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	9A6Y
PDB-Dev ID	PDBDEV_00000327
Structure Title	Integrative model of RPOE-RPOC 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

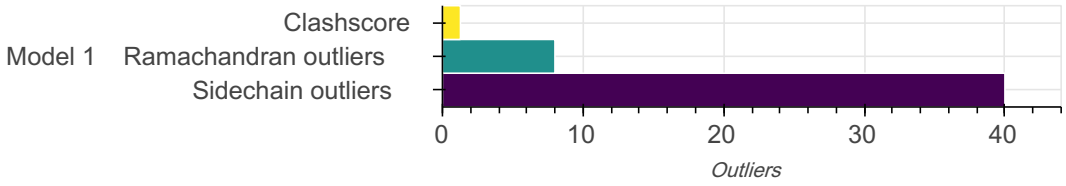
A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the ? symbol.

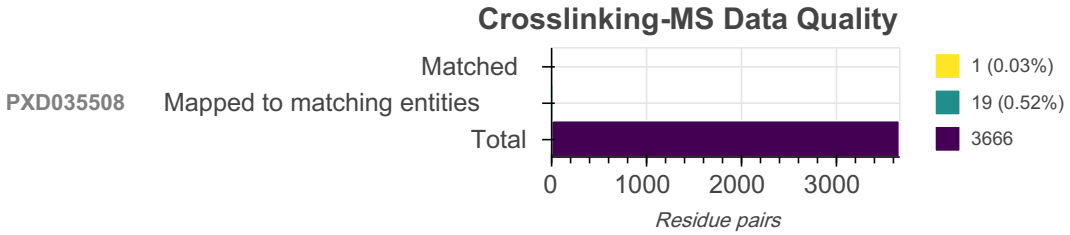
List of references used to build this report is available here.

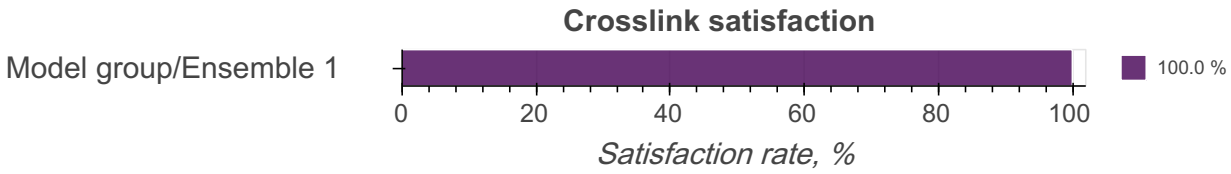
Overall quality

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	RPOE_BACSU	A	173	-	1-173	100.00 / 0.00	Atomic
		2	RPOC_BACSU	B	1199	-	1-1199	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	RPOE_BACSU	dbseq_P12464_target	0.00	True
2	RPOC_BACSU	dbseq_P37871_target	0.00	True

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A6Y	1	1 (100.00%)	1 (100.00%)
PXD035508	3666	19 (0.52%)	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 76 bond angle outliers in this entry (0.51% of 14893 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	153	ILE	C-N-CA	9.17	138.21	121.70	1	1
A	170	GLU	C-N-CA	8.25	136.55	121.70	1	1
A	167	ASP	CA-CB-CG	6.44	119.04	112.60	1	1
A	172	ILE	C-N-CA	6.38	133.19	121.70	1	1
A	170	GLU	O-C-N	6.24	113.01	123.00	1	1
A	159	ASP	C-N-CA	6.22	132.90	121.70	1	1
A	161	ASP	C-N-CA	6.11	132.70	121.70	1	1
B	737	ASN	CA-CB-CG	5.69	118.29	112.60	1	1
B	668	ASP	CA-CB-CG	5.58	118.18	112.60	1	1
A	153	ILE	CA-C-N	5.45	127.10	116.20	1	1
A	162	ASP	CA-CB-CG	5.42	118.02	112.60	1	1
A	153	ILE	O-C-N	5.39	114.37	123.00	1	1
A	170	GLU	CA-C-N	5.36	126.91	116.20	1	1
A	148	ASP	CA-CB-CG	5.27	117.87	112.60	1	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	154	GLU	C-N-CA	5.26	131.18	121.70	1	1
A	157	ILE	CA-CB-CG1	5.24	119.31	110.40	1	1
B	585	ASN	OD1-CG-ND2	5.20	117.40	122.60	1	1
A	163	GLU	O-C-N	5.20	114.68	123.00	1	1
A	172	ILE	O-C-N	5.19	114.70	123.00	1	1
A	158	ILE	CA-CB-CG1	5.16	119.17	110.40	1	1
A	163	GLU	C-N-CA	5.11	130.90	121.70	1	1
B	180	GLN	OE1-CD-NE2	5.03	117.57	122.60	1	1
B	887	ARG	NE-CZ-NH2	5.01	123.71	119.20	1	1
A	27	HIS	CB-CG-CD2	4.99	124.72	131.20	1	1
B	394	GLN	OE1-CD-NE2	4.81	117.79	122.60	1	1
B	439	HIS	CB-CG-CD2	4.76	125.01	131.20	1	1
B	527	GLN	OE1-CD-NE2	4.75	117.85	122.60	1	1
A	5	GLN	OE1-CD-NE2	4.74	117.86	122.60	1	1
B	532	HIS	CB-CG-CD2	4.71	125.08	131.20	1	1
B	713	GLN	OE1-CD-NE2	4.70	117.90	122.60	1	1
A	86	GLN	OE1-CD-NE2	4.68	117.92	122.60	1	1
B	630	ARG	NE-CZ-NH2	4.67	115.00	119.20	1	1
A	159	ASP	O-C-N	4.67	115.54	123.00	1	1
B	1097	HIS	CB-CG-CD2	4.65	125.15	131.20	1	1
A	153	ILE	CA-CB-CG1	4.64	118.29	110.40	1	1
A	150	ASP	CA-CB-CG	4.64	117.24	112.60	1	1
B	925	GLN	OE1-CD-NE2	4.63	117.97	122.60	1	1
B	477	GLN	OE1-CD-NE2	4.61	117.99	122.60	1	1
A	74	GLN	OE1-CD-NE2	4.60	118.00	122.60	1	1
A	149	ASP	CA-CB-CG	4.53	117.13	112.60	1	1
A	155	GLU	N-CA-C	4.49	123.57	111.00	1	1
B	1044	GLN	OE1-CD-NE2	4.47	118.13	122.60	1	1
A	172	ILE	CA-C-N	4.41	125.01	116.20	1	1
B	1190	GLN	OE1-CD-NE2	4.39	118.21	122.60	1	1
B	968	GLN	OE1-CD-NE2	4.38	118.22	122.60	1	1
B	737	ASN	OD1-CG-ND2	4.38	118.22	122.60	1	1
B	753	ASN	OD1-CG-ND2	4.37	118.23	122.60	1	1
B	994	GLN	OE1-CD-NE2	4.36	118.24	122.60	1	1
A	168	GLU	C-N-CA	4.33	129.49	121.70	1	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	169	GLU	N-CA-C	4.32	123.10	111.00	1	1
B	803	ARG	NE-CZ-NH2	4.32	123.09	119.20	1	1
B	553	GLN	OE1-CD-NE2	4.30	118.30	122.60	1	1
B	988	GLN	OE1-CD-NE2	4.26	118.34	122.60	1	1
B	896	HIS	CB-CG-CD2	4.26	125.67	131.20	1	1
B	318	HIS	CB-CG-CD2	4.22	125.71	131.20	1	1
B	167	GLN	OE1-CD-NE2	4.22	118.38	122.60	1	1
A	159	ASP	CA-C-N	4.20	124.60	116.20	1	1
B	66	LYS	C-N-CA	4.19	129.25	121.70	1	1
B	686	GLN	OE1-CD-NE2	4.18	118.42	122.60	1	1
A	34	GLN	OE1-CD-NE2	4.15	118.45	122.60	1	1
B	989	GLN	OE1-CD-NE2	4.14	118.46	122.60	1	1
B	851	GLN	OE1-CD-NE2	4.13	118.47	122.60	1	1
B	809	GLN	OE1-CD-NE2	4.11	118.49	122.60	1	1
B	324	GLN	OE1-CD-NE2	4.11	118.49	122.60	1	1
B	243	GLN	OE1-CD-NE2	4.11	118.49	122.60	1	1
B	141	ASN	OD1-CG-ND2	4.10	118.50	122.60	1	1
B	608	GLN	OE1-CD-NE2	4.10	118.50	122.60	1	1
B	11	ASN	OD1-CG-ND2	4.09	118.51	122.60	1	1
B	670	GLN	OE1-CD-NE2	4.08	118.52	122.60	1	1
B	329	GLN	OE1-CD-NE2	4.06	118.54	122.60	1	1
B	1066	HIS	CB-CG-CD2	4.06	125.92	131.20	1	1
B	1000	ARG	NE-CZ-NH2	4.04	122.84	119.20	1	1
A	133	ASP	CA-CB-CG	4.04	116.64	112.60	1	1
A	111	ASP	CA-CB-CG	4.03	116.63	112.60	1	1
B	951	GLN	OE1-CD-NE2	4.02	118.58	122.60	1	1
A	8	GLN	OE1-CD-NE2	4.01	118.59	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	1.28	28

There are 28 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)
B:808:ALA:CB	B:1070:MET:HE1	0.73	1	1
B:814:ARG:HH12	B:1159:LEU:HD22	0.64	1	1
B:371:MET:CE	B:383:ILE:HG23	0.62	1	1
B:405:ILE:HG23	B:428:PRO:HG2	0.61	1	1
B:751:MET:HE2	B:779:SER:HA	0.58	1	1
A:172:ILE:HG22	A:173:LYS:HB2	0.57	1	1
B:371:MET:HE3	B:383:ILE:HG23	0.57	1	1
B:803:ARG:NH1	B:1124:THR:HG23	0.56	1	1
B:317:SER:HB3	B:321:LYS:HE2	0.56	1	1
B:506:ARG:HH21	B:723:LEU:HD22	0.55	1	1
B:720:LEU:HD11	B:741:PHE:CE2	0.53	1	1
B:503:THR:HG21	B:615:PHE:CD2	0.53	1	1
B:747:MET:HE2	B:749:GLY:HA2	0.52	1	1
B:583:LYS:HE2	B:587:GLU:OE2	0.52	1	1
B:317:SER:HB3	B:321:LYS:CE	0.49	1	1
B:808:ALA:HB2	B:1070:MET:HE1	0.47	1	1
B:726:ILE:HG21	B:741:PHE:CD1	0.47	1	1
B:806:ASP:CG	B:1162:LYS:HZ3	0.47	1	1
B:346:VAL:HG22	B:450:PHE:CE2	0.46	1	1
B:895:PRO:HA	B:1097:HIS:CD2	0.46	1	1
B:803:ARG:HH12	B:1124:THR:HG23	0.45	1	1
B:1086:THR:OG1	B:1088:VAL:HG22	0.45	1	1
B:751:MET:HE3	B:782:GLY:HA3	0.43	1	1
B:1106:VAL:HG13	B:1111:ASN:HB2	0.41	1	1
B:506:ARG:HE	B:723:LEU:CD2	0.41	1	1
A:157:ILE:CG2	A:158:ILE:HG12	0.41	1	1
B:930:PRO:HB3	B:1060:VAL:HG11	0.41	1	1
B:352:MET:HE2	B:476:ALA:HA	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	1368	1309	51	8

There are 8 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	97	ALA	1
A	100	LYS	1
A	117	GLU	1
A	118	ILE	1
A	124	ASP	1
A	145	ASP	1
A	147	ASP	1
B	67	ARG	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	1186	1062	84	40

There are 40 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	1	MET	1
A	3	ILE	1
A	37	LEU	1
A	43	LEU	1
A	46	VAL	1
A	62	LEU	1
A	64	ILE	1
A	124	ASP	1
A	153	ILE	1
A	158	ILE	1
A	172	ILE	1
B	1	MET	1
B	2	LEU	1
B	107	PHE	1
B	108	LYS	1
B	125	LEU	1
B	159	LEU	1
B	195	LEU	1
B	234	ILE	1

Chain	Res	Type	Models (Total)
B	287	MET	1
B	357	LEU	1
B	381	HIS	1
B	408	HIS	1
B	497	LEU	1
B	504	LEU	1
B	666	LEU	1
B	716	LEU	1
B	737	ASN	1
B	761	LEU	1
B	780	THR	1
B	806	ASP	1
B	807	VAL	1
B	824	ILE	1
B	953	LEU	1
B	1040	LEU	1
B	1130	THR	1
B	1167	ILE	1
B	1171	VAL	1
B	1177	MET	1
B	1178	MET	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.

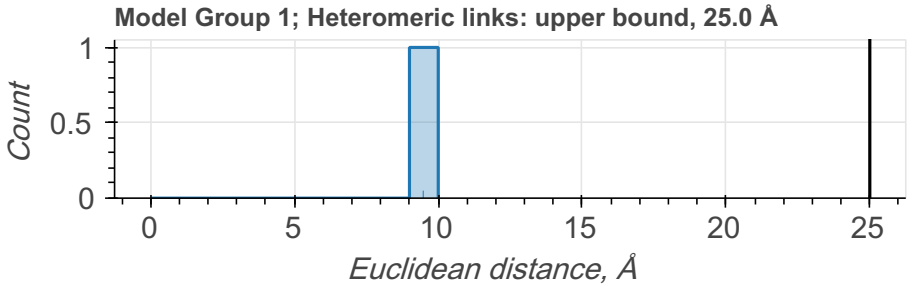
There are 1 crosslinking restraints combined in 1 restraint groups.

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.



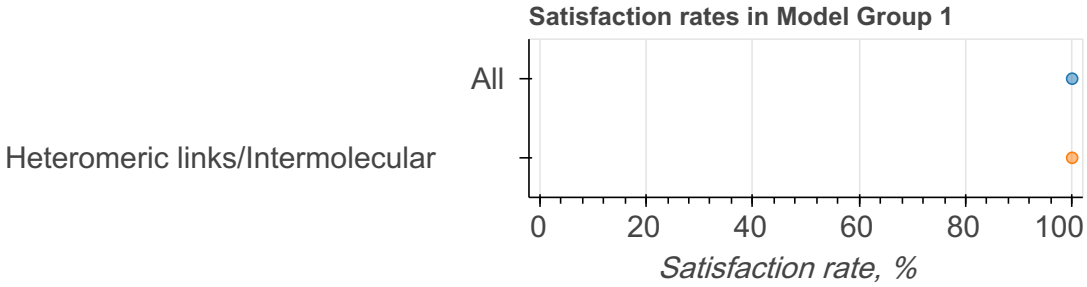
Satisfaction of restraints

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	100.00	0.00	1
				Heteromeric links/ Intermolecular	100.00	0.00	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.