Integrative Structure Validation Report

October 09, 2025 - 04:50 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
PrISM Version dbe5a41
PyMOL Version 2.5.0

PDB ID	9A99 | pdb_00009a99
Structure Title	PTEN-4A model. Model restrained with crosslinking-MS data from "strained" conformation but no CTT. AlphaFold2-generated initial structure
Structure Authors	Dawson, J.E.; Eng, C.
Deposited on	2025-01-08

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 17 model(s). A total of 2 dataset(s) were used to build this entry.

Name	Type	Count
Crosslinking-MS data	Experimental data	1
De Novo model	Starting model	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

Model 8
Clashscore
Ramachandran outliers
Sidechain outliers
0 1 2 3 4 5 6 7 8
Outliers

Model 9
Clashscore
Ramachandran outliers
Sidechain outliers
0 1 2 3 4 5 6 7 8
Outliers

Model 10
Clashscore
Ramachandran outliers
Sidechain outliers
0 1 2 3 4 5 6 7 8
Outliers

Model 11
Clashscore
Ramachandran outliers
Sidechain outliers
0 1 2 3 4 5 6 7 8
Outliers

Model 12
Clashscore
Ramachandran outliers
Sidechain outliers
0 1 2 3 4 5 6 7 8
Outliers

Model 13
Clashscore
Ramachandran outliers
Sidechain outliers
0 1 2 3 4 5 6 7 8
Outliers

Model 14
Clashscore
Ramachandran outliers
Sidechain outliers
0 1 2 3 4 5 6 7 8
Outliers

Model 15
Clashscore
Ramachandran outliers
Sidechain outliers
0 1 2 3 4 5 6 7 8
Outliers

Model 16
Clashscore
Ramachandran outliers
Sidechain outliers
0 1 2 3 4 5 6 7 8
Outliers

Fit to Data Used for Modeling

2. Model Details

2.1. Ensemble information

This entry consists of 1 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	1-17	1	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN, PTEN-4A variant	A	403	-	1-403	100.00 / 100.00	Atomic

2.3. Datasets used for modeling

There are 2 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	De Novo model	Not available	Not available
2	Crosslinking-MS data	PRIDE	PXD060469

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	Protein structure prediction and 4A mutation (S380A/T382A/T383A/S385A)	AlphaFold2, PyMOL	Not available	1	False	False
2	1	Rosetta pre-relaxation	Rosetta relax protocol. Unrestrained. Lowest energy model of 200 used in next step	Not available	200	False	False
3	1	Restrained modeling	Rosetta, crosslinking-MS distance restraints	The individual models were restrained with subsets of experimental crosslink data in order to parse out specific restraints that correspond to a particular model. The crosslink list table provides the full set of experimental data obtained and the crosslink restraint table identifies the subset of crosslink restraints used in the modeling.	2000	False	False

There are 3 software packages reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	AlphaFold 2	ColabFold v1.5.0	Protein Structure Prediction	https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
2	Rosetta	2020.08	Model building and refinement	https://rosettacommons.org/software/
3	Rosetta	3.12	Model building and refinement	https://rosettacommons.org/software/

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).

Crosslinking-MS dataset is not available in the PRIDE Crosslinking database.

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 54 bond angle outliers in this entry (0.07% of 78251 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	361	SER	C-N-CA	4.87	130.47	121.70	10	17
A	369	VAL	C-N-CA	4.67	130.11	121.70	12	17
A	87	GLN	OE1-CD-NE2	4.44	118.16	122.60	3	3
A	370	SER	C-N-CA	4.40	129.62	121.70	10	17

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	0.92	6
2	0.61	4
3	1.23	8
4	0.61	4
5	1.23	8
6	1.53	10
7	0.77	5
8	0.61	4
9	1.07	7
10	0.77	5
11	0.77	5

Model ID	Clash score	Number of clashes
12	0.46	3
13	1.23	8
14	1.23	8
15	0.92	6
16	0.92	6
17	1.69	11

There are 108 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:60:LYS:HA	A:60:LYS:HE2	0.64	3	1
A:50:ILE:C	A:50:ILE:HD13	0.57	7	1
A:125:LYS:NZ	A:386:ASP:OD1	0.53	1	1
A:91:GLU:N	A:91:GLU:OE1	0.52	11	4
A:125:LYS:H	A:130:ARG:HH21	0.52	13	1
A:40:GLU:OE1	A:80:LYS:NZ	0.51	3	2
A:6:LYS:NZ	A:52:ASP:OD2	0.50	3	5
A:40:GLU:O	A:41:ARG:C	0.50	6	2
A:327:LYS:NZ	A:388:GLU:OE2	0.49	17	1
A:94:ASN:ND2	A:272:HIS:ND1	0.49	6	6
A:168:ILE:N	A:168:ILE:HD12	0.48	17	1
A:70:LEU:O	A:130:ARG:NH1	0.46	11	1
A:66:LYS:NZ	A:107:ASP:OD1	0.46	13	3
A:26:THR:OG1	A:159:ARG:NH1	0.46	12	9
A:322:LYS:N	A:341:PHE:O	0.45	1	2
A:102:LYS:HB3	A:103:PRO:HD3	0.45	2	14
A:154:PHE:O	A:155:TYR:C	0.45	1	9
A:322:LYS:NZ	A:334:ASN:OD1	0.45	15	8
A:91:GLU:OE2	A:254:LYS:NZ	0.45	13	1
A:77:ASP:C	A:79:ALA:H	0.44	14	2
A:88:TYR:N	A:89:PRO:HD3	0.44	11	3
A:65:TYR:C	A:65:TYR:CD1	0.43	14	3
A:81:PHE:CE2	A:123:HIS:CE1	0.43	14	1
A:27:TYR:CD1	A:27:TYR:N	0.42	16	2
A:69:ASN:HA	A:123:HIS:CE1	0.42	14	1
A:102:LYS:NZ	A:106:GLU:OE2	0.42	5	3

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:167:THR:O	A:329:ASN:ND2	0.42	1	1
A:280:ILE:HB	A:281:PRO:HD3	0.42	2	3
A:46:TYR:OH	A:386:ASP:N	0.42	7	1
A:148:ALA:HB3	A:176:TYR:CD1	0.42	9	1
A:170:SER:HB2	A:326:ASP:OD2	0.42	17	1
A:280:ILE:N	A:281:PRO:CD	0.41	16	7
A:170:SER:OG	A:173:ARG:NH2	0.41	17	2
A:67:ILE:HD13	A:81:PHE:CD1	0.41	7	1
A:133:VAL:CG2	A:171:GLN:HB3	0.41	9	1
A:177:TYR:CD1	A:177:TYR:N	0.40	2	1
A:96:PRO:O	A:97:GLN:C	0.40	13	1
A:38:PRO:HD3	A:81:PHE:CZ	0.40	4	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	401	373	26	2
2	401	378	21	2
3	401	382	18	1
4	401	378	21	2
5	401	380	18	3
6	401	379	18	4
7	401	377	22	2
8	401	376	23	2
9	401	376	22	3
10	401	379	21	1
11	401	373	24	4
12	401	379	19	3
13	401	374	26	1
14	401	375	22	4
15	401	380	19	2
16	401	381	18	2
17	401	379	19	3

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

Chain	Res	Type	Models (Total)
A	202	THR	17
A	38	PRO	8
A	78	THR	5
A	41	ARG	3
A	93	HIS	3
A	129	GLY	2
A	169	PRO	2
A	327	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	367	366	1	0
2	367	367	0	0
3	367	365	2	0
4	367	366	1	0
5	367	366	1	0
6	367	366	1	0
7	367	366	0	1
8	367	366	1	0
9	367	366	1	0
10	367	367	0	0
11	367	366	1	0
12	367	366	1	0
13	367	367	0	0
14	367	364	3	0
15	367	367	0	0
16	367	366	1	0
17	367	365	2	0

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

Chain	Res	Type	Models (Total)
A	50	ILE	1

4.2. PrISM Precision Analysis

*Regions of **low** **high** precision, defined as the variability among the models that satisfy the input data and calculated as the*

density-weighted root mean-square fluctuation (RMSF) from the bead/atom center of density, annotated and visualized using PrISM. The per-bead precision is computed from the deposited ensemble of superposed integrative models. High- and low-precision regions are then determined by clustering beads of similar precision based on their proximity in the structure. Only coarse-grained beads (or CA atoms for atomic models) of deposited models are used for assessment and visualization, and three projections for each representative model are generated.

PrISM analysis for Ensemble 1 (models deposited/total: 17/17).

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 54 crosslinking restraints combined in 54 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
DSSO	LYS	CA	LYS	CA	upper bound	30.00	46
DHSO	ASP	CA	ASP	CA	upper bound	30.00	1
DHSO	ASP	CA	GLU	CA	upper bound	30.00	5
DHSO	GLU	CA	GLU	CA	upper bound	30.00	2

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=54)
1	1	1	17/17	All	90.74	9.26	54
				Self-links/ Intramolecular	90.74	9.26	54

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.