Integrative Structure Validation Report

October 09, 2025 - 04:42 PM PDT

The following software was used in the production of this report:

IHMValidation Version 3.0
Python-IHM Version 2.5
MolProbity Version 4.5.2
pyHMMER Version 0.11.1

PDB ID	9A38 | pdb_00009a38
PDB-Dev ID	PDBDEV_00000193
Structure Title	Model of E. coli MetQ by in-cell photo-crosslinking MS and deep learning
Structure Authors	Stahl, K.; Graziadei, A.; Dau, T.; Brock, O.; Rappsilber, J.
Deposited on	2023-02-03

This is a PDB-IHM Structure Validation Report.

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.

1. Overview

1.1. Summary

This entry consists of 10 model(s). A total of 1 dataset(s) were used to build this entry.

Name	Type	Count
Crosslinking-MS data	Experimental data	1

1.2. 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

Data Quality

Fit to Data Used for Modeling

2. Model Details

2.1. Ensemble information

This entry consists of 0 distinct ensemble(s).

2.2. 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	1-10	1	P28635	A	271	-	1-271	100.00 / 0.00	Atomic

2.3. 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	jPOSTrepo	JPST001851

2.4. 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	AlphaLink	AlphaLink with 10 msa subsamples	Not available	10	False	False

There is 1 software package reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	AlphaLink	1.00	model building	https://github.com/lhatsk/AlphaLink

3. Data quality

3.2. 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).

PXD036833

Number of entities in the crosslinking-MS dataset:	1102
Number of entities in the entry:	1

Matching entities:

Entity ID	Molecule name	Crosslinking-MS Entity ID	E-value	Exact match
1	P28635	dbseq_P28635_target	0.00	True

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A38	24	24 (100.00%)	24 (100.00%)
PXD036833	611	24 (3.93%)	24 (3.93%)

4. 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.

4.1b. MolProbity Analysis

Excluded volume satisfaction for the models in the entry are listed below. The Analysed column shows the number of particle-partice or particle-atom pairs for which excluded volume was analysed.

Standard geometry: bond outliers

There are no bond length outliers.

Standard geometry: angle outliers

There are 137 bond angle outliers in this entry (0.48% of 28650 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	86	PHE	CA-CB-CG	6.42	107.38	113.80	7	1
A	94	GLN	OE1-CD-NE2	6.18	116.42	122.60	8	8
A	237	ASN	CA-CB-CG	5.85	118.45	112.60	1	2
A	7	THR	C-N-CA	5.73	132.01	121.70	8	1
A	5	PHE	C-N-CA	5.65	131.88	121.70	6	3
A	4	LYS	C-N-CA	5.19	131.04	121.70	5	4
A	195	GLN	OE1-CD-NE2	5.19	117.41	122.60	6	8
A	3	PHE	C-N-CA	5.17	131.00	121.70	2	4
A	93	ASP	CA-CB-CG	5.11	117.71	112.60	8	3
A	73	ASN	CA-CB-CG	5.01	107.59	112.60	8	1
A	229	ASN	CA-CB-CG	5.01	117.61	112.60	8	2
A	132	GLN	OE1-CD-NE2	4.99	117.61	122.60	2	10
A	95	GLN	OE1-CD-NE2	4.98	117.62	122.60	9	4
A	50	GLN	OE1-CD-NE2	4.79	117.81	122.60	7	10
A	87	GLN	OE1-CD-NE2	4.71	117.89	122.60	7	5
A	248	GLN	OE1-CD-NE2	4.60	118.00	122.60	3	10
A	140	THR	CA-CB-OG1	4.52	116.38	109.60	7	1
A	16	GLY	C-N-CA	4.49	129.79	121.70	5	2
A	9	ALA	C-N-CA	4.45	129.70	121.70	3	1
A	44	GLN	OE1-CD-NE2	4.42	118.18	122.60	7	9
A	186	GLN	OE1-CD-NE2	4.38	118.22	122.60	6	7
A	32	HIS	CB-CG-CD2	4.37	125.52	131.20	4	9
A	88	HIS	CB-CG-CD2	4.36	125.54	131.20	5	3
A	208	GLN	OE1-CD-NE2	4.35	118.25	122.60	8	4
A	67	ASN	OD1-CG-ND2	4.20	118.40	122.60	7	5

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	263	PHE	CA-CB-CG	4.18	109.62	113.80	3	1
A	1	MET	C-N-CA	4.16	129.19	121.70	7	1
A	251	GLN	OE1-CD-NE2	4.15	118.45	122.60	6	3
A	84	ASN	OD1-CG-ND2	4.14	118.46	122.60	3	1
A	140	THR	CA-CB-CG2	4.13	117.53	110.50	5	1
A	25	GLN	OE1-CD-NE2	4.13	118.47	122.60	7	4
A	128	GLN	OE1-CD-NE2	4.11	118.49	122.60	7	5
A	141	ASN	CA-CB-CG	4.11	116.71	112.60	8	1
A	24	GLY	N-CA-C	4.08	101.47	113.30	6	3

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	2.15	9
2	3.10	13
3	3.34	14
4	1.91	8
5	1.43	6
6	2.62	11
7	2.86	12
8	4.05	17
9	2.15	9
10	1.43	6

There are 105 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:144:ARG:HD2	A:227:TYR:CZ	0.82	8	1
A:96:LEU:HD13	A:103:LEU:HD12	0.73	10	1
A:163:LEU:HD22	A:226:PRO:CG	0.66	1	4
A:112:TYR:CE1	A:263:PHE:CZ	0.64	2	1
A:144:ARG:HD2	A:227:TYR:CE2	0.64	8	1
A:110:PHE:CE1	A:112:TYR:CZ	0.63	3	2
A:140:THR:HG22	A:227:TYR:CZ	0.62	8	1
A:110:PHE:CE1	A:112:TYR:CE2	0.60	3	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:69:TYR:CE2	A:86:PHE:CE2	0.60	9	2
A:69:TYR:CE2	A:86:PHE:CE1	0.59	7	1
A:69:TYR:CD1	A:86:PHE:CE1	0.57	1	1
A:110:PHE:CE1	A:112:TYR:CE1	0.56	2	1
A:144:ARG:HD2	A:227:TYR:CE1	0.56	8	1
A:163:LEU:HD22	A:226:PRO:HG3	0.56	1	4
A:112:TYR:CZ	A:263:PHE:CZ	0.55	3	1
A:69:TYR:CE1	A:86:PHE:CE1	0.55	10	3
A:73:ASN:ND2	A:95:GLN:HE22	0.54	5	1
A:114:ILE:HG22	A:202:ASN:HA	0.54	1	5
A:38:ILE:HG13	A:86:PHE:CE2	0.54	9	1
A:88:HIS:CE1	A:229:ASN:OD1	0.53	8	2
A:88:HIS:CE1	A:91:TYR:HB2	0.53	7	3
A:69:TYR:CE2	A:86:PHE:CZ	0.53	9	1
A:88:HIS:CD2	A:140:THR:HG21	0.52	3	1
A:92:LEU:HD11	A:232:VAL:HG11	0.52	6	1
A:269:LYS:HE3	A:271:TRP:CE2	0.52	4	4
A:112:TYR:CZ	A:263:PHE:CE2	0.52	3	1
A:144:ARG:HD2	A:227:TYR:CD1	0.52	4	1
A:110:PHE:HE1	A:112:TYR:CZ	0.49	2	1
A:96:LEU:HD21	A:103:LEU:H	0.48	7	1
A:72:PRO:HG2	A:86:PHE:CE2	0.48	3	1
A:90:PRO:HG3	A:163:LEU:HD13	0.47	7	2
A:109:THR:OG1	A:229:ASN:HB3	0.47	8	1
A:69:TYR:CE1	A:86:PHE:CZ	0.47	1	1
A:111:VAL:HG11	A:271:TRP:CH2	0.47	10	1
A:212:THR:HG21	A:264:ASN:HB3	0.47	9	1
A:72:PRO:CG	A:86:PHE:CZ	0.46	3	1
A:87:GLN:NE2	A:91:TYR:CD1	0.46	8	1
A:92:LEU:HA	A:96:LEU:HD12	0.46	4	1
A:144:ARG:CD	A:227:TYR:CZ	0.46	8	1
A:72:PRO:HG2	A:86:PHE:CZ	0.45	3	2
A:251:GLN:HA	A:271:TRP:CZ2	0.45	7	3
A:116:GLY:HA3	A:219:PHE:CE1	0.45	5	1
A:88:HIS:CG	A:227:TYR:CE2	0.45	8	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:163:LEU:HD22	A:226:PRO:HG2	0.45	1	1
A:112:TYR:OH	A:263:PHE:CE2	0.44	3	1
A:73:ASN:HD21	A:87:GLN:NE2	0.44	6	2
A:112:TYR:CE1	A:263:PHE:CE2	0.44	9	1
A:70:VAL:HG22	A:91:TYR:OH	0.44	6	1
A:69:TYR:HE2	A:86:PHE:CE1	0.44	7	1
A:151:LYS:HZ3	A:222:ASP:CG	0.44	4	1
A:110:PHE:C	A:110:PHE:CD1	0.43	2	2
A:114:ILE:CG2	A:200:VAL:HG12	0.43	5	2
A:96:LEU:HD21	A:102:LYS:HA	0.43	6	1
A:110:PHE:CE1	A:271:TRP:CH2	0.43	7	1
A:38:ILE:HG12	A:86:PHE:CD2	0.43	3	1
A:110:PHE:CD1	A:112:TYR:CZ	0.43	3	1
A:38:ILE:HG13	A:69:TYR:CE1	0.43	8	1
A:110:PHE:CE1	A:271:TRP:HH2	0.43	7	1
A:163:LEU:HD13	A:226:PRO:HG2	0.43	8	1
A:94:GLN:HE21	A:98:ASP:CG	0.43	7	2
A:111:VAL:HG22	A:228:VAL:HG23	0.42	5	1
A:163:LEU:CD2	A:226:PRO:HG3	0.42	8	1
A:144:ARG:NH2	A:221:GLU:OE1	0.42	9	1
A:66:PHE:CG	A:72:PRO:HG3	0.41	2	1
A:99:ARG:HD2	A:101:TYR:CE2	0.41	3	1
A:45:VAL:HG11	A:110:PHE:CE1	0.41	6	1
A:114:ILE:HG23	A:144:ARG:CZ	0.41	8	1
A:269:LYS:HE2	A:271:TRP:NE1	0.41	9	1
A:112:TYR:HE1	A:263:PHE:CE2	0.41	2	1
A:112:TYR:CE1	A:263:PHE:CD1	0.41	5	1
A:42:GLU:OE1	A:86:PHE:CD2	0.41	7	1
A:72:PRO:HB2	A:86:PHE:HZ	0.41	8	1
A:69:TYR:CD2	A:86:PHE:CE2	0.41	9	1
A:184:ALA:HB1	A:205:TYR:CE2	0.41	6	1
A:144:ARG:HE	A:221:GLU:CD	0.40	1	1
A:88:HIS:CE1	A:227:TYR:CE2	0.40	8	1
A:41:ALA:HB1	A:263:PHE:CZ	0.40	2	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	269	251	13	5
2	269	247	16	6
3	269	249	11	9
4	269	250	12	7
5	269	249	13	7
6	269	250	11	8
7	269	248	16	5
8	269	248	12	9
9	269	246	15	8
10	269	248	15	6

There are 21 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	20	LEU	10
A	23	CYS	10
A	18	LEU	9
A	21	VAL	7
A	17	SER	5
A	5	PHE	4
A	15	ILE	4
A	25	GLN	4
A	4	LYS	3
A	6	LYS	2
A	19	ALA	2
A	2	ALA	1
A	3	PHE	1
A	7	THR	1
A	8	PHE	1
A	9	ALA	1
A	10	ALA	1
A	11	VAL	1
A	22	GLY	1
A	24	GLY	1
A	269	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	224	215	6	3
2	224	219	0	5
3	224	216	3	5
4	224	215	7	2
5	224	218	3	3
6	224	215	5	4
7	224	217	4	3
8	224	217	2	5
9	224	217	3	4
10	224	219	3	2

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

Chain	Res	Type	Models (Total)
A	7	THR	10
A	14	LEU	10
A	17	SER	5
A	18	LEU	5
A	5	PHE	2
A	91	TYR	2
A	140	THR	1
A	227	TYR	1

5. Fit to Data Used for Modeling Assessment

5.2. Crosslinking-MS

5.2.1. Restraint types

This table summarizes information about crosslinker(s) used for data generation, and how crosslinking information was translated into actual modeling restraints. 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 24 crosslinking restraints combined in 24 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
L-Photo-Leucine	ILE	CA	LEU	CA	upper bound	10.00	1
L-Photo-Leucine	GLU	CA	LEU	CA	upper bound	10.00	4
L-Photo-Leucine	LEU	CA	VAL	CA	upper bound	10.00	5
L-Photo-Leucine	LEU	CA	PRO	CA	upper bound	10.00	1
L-Photo-Leucine	LEU	CA	TYR	CA	upper bound	10.00	4
L-Photo-Leucine	LEU	CA	LEU	CA	upper bound	10.00	3
L-Photo-Leucine	HIS	CA	LEU	CA	upper bound	10.00	1
L-Photo-Leucine	LEU	CA	SER	CA	upper bound	10.00	1
L-Photo-Leucine	ASP	CA	LEU	CA	upper bound	10.00	2
L-Photo-Leucine	LEU	CA	LYS	CA	upper bound	10.00	1
L-Photo-Leucine	GLY	CA	LEU	CA	upper bound	10.00	1

Distograms of individual restraints

Distograms (i.e., histogram plots of distances) provide an overview of distributions of distances between residues for which chemical crosslinks were identified. The shift of the distogram relative to the threshold value may indicate a poor model. 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.

5.2.2. 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=24)
1	1	1	10/10	All	50.00	50.00	24
				Self-links/ Intramolecular	50.00	50.00	24

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.

6. Fit to Data Used for Validation Assessment

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.