

# Integrative Structure Validation Report ?

April 09, 2025 - 11:36 PM PDT

*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*

PDB ID	9A7T
PDB-Dev ID	PDBDEV_00000358
Structure Title	Integrative model of RPOB-GREA by crosslinking MS and deep learning
Structure Authors	Kolja Stahl; Oliver Brock; Juri Rappsilber
Deposited on	2024-01-24

*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](mailto:helpdesk@pdb-ihm.org)*

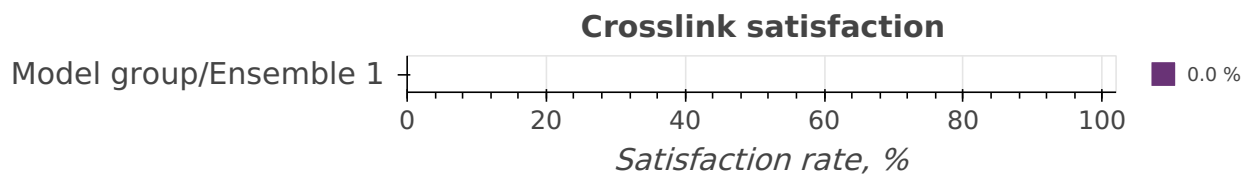
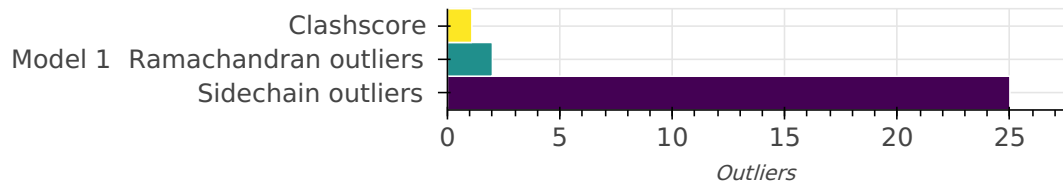
*A user guide is available at [https://pdb-ihm.org/validation\\_help.html](https://pdb-ihm.org/validation_help.html) with specific help available everywhere you see the ? symbol.*

*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	GREa_BACSU	B	157	-	1-157	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	<a href="#">AlphaLink2</a>	1.00	model building	<a href="https://github.com/Rappsilber-Laboratory/AlphaLink2">https://github.com/Rappsilber-Laboratory/AlphaLink2</a>

## 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 is not available in the [PRIDE Crosslinking](#) database.

## 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 46 bond angle outliers in this entry (0.32% of 14569 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	7.44	135.09	121.70	1	1
A	1189	VAL	CA-CB-CG1	6.83	122.01	110.40	1	1
A	1183	ALA	O-C-N	5.79	113.74	123.00	1	1
A	1189	VAL	O-C-N	5.54	114.14	123.00	1	1
A	1101	ASP	CA-CB-CG	5.42	118.02	112.60	1	1
A	720	GLN	OE1-CD-NE2	5.37	117.23	122.60	1	1
A	612	ASP	CA-CB-CG	5.18	117.78	112.60	1	1
A	438	GLN	OE1-CD-NE2	5.11	117.49	122.60	1	1
A	121	ASP	CA-CB-CG	5.01	117.61	112.60	1	1
A	701	GLN	OE1-CD-NE2	4.92	117.68	122.60	1	1
A	390	ASN	CA-CB-CG	4.90	117.50	112.60	1	1
A	578	ASN	OD1-CG-ND2	4.87	117.73	122.60	1	1
A	578	ASN	CA-CB-CG	4.83	117.43	112.60	1	1
A	185	ASP	CA-CB-CG	4.82	117.42	112.60	1	1
A	1189	VAL	C-N-CA	4.70	130.17	121.70	1	1
A	197	ARG	NE-CZ-NH1	4.70	126.20	121.50	1	1
A	1183	ALA	CA-C-N	4.67	125.55	116.20	1	1
A	538	ASN	CA-CB-CG	4.67	117.27	112.60	1	1
B	15	GLN	OE1-CD-NE2	4.66	117.94	122.60	1	1
A	915	GLN	OE1-CD-NE2	4.64	117.96	122.60	1	1
A	469	GLN	OE1-CD-NE2	4.59	118.01	122.60	1	1
A	355	GLN	OE1-CD-NE2	4.55	118.05	122.60	1	1
A	419	GLN	OE1-CD-NE2	4.55	118.05	122.60	1	1
A	1189	VAL	CA-C-N	4.53	125.26	116.20	1	1
B	19	GLN	OE1-CD-NE2	4.48	118.12	122.60	1	1
A	50	GLN	OE1-CD-NE2	4.46	118.14	122.60	1	1
A	10	ARG	NE-CZ-NH2	4.43	115.21	119.20	1	1
A	1191	THR	N-CA-C	4.38	123.26	111.00	1	1
A	965	ARG	CD-NE-CZ	4.36	130.51	124.40	1	1
A	280	HIS	CB-CG-CD2	4.35	125.54	131.20	1	1
A	773	ASP	CA-CB-CG	4.35	116.95	112.60	1	1
A	1185	VAL	C-N-CA	4.34	129.51	121.70	1	1
A	576	GLN	OE1-CD-NE2	4.28	118.32	122.60	1	1
A	1190	VAL	CG1-CB-CG2	4.28	101.39	110.80	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	58	GLN	OE1-CD-NE2	4.27	118.33	122.60	1	1
A	375	ASN	OD1-CG-ND2	4.21	118.39	122.60	1	1
B	3	GLN	OE1-CD-NE2	4.21	118.39	122.60	1	1
A	1165	GLN	OE1-CD-NE2	4.13	118.47	122.60	1	1
A	13	GLN	OE1-CD-NE2	4.12	118.48	122.60	1	1
A	403	HIS	CB-CG-CD2	4.11	125.86	131.20	1	1
A	308	GLN	OE1-CD-NE2	4.05	118.55	122.60	1	1
A	870	ARG	NH1-CZ-NH2	4.05	114.04	119.30	1	1
A	1188	ASP	C-CA-CB	4.04	117.77	110.10	1	1
A	28	ASN	OD1-CG-ND2	4.03	118.57	122.60	1	1
A	359	GLN	OE1-CD-NE2	4.02	118.58	122.60	1	1
A	652	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	1.09	23

There are 23 clashes. The table below contains the detailed list of all clashes based on a MolProbity analysis. Bad clashes are  $\geq 0.4$  Angstrom.

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:849:LYS:HE2	A:852:THR:HG21	0.77	1	1
A:1188:ASP:HB3	A:1189:VAL:HG23	0.68	1	1
A:1189:VAL:HG13	A:1190:VAL:HG23	0.66	1	1
A:849:LYS:CE	A:852:THR:HG21	0.64	1	1
A:777:MET:HG3	A:781:LEU:HD12	0.61	1	1
A:233:LEU:HD21	A:337:VAL:HG23	0.58	1	1
A:1183:ALA:HB1	A:1184:ASP:HB2	0.57	1	1
A:471:MET:HE1	A:483:LYS:HE2	0.54	1	1
A:978:MET:HE3	A:982:TYR:CZ	0.53	1	1
A:791:HIS:CE1	A:1043:MET:HE1	0.53	1	1
A:139:VAL:HG21	A:410:ARG:HE	0.51	1	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:1189:VAL:CG1	A:1190:VAL:HG23	0.49	1	1
A:48:MET:HA	A:48:MET:HE2	0.47	1	1
A:48:MET:CE	A:48:MET:HA	0.46	1	1
A:849:LYS:NZ	A:852:THR:HG21	0.46	1	1
A:90:ALA:HB3	A:118:ILE:HD11	0.45	1	1
A:981:ARG:HH11	A:1034:ILE:HD11	0.45	1	1
A:32:ILE:HD12	A:531:LEU:HD22	0.43	1	1
A:917:ARG:HH11	A:1043:MET:HE3	0.43	1	1
A:511:TYR:CE2	A:618:VAL:CG2	0.41	1	1
A:781:LEU:HD11	A:954:ASP:HB3	0.41	1	1
A:269:ASN:OD1	A:374:LYS:HE2	0.41	1	1
A:674:VAL:HG21	A:712:LEU:HD23	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	1346	1307	37	2

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

Chain	Res	Type	Models (Total)
A	268	ALA	1
A	1187	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	1161	1087	49	25

There are 25 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	2	THR	1
A	48	MET	1
A	161	THR	1
A	174	THR	1
A	175	ASP	1

Chain	Res	Type	Models (Total)
A	189	LYS	1
A	192	VAL	1
A	211	ILE	1
A	255	LEU	1
A	319	LEU	1
A	578	ASN	1
A	590	ASP	1
A	645	ARG	1
A	753	LEU	1
A	846	VAL	1
A	881	HIS	1
A	1044	VAL	1
A	1048	LEU	1
A	1169	LEU	1
A	1180	THR	1
A	1191	THR	1
B	34	GLU	1
B	89	LEU	1
B	150	LEU	1
B	151	VAL	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.

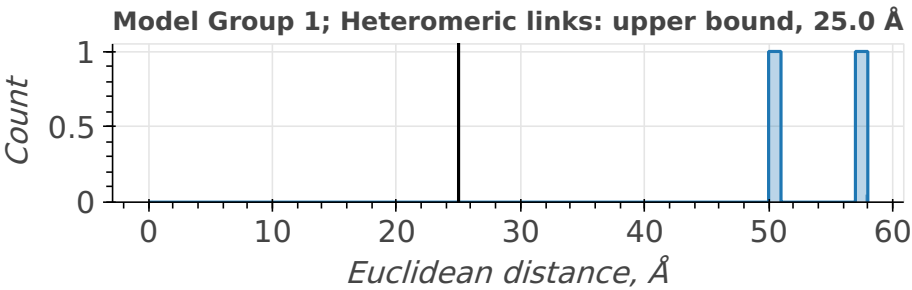
There are 2 crosslinking restraints combined in 2 restraint groups.

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

#### 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.



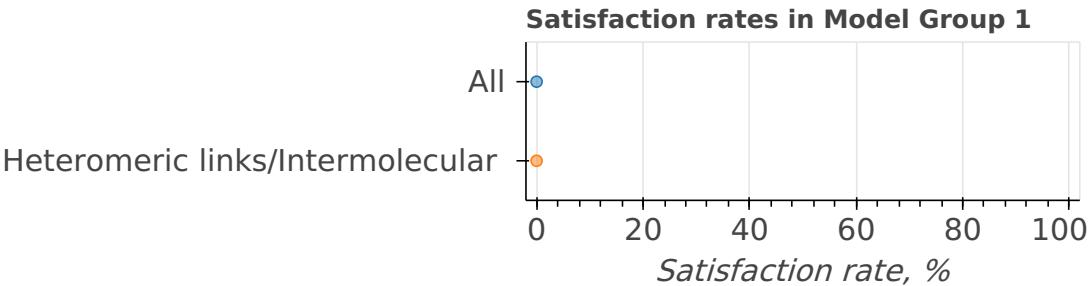
Satisfaction of restraints

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	0.00	100.00	2
				Heteromeric links/ Intermolecular	0.00	100.00	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.*