

# Integrative Structure Validation Report

October 09, 2025 - 04:49 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*

PDB ID	9A90   pdb_00009a90
Structure Title	PTEN model. Model restrained with crosslinking-MS data from "strained" conformation and CTT. AlphaFold2-generated initial structure
Structure Authors	Dawson, J.E.; Eng, C.
Deposited on	2024-12-23

*This is a PDB-IHM Structure Validation Report.*

*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](#).*

## 1. Overview

### 1.1. Summary

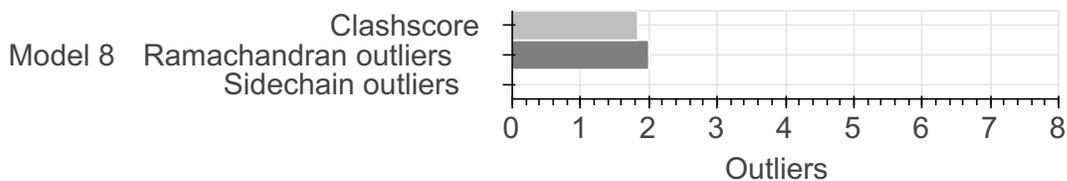
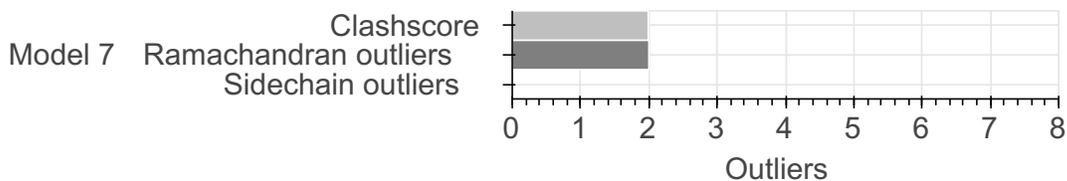
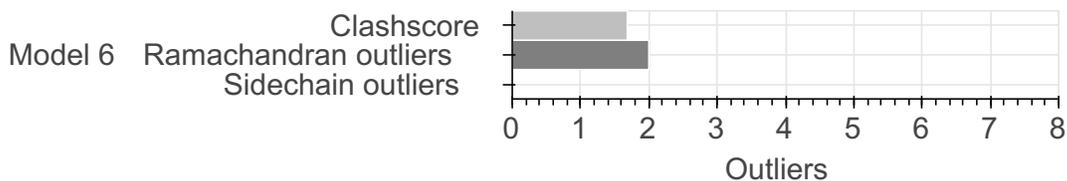
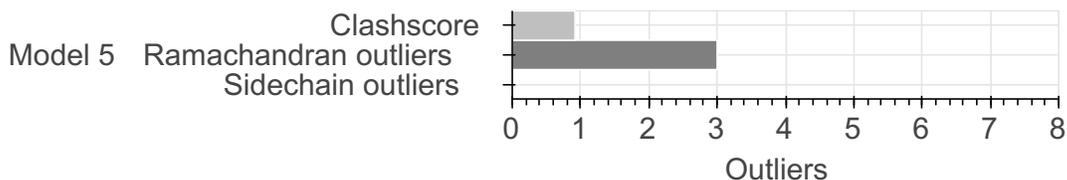
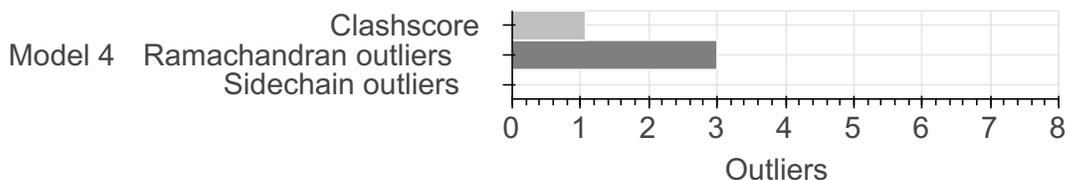
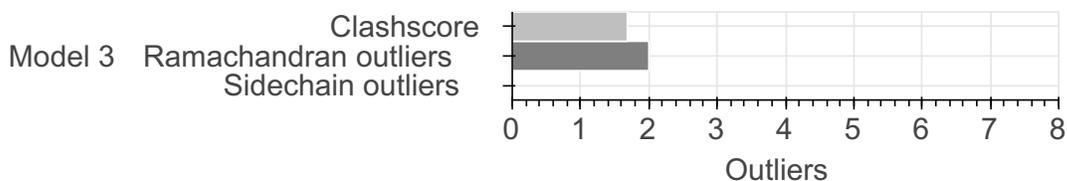
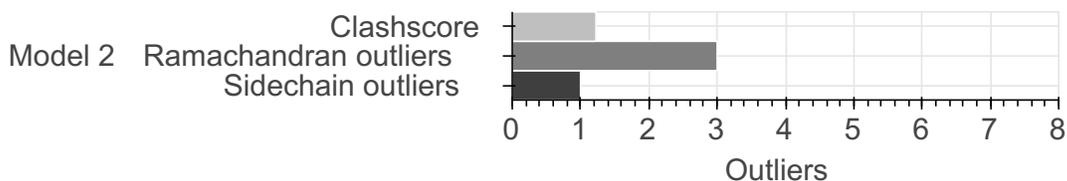
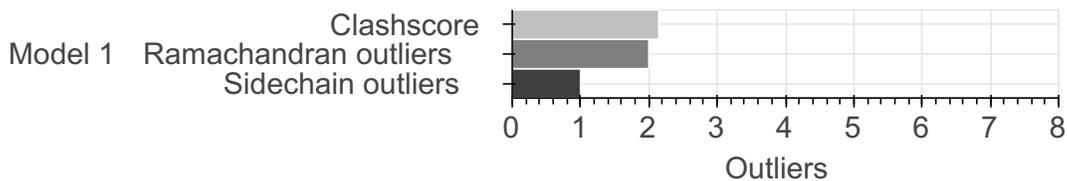
*This entry consists of 17 model(s). A total of 2 dataset(s) were used to build this entry.*

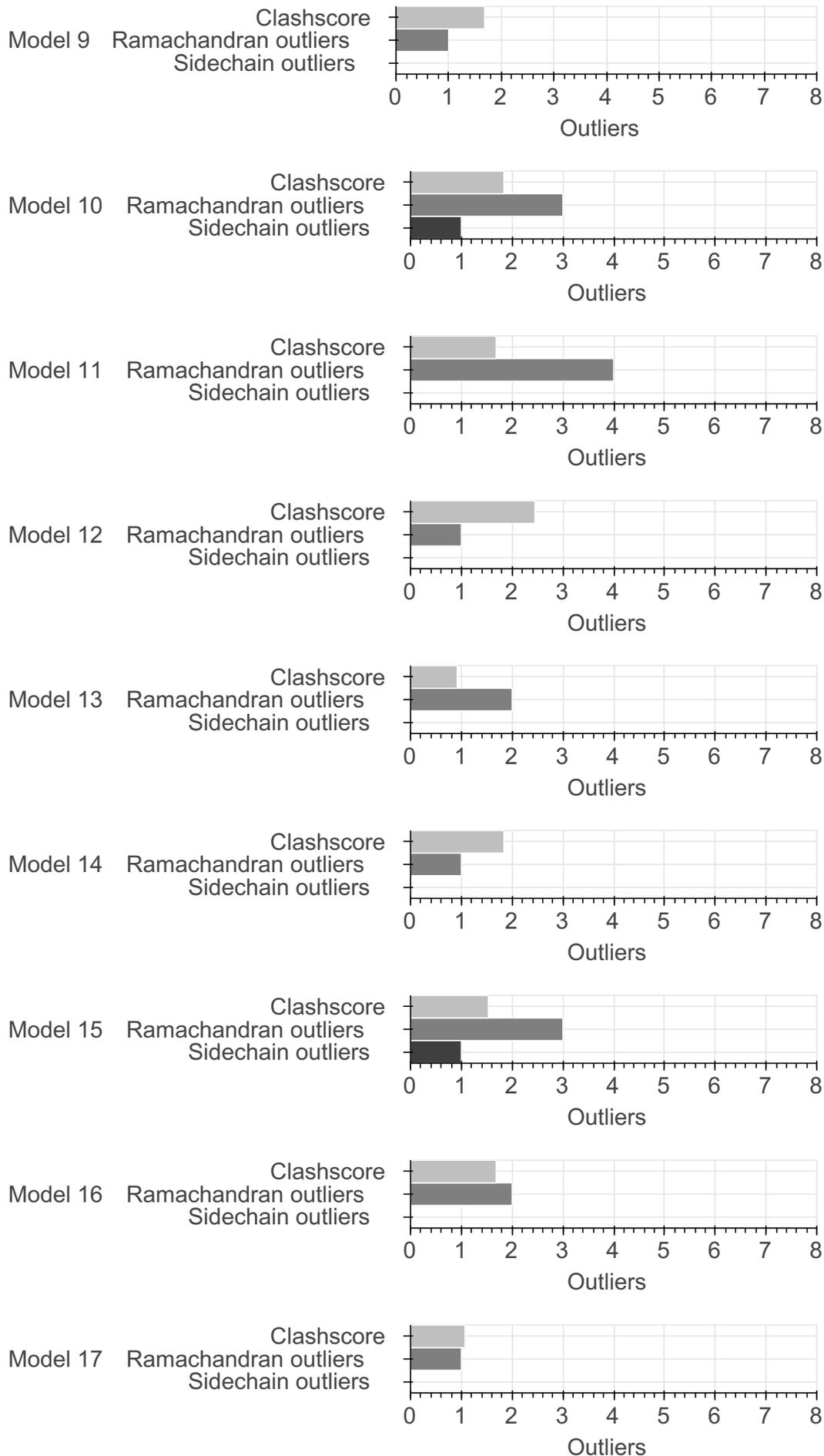
Name	Type	Count
Crosslinking-MS data	Experimental data	1
De Novo model	Starting model	1

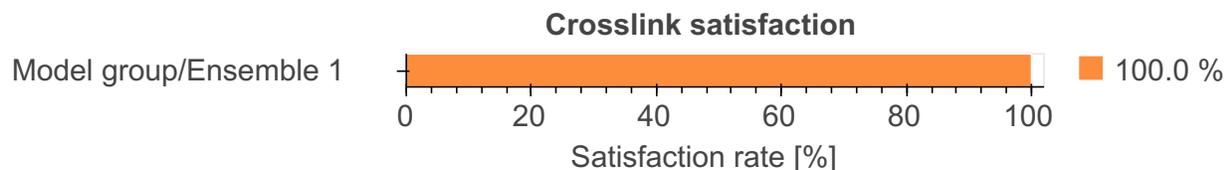
### 1.2. Overall quality

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





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	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	Crosslinking-MS data	PRIDE	<a href="#">PXD059974</a>
2	De Novo model	AlphaFoldDB	<a href="#">AF-P60484-F1</a>

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	AlphaFold2	Not available	1	False	False

Step number	Protocol ID	Method name	Method type	Method description	Number of computed models	Multi state modeling	Multi scale modeling
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
2	<a href="#">AlphaFold2</a>	ColabFold v1.5.0	Protein Structure Prediction	<a href="https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb">https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb</a>
1	<a href="#">Rosetta</a>	2020.08	Model building and refinement	<a href="https://rosettacommons.org/software/">https://rosettacommons.org/software/</a>
3	<a href="#">Rosetta</a>	3.12	Model building and refinement	<a href="https://rosettacommons.org/software/">https://rosettacommons.org/software/</a>

### 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-particle 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 66 bond angle outliers in this entry (0.08% of 78387 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	4.86	130.45	121.70	7	17
A	369	VAL	C-N-CA	4.65	130.06	121.70	9	17
A	87	GLN	OE1-CD-NE2	4.44	118.16	122.60	10	15
A	370	SER	C-N-CA	4.41	129.64	121.70	6	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	2.14	14
2	1.23	8
3	1.68	11
4	1.07	7
5	0.92	6
6	1.68	11
7	1.99	13
8	1.84	12
9	1.68	11
10	1.84	12
11	1.68	11
12	2.45	16
13	0.92	6
14	1.84	12
15	1.53	10
16	1.68	11
17	1.07	7

There are 178 clashes. The table below contains the detailed list of all clashes based on a MolProbity analysis. Bad clashes are  $\geq 0.4$  Angstrom. The output is limited to 100 rows.

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:166:VAL:O	A:166:VAL:HG23	0.73	8	1
A:138:TYR:C	A:138:TYR:CD1	0.69	2	1
A:332:LYS:HA	A:332:LYS:HE2	0.66	17	1
A:322:LYS:C	A:322:LYS:HD3	0.65	12	1
A:221:LYS:HG3	A:221:LYS:O	0.65	8	1
A:155:TYR:HD1	A:155:TYR:N	0.62	15	1
A:155:TYR:CD1	A:155:TYR:N	0.62	15	1
A:330:LYS:NZ	A:386:ASP:OD1	0.61	14	1
A:42:LEU:C	A:42:LEU:HD23	0.58	3	1
A:124:CYS:N	A:131:THR:OG1	0.57	7	1
A:84:ARG:HG2	A:84:ARG:O	0.57	2	1
A:267:LYS:HD2	A:267:LYS:N	0.55	15	1
A:266:LYS:NZ	A:268:ASP:OD1	0.55	14	1
A:10:SER:O	A:11:ARG:C	0.55	9	1
A:92:ASP:OD2	A:327:LYS:NZ	0.54	13	1
A:329:ASN:O	A:332:LYS:NZ	0.54	6	1
A:254:LYS:NZ	A:256:GLU:OE2	0.53	1	1
A:168:ILE:N	A:168:ILE:HD12	0.53	3	2
A:13:LYS:NZ	A:34:ALA:O	0.53	7	1
A:196:HIS:NE2	A:348:THR:OG1	0.53	6	1
A:201:GLU:O	A:202:THR:CB	0.52	8	1
A:149:GLN:OE1	A:176:TYR:OH	0.51	9	1
A:162:ASP:OD2	A:164:LYS:NZ	0.51	4	1
A:25:LEU:C	A:25:LEU:HD23	0.50	14	3
A:254:LYS:NZ	A:256:GLU:OE1	0.50	16	1
A:316:LEU:N	A:316:LEU:HD12	0.50	14	8
A:37:PHE:O	A:39:ALA:N	0.50	11	2
A:236:ASP:OD2	A:237:LYS:NZ	0.50	15	2
A:6:LYS:NZ	A:52:ASP:OD2	0.50	12	3
A:19:ASP:OD2	A:60:LYS:NZ	0.49	1	3
A:123:HIS:CD2	A:123:HIS:N	0.49	1	1
A:126:ALA:HB2	A:327:LYS:H	0.49	11	1
A:102:LYS:HB3	A:103:PRO:HD3	0.49	7	10

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:237:LYS:NZ	A:310:ASP:OD2	0.49	8	1
A:322:LYS:N	A:341:PHE:O	0.49	12	1
A:173:ARG:O	A:177:TYR:N	0.49	2	1
A:269:LYS:HD2	A:269:LYS:N	0.49	16	1
A:142:ARG:HG2	A:142:ARG:O	0.48	8	1
A:93:HIS:H	A:130:ARG:HG3	0.48	11	1
A:73:GLU:N	A:73:GLU:OE1	0.48	12	1
A:73:GLU:CD	A:73:GLU:H	0.48	12	1
A:140:LEU:C	A:140:LEU:HD23	0.48	9	1
A:102:LYS:HB2	A:103:PRO:HD3	0.48	1	7
A:354:PRO:O	A:356:ASN:N	0.48	1	8
A:350:THR:O	A:351:VAL:C	0.48	8	1
A:152:LEU:O	A:166:VAL:N	0.48	9	1
A:188:TYR:C	A:188:TYR:CD1	0.48	10	2
A:353:GLU:C	A:354:PRO:O	0.47	1	8
A:144:LYS:CD	A:144:LYS:N	0.47	7	1
A:69:ASN:C	A:69:ASN:OD1	0.47	1	1
A:196:HIS:CD2	A:348:THR:HG1	0.47	6	1
A:40:GLU:OE2	A:80:LYS:NZ	0.46	14	1
A:345:LEU:N	A:345:LEU:HD12	0.46	1	2
A:269:LYS:NZ	A:384:ASP:OD2	0.46	9	1
A:206:PHE:CG	A:207:SER:N	0.46	17	1
A:11:ARG:NH2	A:332:LYS:O	0.46	12	1
A:154:PHE:O	A:155:TYR:C	0.46	9	2
A:154:PHE:O	A:156:GLY:N	0.46	3	2
A:139:LEU:N	A:139:LEU:HD23	0.45	10	1
A:168:ILE:H	A:168:ILE:HD12	0.45	3	1
A:322:LYS:NZ	A:334:ASN:OD1	0.45	10	1
A:65:TYR:C	A:65:TYR:CD1	0.45	16	1
A:124:CYS:N	A:131:THR:HG1	0.45	7	1
A:58:ASP:OD2	A:62:LYS:NZ	0.45	7	1
A:188:TYR:HD1	A:188:TYR:O	0.45	10	1
A:252:ASP:C	A:252:ASP:OD1	0.44	11	1
A:322:LYS:HD3	A:322:LYS:O	0.44	12	1
A:47:ARG:O	A:48:ASN:C	0.44	15	2

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:70:LEU:O	A:130:ARG:NH1	0.44	9	1
A:316:LEU:CD1	A:316:LEU:N	0.44	14	3
A:335:ARG:NH2	A:381:ASP:OD1	0.44	6	1
A:41:ARG:HG2	A:42:LEU:H	0.44	8	1
A:10:SER:O	A:12:ASN:N	0.44	15	1
A:27:TYR:CD1	A:27:TYR:N	0.43	5	1
A:91:GLU:O	A:92:ASP:C	0.43	14	1
A:13:LYS:NZ	A:24:ASP:OD1	0.43	17	2
A:40:GLU:O	A:43:GLU:HG3	0.43	16	1
A:10:SER:O	A:13:LYS:N	0.43	4	1
A:326:ASP:OD1	A:327:LYS:N	0.43	12	1
A:225:TYR:C	A:225:TYR:CD1	0.43	1	1
A:280:ILE:N	A:281:PRO:CD	0.42	5	13
A:118:HIS:O	A:142:ARG:NE	0.42	3	1
A:206:PHE:O	A:207:SER:C	0.42	11	1
A:144:LYS:HD3	A:144:LYS:N	0.42	7	1
A:168:ILE:HG22	A:170:SER:H	0.42	3	1
A:92:ASP:OD1	A:130:ARG:NH1	0.42	17	1
A:287:SER:O	A:288:GLU:C	0.42	8	3
A:192:ALA:HB2	A:365:VAL:H	0.42	16	1
A:259:HIS:O	A:260:LYS:C	0.42	2	2
A:334:ASN:H	A:389:ASN:HB3	0.42	12	1
A:92:ASP:O	A:274:TRP:NE1	0.42	13	1
A:166:VAL:O	A:166:VAL:HG12	0.42	9	1
A:221:LYS:CG	A:221:LYS:O	0.41	8	1
A:271:PHE:HB3	A:328:ALA:HB2	0.41	6	3
A:92:ASP:OD2	A:221:LYS:NZ	0.41	6	1
A:72:ALA:CB	A:90:PHE:H	0.41	8	1
A:201:GLU:O	A:202:THR:HB	0.41	10	2
A:80:LYS:N	A:381:ASP:O	0.41	7	1
A:330:LYS:NZ	A:384:ASP:OD1	0.41	7	1
A:37:PHE:HA	A:38:PRO:HD3	0.41	10	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	381	18	2
2	401	370	28	3
3	401	376	23	2
4	401	377	21	3
5	401	377	21	3
6	401	378	21	2
7	401	366	33	2
8	401	379	20	2
9	401	375	25	1
10	401	375	23	3
11	401	374	23	4
12	401	379	21	1
13	401	380	19	2
14	401	377	23	1
15	401	378	20	3
16	401	378	21	2
17	401	375	25	1

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

Chain	Res	Type	Models (Total)
A	202	THR	13
A	11	ARG	10
A	127	GLY	4
A	354	PRO	4
A	38	PRO	3
A	34	ALA	1
A	41	ARG	1
A	207	SER	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	371	367	3	1
2	371	367	3	1
3	371	369	2	0

Model ID	Analysed	Favored	Allowed	Outliers
4	371	369	2	0
5	371	368	3	0
6	371	370	1	0
7	371	369	2	0
8	371	368	3	0
9	371	371	0	0
10	371	368	2	1
11	371	369	2	0
12	371	369	2	0
13	371	369	2	0
14	371	369	2	0
15	371	367	3	1
16	371	368	3	0
17	371	370	1	0

There are 4 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	123	HIS	1
A	138	TYR	1
A	155	TYR	1
A	316	LEU	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.

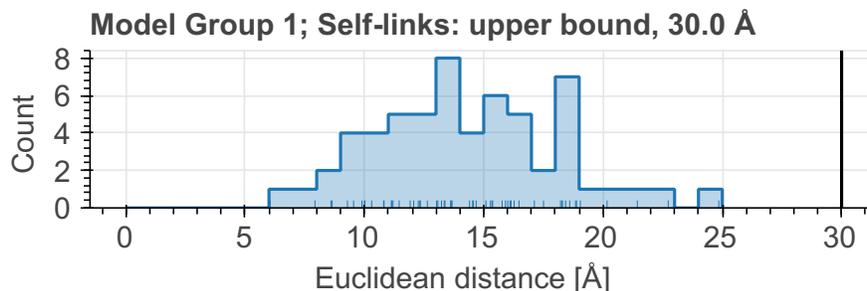
There are 59 crosslinking restraints combined in 59 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
DSSO	LYS	CA	LYS	CA	upper bound	30.00	44
DHSO	GLU	CA	GLU	CA	upper bound	30.00	9
DHSO	ASP	CA	GLU	CA	upper bound	30.00	4

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
DHSO	ASP	CA	ASP	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=59)
1	1	1	17/17	All	100.00	0.00	59
				Self-links/ Intramolecular	100.00	0.00	59

#### 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.*