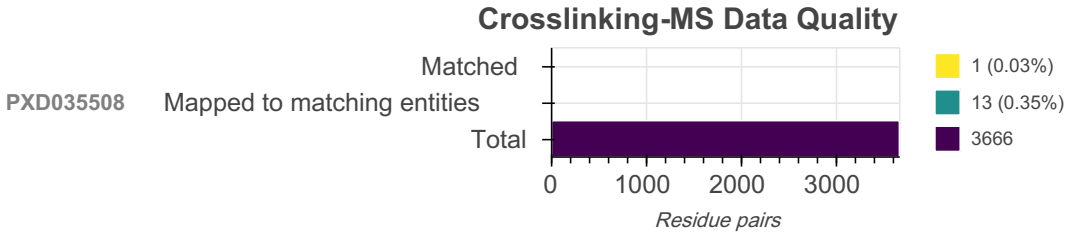
List of references used to build this report is available here.

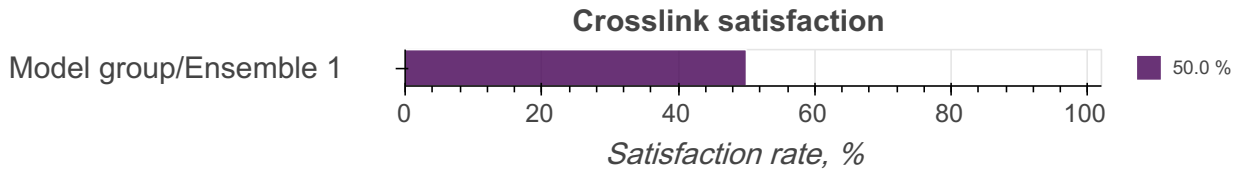
Overall quality

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 1 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	PARC_BACSU	A	806	-	1-806	100.00 / 0.00	Atomic
		2	PARE_BACSU	B	655	-	1-655	100.00 / 0.00	Atomic

Datasets used for modeling

There is 1 unique dataset used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	Crosslinking-MS data	PRIDE	PXD035508

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	AlphaLink2	AlphaLink2	None	1	False	False

There is 1 software package reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	AlphaLink2	1.00	model building	https://github.com/Rappsilber-Laboratory/AlphaLink2

Data quality

Crosslinking-MS

At the moment, data validation is only available for crosslinking-MS data deposited as a fully compliant dataset in the PRIDE Crosslinking database. Correspondence between crosslinking-MS and entry entities is established using pyHMMER. Only residue pairs that passed the reported threshold are used for the analysis. The values in the report have to be interpreted in the context of the experiment (i.e. only a minor fraction of in-situ or in-vivo dataset can be used for modeling).

Crosslinking-MS dataset (PRIDE ID)	PXD035508
Number of entities in the crosslinking-MS dataset:	810
Number of entities in the entry:	2
Matching entities:	

Entity ID	Molecule name	Crosslinking-MS Entity ID	E-value	Exact match
1	PARC_BACSU	dbseq_Q45066_target	0.00	True
2	PARE_BACSU	dbseq_Q59192_target	0.00	True

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A68	2	2 (100.00%)	1 (50.00%)
PXD035508	3666	13 (0.35%)	1 (0.03%)

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 no bond length outliers.

Standard geometry: angle outliers

There are 47 bond angle outliers in this entry (0.30% of 15826 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	29	GLN	OE1-CD-NE2	5.48	117.12	122.60	1	1
A	806	GLN	OE1-CD-NE2	5.32	117.28	122.60	1	1
A	93	GLN	OE1-CD-NE2	5.13	117.47	122.60	1	1
A	75	ASN	OD1-CG-ND2	5.12	117.48	122.60	1	1
A	537	ASP	CA-CB-CG	5.05	117.65	112.60	1	1
B	586	GLN	OE1-CD-NE2	5.00	117.60	122.60	1	1
A	426	GLN	OE1-CD-NE2	4.91	117.69	122.60	1	1
A	256	GLN	OE1-CD-NE2	4.86	117.74	122.60	1	1
B	52	ASP	CA-CB-CG	4.81	117.41	112.60	1	1
A	198	GLN	OE1-CD-NE2	4.74	117.86	122.60	1	1
A	479	ASP	CA-CB-CG	4.54	117.14	112.60	1	1
B	214	GLN	OE1-CD-NE2	4.53	118.07	122.60	1	1
A	702	GLN	OE1-CD-NE2	4.51	118.09	122.60	1	1
B	439	GLN	OE1-CD-NE2	4.50	118.10	122.60	1	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	648	HIS	CB-CG-CD2	4.48	125.38	131.20	1	1
B	383	GLN	OE1-CD-NE2	4.46	118.14	122.60	1	1
A	578	GLN	OE1-CD-NE2	4.41	118.19	122.60	1	1
A	31	ARG	NE-CZ-NH2	4.41	123.17	119.20	1	1
A	3	GLN	OE1-CD-NE2	4.41	118.19	122.60	1	1
A	31	ARG	NH1-CZ-NH2	4.36	113.63	119.30	1	1
B	334	GLN	OE1-CD-NE2	4.31	118.29	122.60	1	1
B	86	HIS	CB-CG-CD2	4.28	125.63	131.20	1	1
B	75	GLN	OE1-CD-NE2	4.27	118.33	122.60	1	1
B	6	GLN	OE1-CD-NE2	4.26	118.34	122.60	1	1
B	338	GLN	OE1-CD-NE2	4.25	118.35	122.60	1	1
A	756	GLN	OE1-CD-NE2	4.25	118.35	122.60	1	1
B	413	THR	CA-CB-OG1	4.23	103.26	109.60	1	1
B	473	ASN	OD1-CG-ND2	4.20	118.40	122.60	1	1
A	783	ASN	OD1-CG-ND2	4.20	118.40	122.60	1	1
A	701	GLN	OE1-CD-NE2	4.19	118.41	122.60	1	1
B	227	PHE	CA-CB-CG	4.18	117.98	113.80	1	1
A	801	ARG	NH1-CZ-NH2	4.17	113.88	119.30	1	1
B	105	LYS	C-N-CA	4.16	129.19	121.70	1	1
B	67	HIS	CB-CG-CD2	4.13	125.84	131.20	1	1
B	282	HIS	CB-CG-CD2	4.13	125.84	131.20	1	1
B	15	GLN	OE1-CD-NE2	4.12	118.48	122.60	1	1
B	632	ASN	OD1-CG-ND2	4.12	118.48	122.60	1	1
A	46	ARG	CD-NE-CZ	4.11	130.15	124.40	1	1
A	291	ARG	NH1-CZ-NH2	4.11	113.96	119.30	1	1
B	249	HIS	CB-CG-CD2	4.09	125.88	131.20	1	1
A	529	GLN	OE1-CD-NE2	4.08	118.52	122.60	1	1
B	108	GLN	OE1-CD-NE2	4.07	118.53	122.60	1	1
B	609	ARG	NH1-CZ-NH2	4.07	114.01	119.30	1	1
A	637	ASN	OD1-CG-ND2	4.06	118.54	122.60	1	1
A	711	GLN	OE1-CD-NE2	4.05	118.55	122.60	1	1
B	646	ASN	OD1-CG-ND2	4.03	118.57	122.60	1	1
B	5	GLN	OE1-CD-NE2	4.03	118.57	122.60	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	0.78	18

There are 18 clashes. The table below contains the detailed list of all clashes based on a MolProbity analysis. Bad clashes are >= 0.4 Angstrom.

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:73:ILE:HD11	A:81:ASP:HA	0.66	1	1
A:381:LEU:HD22	A:425:LEU:CD1	0.56	1	1
A:357:LYS:HE2	A:463:LEU:HD21	0.55	1	1
A:26:TYR:CB	B:505:THR:HG21	0.54	1	1
A:700:LEU:HD22	A:720:LEU:HD11	0.53	1	1
A:787:PHE:CE2	A:788:PHE:CD2	0.48	1	1
B:206:GLU:CD	B:215:ARG:HH21	0.48	1	1
A:759:MET:HE2	A:802:LEU:HD11	0.46	1	1
A:368:LEU:HD11	A:372:LYS:HE3	0.46	1	1
A:661:HIS:CE1	A:745:ARG:HB3	0.44	1	1
A:338:ARG:HH12	B:637:LEU:HD21	0.44	1	1
B:99:VAL:HB	B:101:HIS:CE1	0.43	1	1
B:42:LEU:HD11	B:179:PHE:CZ	0.42	1	1
A:112:ASP:OD1	A:268:LYS:HE3	0.41	1	1
A:26:TYR:HB2	B:505:THR:HG21	0.41	1	1
B:44:TYR:CE1	B:192:ARG:HG2	0.40	1	1
B:413:THR:HG22	B:441:ARG:NH2	0.40	1	1
A:12:LEU:HD23	B:610:VAL:HG11	0.40	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	1457	1414	41	2

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

Chain	Res	Type	Models (Total)
B	106	PHE	1
B	340	LYS	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	1252	1188	41	23

There are 23 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	27	ILE	1
A	210	VAL	1
A	425	LEU	1
A	433	THR	1
A	496	LEU	1
A	585	THR	1
A	706	ILE	1
A	720	LEU	1
A	763	GLU	1
A	777	THR	1
A	779	ASP	1
B	7	PHE	1
B	154	VAL	1
B	301	VAL	1
B	331	GLU	1
B	346	SER	1
B	408	LEU	1
B	417	SER	1
B	463	LEU	1
B	505	THR	1
B	558	GLU	1
B	572	THR	1
B	649	LEU	1

Fit of model to data used for modeling

Fit of model(s) to crosslinking-MS data

Restraint types

Restraint types are summarized in the table below. Restraints assigned "by-residue" are interpreted as between CA

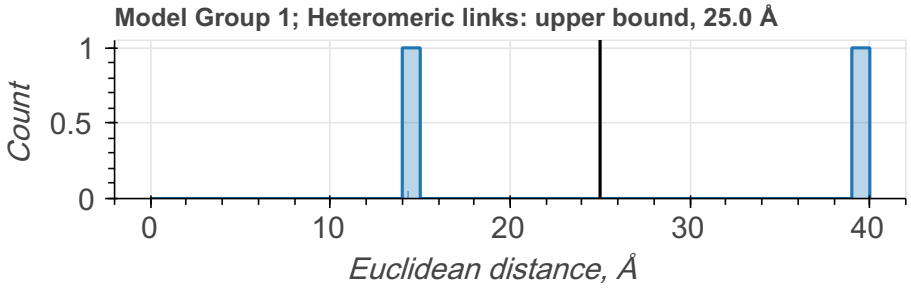
atoms. Restraints between coarse-grained beads are indicated as "coarse-grained". Restraint group represents a set of crosslinking restraints applied collectively in the modeling.

There are 2 crosslinking restraints combined in 2 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
SDA	LYS	CA	SER	CA	upper bound	25.0	1
SDA	LYS	CA	LYS	CA	upper bound	25.0	1

Distograms of individual restraints

Restraints with identical thresholds are grouped into one plot. Only the best distance per restraint per model group/ensemble is plotted. Inter- and intramolecular (including self-links) restraints are also grouped into one plot. Distance for a restraint between coarse-grained beads is calculated as a minimal distance between shells; if beads intersect, the distance will be reported as 0.0. A bead with the highest available resolution for a given residue is used for the assessment.



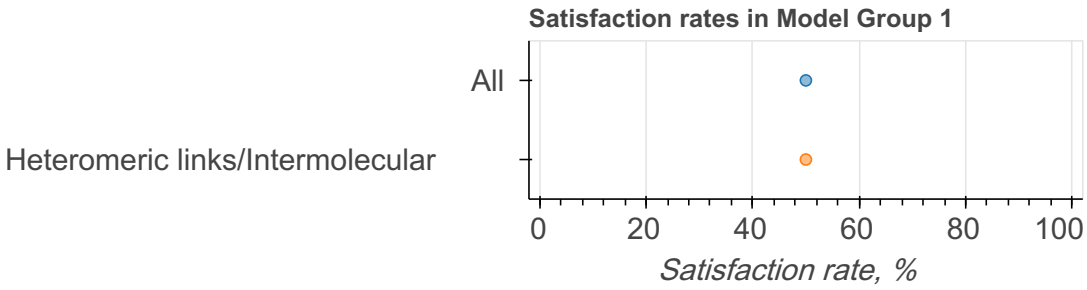
Satisfaction of restraints

Satisfaction of restraints is calculated on a restraint group (a set of crosslinking restraints applied collectively in the modeling) level. Satisfaction of a restraint group depends on satisfaction of individual restraints in the group and the conditionality (all/any). A restraint group is considered satisfied, if the condition was met in at least one model of the model group/ensemble. The number of measured restraints can be smaller than the total number of restraint groups if crosslinks involve non-modeled residues. Only deposited models are used for validation right now.

State group	State	Model group	# of Deposited models/Total	Restraint group type	Satisfied (%)	Violated (%)	Count (Total=2)
1	1	1	1/1	All	50.00	50.00	2
				Heteromeric links/ Intermolecular	50.00	50.00	2

Per-model satisfaction rates in ensembles

Every point represents one model in a model group/ensemble. Where possible, boxplots with quartile marks are also plotted.



Fit of model to data used for validation

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 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.