

Integrative Structure Validation Report

February 18, 2025 - 08:42 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	9A6F
PDB-Dev ID	PDBDEV_00000308
Structure Title	Integrative model of CODY-YPPF 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

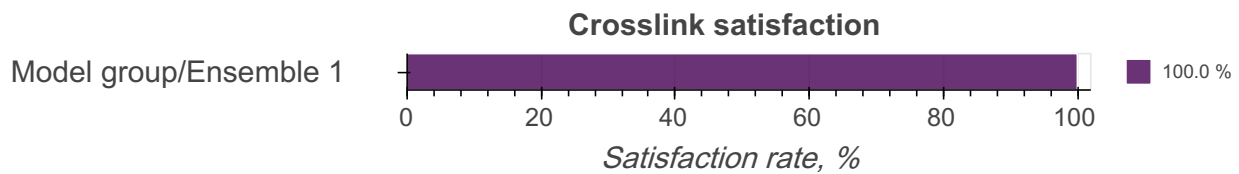
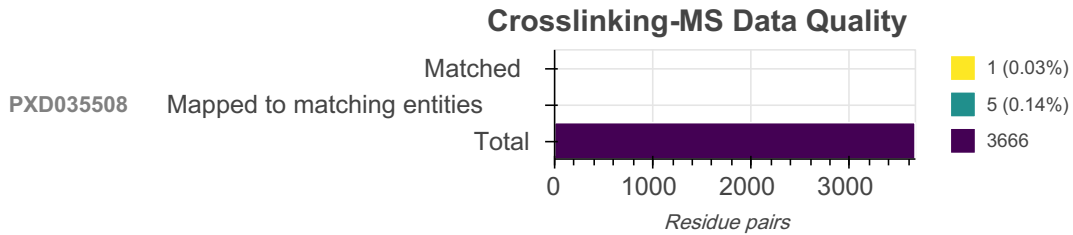
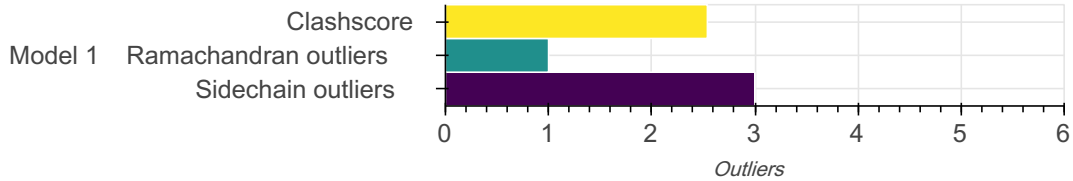
A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the  symbol.

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	CODY_BACSU	A	259	-	1-259	100.00 / 0.00	Atomic
		2	YPPF_BACSU	B	62	-	1-62	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	CODY_BACSU	dbseq_P39779_target	0.00	True
2	YPPF_BACSU	dbseq_P50834_target	0.00	True

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A6F	1	1 (100.00%)	1 (100.00%)
PXD035508	3666	5 (0.14%)	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 13 bond angle outliers in this entry (0.38% of 3456 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	96	THR	OG1-CB-CG2	5.07	99.16	109.30	1	1
A	130	ARG	NH1-CZ-NH2	4.96	112.85	119.30	1	1
A	56	GLN	OE1-CD-NE2	4.67	117.93	122.60	1	1
A	132	GLN	OE1-CD-NE2	4.62	117.98	122.60	1	1
A	107	GLN	OE1-CD-NE2	4.55	118.05	122.60	1	1
A	130	ARG	NE-CZ-NH2	4.42	123.18	119.20	1	1
A	5	GLN	OE1-CD-NE2	4.26	118.34	122.60	1	1
B	52	ASN	OD1-CG-ND2	4.26	118.34	122.60	1	1
A	15	GLN	OE1-CD-NE2	4.25	118.35	122.60	1	1
A	173	GLN	OE1-CD-NE2	4.22	118.38	122.60	1	1
A	139	ASP	CA-CB-CG	4.17	116.77	112.60	1	1
B	62	ASN	OD1-CG-ND2	4.16	118.44	122.60	1	1
A	134	GLN	OE1-CD-NE2	4.14	118.46	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	2.54	13

There are 13 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:1:MET:HE1	B:37:LEU:HD21	0.73	1	1
A:201:LEU:HD11	A:206:ILE:HD12	0.66	1	1
A:201:LEU:HD11	A:206:ILE:CD1	0.60	1	1
A:96:THR:HG22	A:98:PHE:HD2	0.56	1	1
B:1:MET:HE3	B:5:TYR:CD2	0.53	1	1
A:96:THR:C	A:98:PHE:H	0.46	1	1
A:75:TYR:CE1	A:98:PHE:CE1	0.45	1	1
A:25:LYS:HE2	A:29:GLU:OE2	0.44	1	1
A:96:THR:HG22	A:98:PHE:CD2	0.44	1	1
A:62:MET:SD	A:65:MET:HE2	0.42	1	1
B:1:MET:HE1	B:37:LEU:CD2	0.41	1	1
A:186:ALA:HB2	A:212:ILE:CD1	0.41	1	1
B:25:MET:HA	B:49:LEU:HD13	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	317	309	7	1

There are 1 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	97	ALA	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	280	273	4	3

There are 3 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	105	LEU	1
A	131	LEU	1
A	256	LEU	1

Fit of model to data used for modeling ?

Fit of model(s) to crosslinking-MS data

Restraint types

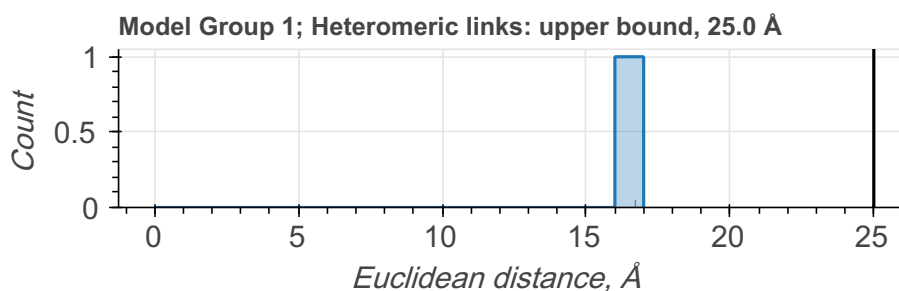
Restraint types are summarized in the table below. Restraints assigned "by-residue" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "coarse-grained". *Restraint group* represents a set of crosslinking restraints applied collectively in the modeling.

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	MET	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.



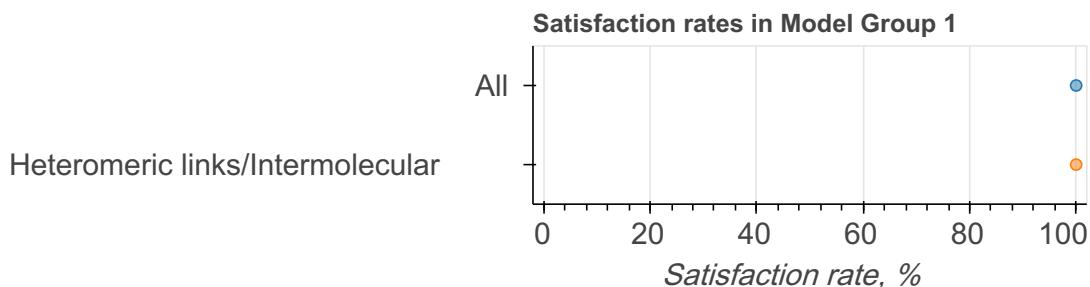
Satisfaction of restraints

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.