Integrative Structure Validation Report

February 18, 2025 - 08:43 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	9A6Z
PDB-Dev ID	PDBDEV_00000328
Structure Title	Integrative model of RPOB-DNLJ 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

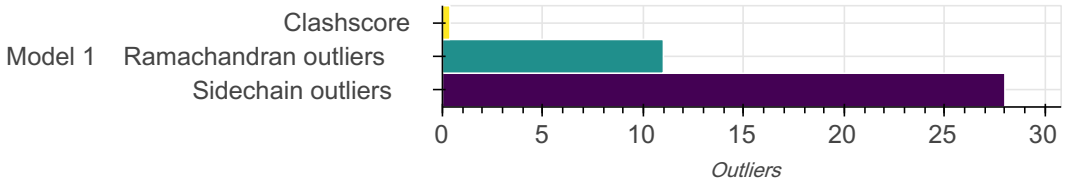
A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the ? symbol.

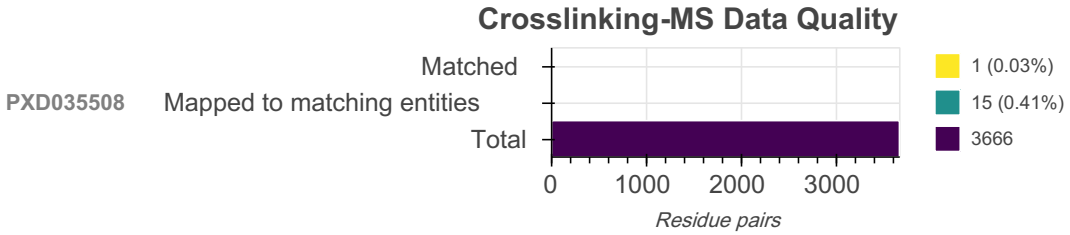
List of references used to build this report is available here.

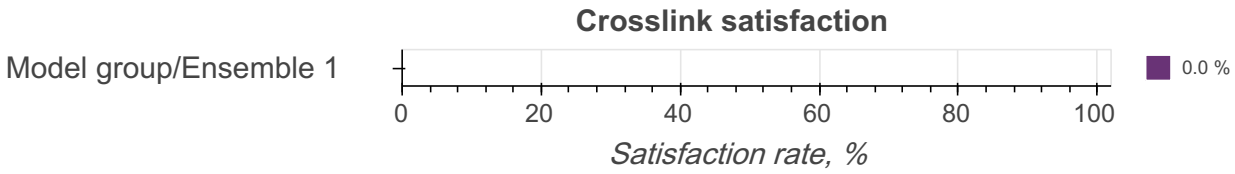
Overall quality

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	RPOB_BACSU	A	1193	-	1-1193	100.00 / 0.00	Atomic
		2	DNLJ_BACSU	B	668	-	1-668	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	RPOB_BACSU	dbseq_P37870_target	0.00	True
2	DNLJ_BACSU	dbseq_O31498_target	0.00	True

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A6Z	1	1 (100.00%)	1 (100.00%)
PXD035508	3666	15 (0.41%)	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 88 bond angle outliers in this entry (0.44% of 20100 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	1183	ALA	C-N-CA	8.50	137.00	121.70	1	1
A	1187	ARG	C-N-CA	8.05	136.19	121.70	1	1
A	606	ASN	CA-CB-CG	7.10	119.70	112.60	1	1
A	1192	LYS	C-N-CA	7.03	134.35	121.70	1	1
A	1180	THR	C-N-CA	6.41	133.24	121.70	1	1
A	773	ASP	CA-CB-CG	6.12	118.72	112.60	1	1
B	92	ASP	CA-CB-CG	5.90	118.50	112.60	1	1
A	197	ARG	NE-CZ-NH2	5.87	124.49	119.20	1	1
A	1180	THR	O-C-N	5.81	113.70	123.00	1	1
B	413	ARG	NE-CZ-NH1	5.79	127.29	121.50	1	1
A	1180	THR	CA-C-N	5.48	127.16	116.20	1	1
A	175	ASP	CA-CB-CG	5.39	117.99	112.60	1	1
A	128	ASN	CA-CB-CG	5.14	117.74	112.60	1	1
A	1192	LYS	CA-C-N	5.06	126.33	116.20	1	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	572	GLN	OE1-CD-NE2	5.05	117.55	122.60	1	1
A	51	ASP	CA-CB-CG	5.04	117.64	112.60	1	1
A	438	GLN	OE1-CD-NE2	5.03	117.57	122.60	1	1
B	479	GLN	OE1-CD-NE2	5.02	117.58	122.60	1	1
B	270	GLN	OE1-CD-NE2	4.99	117.61	122.60	1	1
A	355	GLN	OE1-CD-NE2	4.94	117.66	122.60	1	1
A	390	ASN	CA-CB-CG	4.91	117.51	112.60	1	1
B	293	GLN	OE1-CD-NE2	4.89	117.71	122.60	1	1
B	42	GLN	OE1-CD-NE2	4.84	117.76	122.60	1	1
A	701	GLN	OE1-CD-NE2	4.81	117.79	122.60	1	1
A	576	GLN	OE1-CD-NE2	4.80	117.80	122.60	1	1
B	496	GLN	OE1-CD-NE2	4.76	117.84	122.60	1	1
A	121	ASP	CA-CB-CG	4.74	117.34	112.60	1	1
A	469	GLN	OE1-CD-NE2	4.68	117.92	122.60	1	1
A	507	HIS	CB-CG-CD2	4.68	125.12	131.20	1	1
A	967	ASN	OD1-CG-ND2	4.65	117.95	122.60	1	1
B	390	GLN	OE1-CD-NE2	4.65	117.95	122.60	1	1
A	612	ASP	CA-CB-CG	4.61	117.21	112.60	1	1
B	111	ASN	OD1-CG-ND2	4.60	118.00	122.60	1	1
A	419	GLN	OE1-CD-NE2	4.59	118.01	122.60	1	1
A	375	ASN	OD1-CG-ND2	4.57	118.03	122.60	1	1
A	1185	VAL	CA-CB-CG2	4.56	118.15	110.40	1	1
A	403	HIS	CB-CG-CD2	4.54	125.29	131.20	1	1
B	351	ARG	NE-CZ-NH1	4.52	126.02	121.50	1	1
A	825	ASP	CA-CB-CG	4.52	117.12	112.60	1	1
B	102	GLN	OE1-CD-NE2	4.51	118.09	122.60	1	1
A	280	HIS	CB-CG-CD2	4.51	125.34	131.20	1	1
A	50	GLN	OE1-CD-NE2	4.48	118.12	122.60	1	1
A	1179	GLU	C-N-CA	4.48	129.76	121.70	1	1
A	606	ASN	OD1-CG-ND2	4.46	118.14	122.60	1	1
B	617	GLN	OE1-CD-NE2	4.46	118.14	122.60	1	1
B	456	GLN	OE1-CD-NE2	4.44	118.16	122.60	1	1
A	258	ARG	NE-CZ-NH2	4.43	123.18	119.20	1	1
A	1179	GLU	N-CA-C	4.40	123.32	111.00	1	1
A	359	GLN	OE1-CD-NE2	4.37	118.23	122.60	1	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	437	HIS	CB-CG-CD2	4.37	125.53	131.20	1	1
A	308	GLN	OE1-CD-NE2	4.36	118.24	122.60	1	1
A	652	GLN	OE1-CD-NE2	4.35	118.25	122.60	1	1
A	1011	ARG	CD-NE-CZ	4.34	130.48	124.40	1	1
A	1165	GLN	OE1-CD-NE2	4.34	118.26	122.60	1	1
A	1192	LYS	O-C-N	4.33	116.07	123.00	1	1
A	582	ASP	CA-CB-CG	4.31	116.91	112.60	1	1
B	62	GLN	OE1-CD-NE2	4.31	118.29	122.60	1	1
A	1188	ASP	N-CA-C	4.27	122.96	111.00	1	1
B	650	GLN	OE1-CD-NE2	4.27	118.33	122.60	1	1
A	1042	HIS	CB-CG-CD2	4.27	125.65	131.20	1	1
B	195	ASN	CA-CB-CG	4.27	116.87	112.60	1	1
B	73	GLN	OE1-CD-NE2	4.26	118.34	122.60	1	1
A	720	GLN	OE1-CD-NE2	4.26	118.34	122.60	1	1
A	1178	GLU	N-CA-CB	4.25	103.27	110.50	1	1
B	8	GLN	OE1-CD-NE2	4.22	118.38	122.60	1	1
B	237	GLN	OE1-CD-NE2	4.21	118.39	122.60	1	1
B	235	GLN	OE1-CD-NE2	4.21	118.39	122.60	1	1
A	791	HIS	CB-CG-CD2	4.20	125.74	131.20	1	1
A	538	ASN	CA-CB-CG	4.18	116.78	112.60	1	1
B	661	GLN	OE1-CD-NE2	4.18	118.42	122.60	1	1
A	657	PHE	CA-CB-CG	4.16	117.96	113.80	1	1
B	632	ASN	CA-CB-CG	4.15	116.75	112.60	1	1
A	447	GLN	OE1-CD-NE2	4.14	118.46	122.60	1	1
B	28	ASP	CA-CB-CG	4.14	116.74	112.60	1	1
A	1069	GLN	OE1-CD-NE2	4.14	118.46	122.60	1	1
B	77	HIS	CB-CG-CD2	4.12	125.84	131.20	1	1
A	311	ASP	CA-CB-CG	4.11	116.71	112.60	1	1
A	970	GLN	OE1-CD-NE2	4.10	118.50	122.60	1	1
A	446	GLN	OE1-CD-NE2	4.07	118.53	122.60	1	1
A	13	GLN	OE1-CD-NE2	4.07	118.53	122.60	1	1
B	291	ASP	CA-CB-CG	4.07	116.67	112.60	1	1
A	28	ASN	OD1-CG-ND2	4.06	118.54	122.60	1	1
A	337	VAL	CA-CB-CG1	4.05	117.28	110.40	1	1
B	490	ASN	OD1-CG-ND2	4.04	118.56	122.60	1	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	430	GLN	OE1-CD-NE2	4.04	118.56	122.60	1	1
A	4	GLN	OE1-CD-NE2	4.04	118.56	122.60	1	1
B	653	ASN	OD1-CG-ND2	4.01	118.59	122.60	1	1
A	578	ASN	OD1-CG-ND2	4.00	118.60	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	0.38	11

There are 11 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:90:ALA:HB2	A:118:ILE:HD11	0.73	1	1
A:90:ALA:CB	A:118:ILE:HD11	0.59	1	1
A:777:MET:HG3	A:781:LEU:HD12	0.49	1	1
A:791:HIS:CE1	A:1043:MET:HE1	0.49	1	1
A:1192:LYS:H	A:1193:GLU:C	0.46	1	1
A:781:LEU:HD11	A:954:ASP:HB3	0.45	1	1
B:637:ILE:HD13	B:662:LEU:HD13	0.44	1	1
B:113:GLU:CD	B:286:LYS:HZ2	0.43	1	1
B:19:LYS:HE2	B:23:GLU:OE2	0.43	1	1
A:1179:GLU:HG2	A:1180:THR:HG23	0.41	1	1
A:85:ASP:CG	A:910:ARG:HH22	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	1857	1800	46	11

There are 11 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	268	ALA	1
A	1061	GLN	1
A	1170	ALA	1

Chain	Res	Type	Models (Total)
A	1175	GLU	1
A	1177	PRO	1
A	1183	ALA	1
A	1184	ASP	1
A	1188	ASP	1
A	1189	VAL	1
A	1191	THR	1
B	528	GLU	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	1600	1523	49	28

There are 28 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	2	THR	1
A	23	VAL	1
A	48	MET	1
A	57	ASP	1
A	255	LEU	1
A	319	LEU	1
A	337	VAL	1
A	443	ILE	1
A	606	ASN	1
A	645	ARG	1
A	753	LEU	1
A	920	SER	1
A	1044	VAL	1
A	1179	GLU	1
A	1184	ASP	1
A	1185	VAL	1
A	1188	ASP	1
A	1189	VAL	1

Chain	Res	Type	Models (Total)
A	1190	VAL	1
B	83	SER	1
B	95	ASP	1
B	217	LEU	1
B	227	LEU	1
B	316	VAL	1
B	418	VAL	1
B	420	LEU	1
B	449	LEU	1
B	625	LEU	1

Fit of model to data used for modeling

Fit of model(s) to crosslinking-MS data

Restraint types

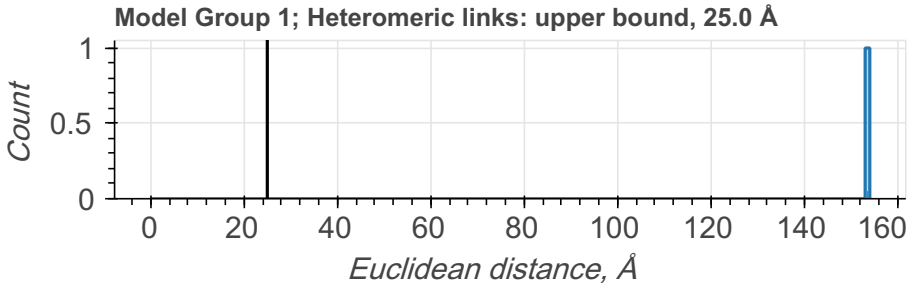
Restraint types are summarized in the table below. Restraints assigned "by-residue" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "coarse-grained". Restraint group represents a set of crosslinking restraints applied collectively in the modeling.

There are 1 crosslinking restraints combined in 1 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
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.



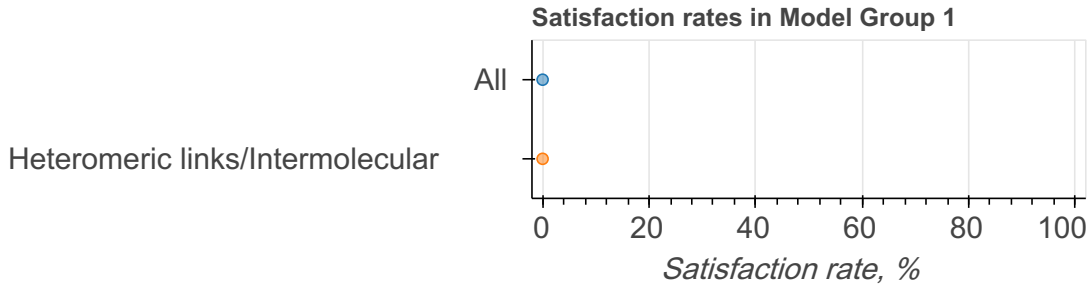
Satisfaction of restraints

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=1)
1	1	1	1/1	All	0.00	100.00	1
				Heteromeric links/ Intermolecular	0.00	100.00	1

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.