

Integrative Structure Validation Report ?

February 18, 2025 - 08:40 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	9A4Y
PDB-Dev ID	PDBDEV_00000255
Structure Title	Integrative model of RPOA-RPOB 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

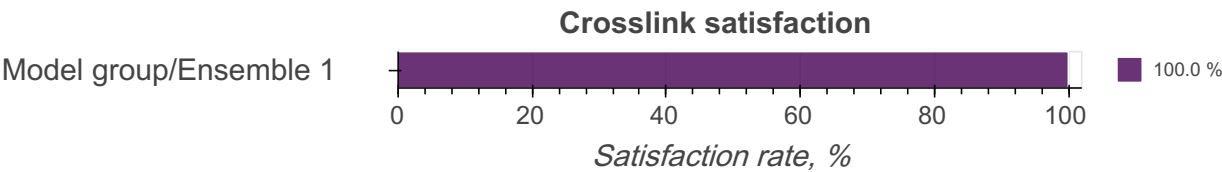
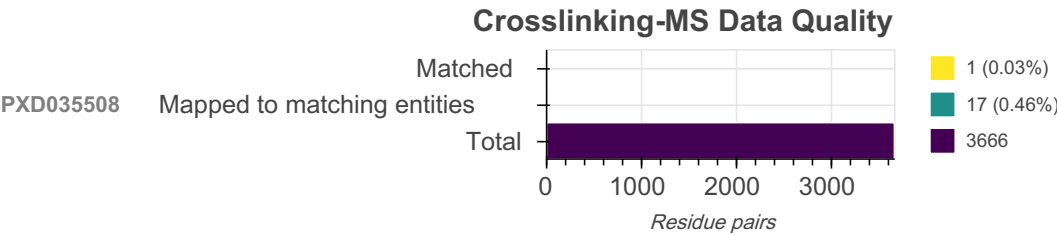
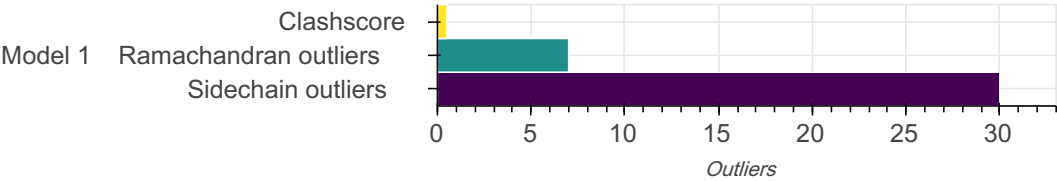
A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the ? symbol.

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	RPOA_BACSU	A	314	-	1-314	100.00 / 0.00	Atomic
		2	RPOB_BACSU	B	1193	-	1-1193	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	RPOA_BACSU	dbseq_P20429_target	0.00	True
2	RPOB_BACSU	dbseq_P37870_target	0.00	True

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A4Y	1	1 (100.00%)	1 (100.00%)
PXD035508	3666	17 (0.46%)	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 71 bond angle outliers in this entry (0.44% of 16269 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	1191	THR	C-N-CA	8.43	136.87	121.70	1	1
B	1184	ASP	C-N-CA	8.40	136.82	121.70	1	1
B	1187	ARG	C-N-CA	7.54	135.28	121.70	1	1
B	175	ASP	CA-CB-CG	6.13	118.73	112.60	1	1
B	1188	ASP	CA-CB-CG	5.83	118.43	112.60	1	1
B	1188	ASP	N-CA-C	5.70	126.95	111.00	1	1
B	1185	VAL	N-CA-C	5.58	126.61	111.00	1	1
B	1181	ALA	C-N-CA	5.55	131.70	121.70	1	1
B	128	ASN	CA-CB-CG	5.35	117.95	112.60	1	1
B	311	ASP	CA-CB-CG	5.13	117.73	112.60	1	1
B	1186	GLU	N-CA-C	5.07	125.19	111.00	1	1
A	74	ASP	CA-CB-CG	5.02	117.62	112.60	1	1
B	390	ASN	CA-CB-CG	5.00	117.60	112.60	1	1
B	576	GLN	OE1-CD-NE2	4.98	117.62	122.60	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	192	ASP	CA-CB-CG	4.94	117.54	112.60	1	1
B	773	ASP	CA-CB-CG	4.90	117.50	112.60	1	1
B	258	ARG	NE-CZ-NH2	4.90	123.61	119.20	1	1
B	469	GLN	OE1-CD-NE2	4.90	117.70	122.60	1	1
B	701	GLN	OE1-CD-NE2	4.90	117.70	122.60	1	1
B	438	GLN	OE1-CD-NE2	4.89	117.71	122.60	1	1
B	121	ASP	CA-CB-CG	4.84	117.44	112.60	1	1
B	280	HIS	CB-CG-CD2	4.83	124.92	131.20	1	1
B	1191	THR	CA-C-N	4.79	125.79	116.20	1	1
B	355	GLN	OE1-CD-NE2	4.76	117.84	122.60	1	1
A	159	GLN	OE1-CD-NE2	4.73	117.87	122.60	1	1
A	115	ASP	CA-CB-CG	4.69	117.29	112.60	1	1
B	1187	ARG	N-CA-C	4.66	124.05	111.00	1	1
B	967	ASN	OD1-CG-ND2	4.61	117.99	122.60	1	1
B	51	ASP	CA-CB-CG	4.59	117.19	112.60	1	1
A	244	GLN	OE1-CD-NE2	4.54	118.06	122.60	1	1
B	910	ARG	NE-CZ-NH2	4.54	115.12	119.20	1	1
B	507	HIS	CB-CG-CD2	4.51	125.34	131.20	1	1
B	50	GLN	OE1-CD-NE2	4.46	118.14	122.60	1	1
B	582	ASP	CA-CB-CG	4.45	117.05	112.60	1	1
A	122	ASN	CA-CB-CG	4.44	117.04	112.60	1	1
B	538	ASN	CA-CB-CG	4.43	117.03	112.60	1	1
B	375	ASN	OD1-CG-ND2	4.42	118.18	122.60	1	1
B	4	GLN	OE1-CD-NE2	4.39	118.21	122.60	1	1
A	63	HIS	CB-CG-CD2	4.38	125.51	131.20	1	1
B	1101	ASP	CA-CB-CG	4.38	116.98	112.60	1	1
B	1181	ALA	O-C-N	4.37	116.00	123.00	1	1
B	308	GLN	OE1-CD-NE2	4.37	118.23	122.60	1	1
B	447	GLN	OE1-CD-NE2	4.37	118.23	122.60	1	1
A	187	GLN	OE1-CD-NE2	4.36	118.24	122.60	1	1
B	419	GLN	OE1-CD-NE2	4.35	118.25	122.60	1	1
A	232	GLN	OE1-CD-NE2	4.35	118.25	122.60	1	1
B	1191	THR	O-C-N	4.34	116.06	123.00	1	1
B	791	HIS	CB-CG-CD2	4.34	125.56	131.20	1	1
A	143	GLN	OE1-CD-NE2	4.32	118.28	122.60	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	11	HIS	CB-CG-CD2	4.28	125.64	131.20	1	1
B	359	GLN	OE1-CD-NE2	4.26	118.34	122.60	1	1
A	254	GLU	CB-CG-CD	4.20	119.75	112.60	1	1
B	720	GLN	OE1-CD-NE2	4.17	118.43	122.60	1	1
B	1042	HIS	CB-CG-CD2	4.17	125.78	131.20	1	1
A	190	ASN	OD1-CG-ND2	4.16	118.44	122.60	1	1
B	402	ASP	CA-CB-CG	4.15	108.45	112.60	1	1
B	1165	GLN	OE1-CD-NE2	4.14	118.46	122.60	1	1
B	606	ASN	CA-CB-CG	4.13	116.73	112.60	1	1
B	652	GLN	OE1-CD-NE2	4.12	118.48	122.60	1	1
A	275	GLN	OE1-CD-NE2	4.12	118.48	122.60	1	1
B	881	HIS	CB-CG-CD2	4.12	125.85	131.20	1	1
B	578	ASN	OD1-CG-ND2	4.10	118.50	122.60	1	1
B	1183	ALA	N-CA-C	4.09	122.47	111.00	1	1
A	314	ASP	CA-CB-CG	4.08	116.68	112.60	1	1
B	1069	GLN	OE1-CD-NE2	4.07	118.53	122.60	1	1
B	646	GLN	OE1-CD-NE2	4.06	118.54	122.60	1	1
A	179	GLN	OE1-CD-NE2	4.05	118.55	122.60	1	1
B	1189	VAL	C-N-CA	4.04	128.97	121.70	1	1
B	466	GLN	OE1-CD-NE2	4.02	118.58	122.60	1	1
B	1181	ALA	CA-C-N	4.01	124.22	116.20	1	1
B	657	PHE	CA-CB-CG	4.01	117.81	113.80	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	0.51	12

There are 12 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:90:ALA:HB2	B:118:ILE:HD11	0.70	1	1
B:791:HIS:CE1	B:1043:MET:HE1	0.60	1	1
B:777:MET:HG3	B:781:LEU:HD12	0.55	1	1
B:1189:VAL:HG12	B:1190:VAL:H	0.46	1	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
B:48:MET:HA	B:48:MET:HE2	0.44	1	1
B:781:LEU:HD11	B:954:ASP:HB3	0.44	1	1
B:85:ASP:CG	B:910:ARG:HH22	0.44	1	1
B:769:TYR:HB2	B:776:ILE:HD11	0.43	1	1
B:1191:THR:HB	B:1192:LYS:HB2	0.42	1	1
B:48:MET:CE	B:48:MET:HA	0.41	1	1
B:471:MET:HE1	B:483:LYS:HE2	0.41	1	1
B:764:MET:HE1	B:1026:PHE:CZ	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	1503	1465	31	7

There are 7 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
B	268	ALA	1
B	1041	ALA	1
B	1061	GLN	1
B	1177	PRO	1
B	1178	GLU	1
B	1182	SER	1
B	1185	VAL	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	1299	1233	36	30

There are 30 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	76	THR	1
A	113	THR	1
A	124	ASP	1
A	196	LEU	1

Chain	Res	Type	Models (Total)
A	283	GLU	1
A	291	LEU	1
B	2	THR	1
B	23	VAL	1
B	48	MET	1
B	59	THR	1
B	175	ASP	1
B	192	VAL	1
B	223	LYS	1
B	255	LEU	1
B	319	LEU	1
B	334	ASN	1
B	353	THR	1
B	570	ASP	1
B	753	LEU	1
B	802	THR	1
B	846	VAL	1
B	881	HIS	1
B	920	SER	1
B	1044	VAL	1
B	1048	LEU	1
B	1169	LEU	1
B	1177	PRO	1
B	1185	VAL	1
B	1189	VAL	1
B	1191	THR	1

Fit of model to data used for modeling ?

Fit of model(s) to crosslinking-MS data

Restraint types

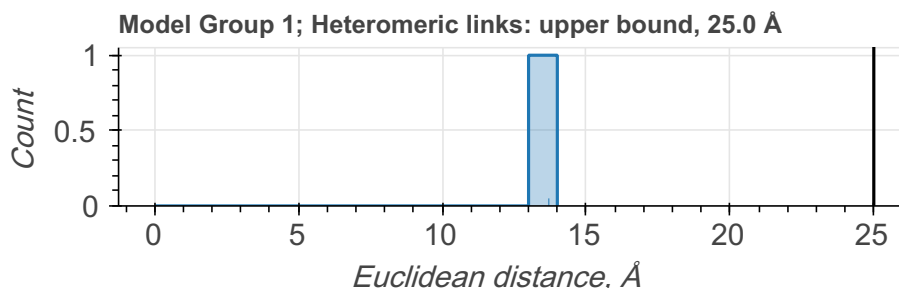
Restraint types are summarized in the table below. Restraints assigned "*by-residue*" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "*coarse-grained*". *Restraint group* represents a set of crosslinking restraints applied collectively in the modeling.

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.



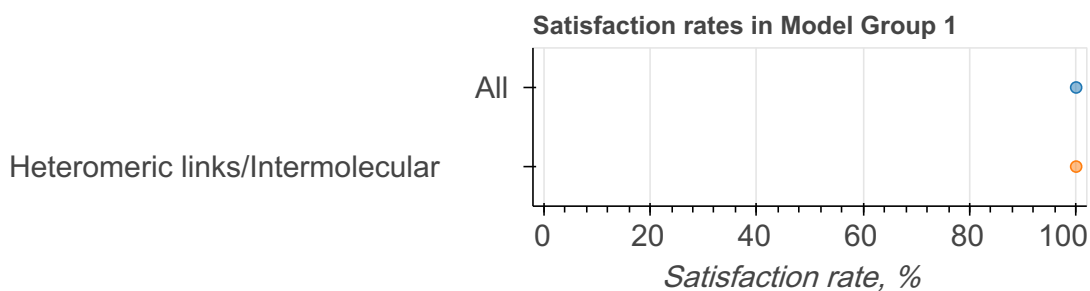
Satisfaction of restraints

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 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.