Integrative Structure Validation Report

April 01, 2026 - 08:55 AM PDT

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

IHMValidation Version 3.1
Python-IHM Version 2.9
MolProbity Version 4.5.2
PrISM Version dbe5a41
PyMOL Version 2.5.0

PDB ID	9AAL | pdb_00009aal
Structure Title	Pennycress biotin carboxylase tetramer integrative model
Structure Authors	Van Doren, S.R.; You, Y.; Madison, H.J.; Pasa-Tolic, L.; Yokom, A.L.
Deposited on	2026-03-13

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

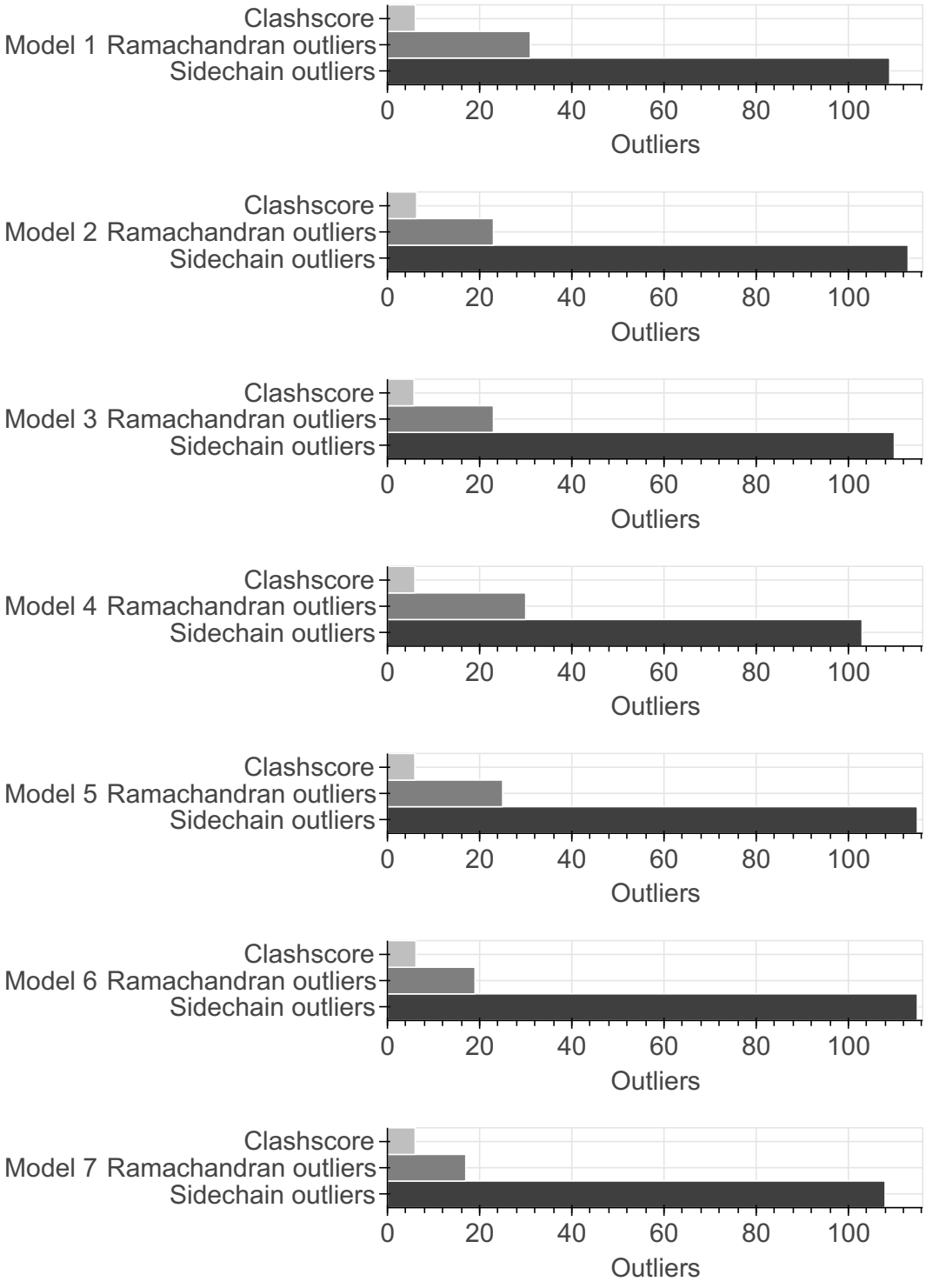
This entry consists of 7 model(s). A total of 2 dataset(s) were used to build this entry.

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

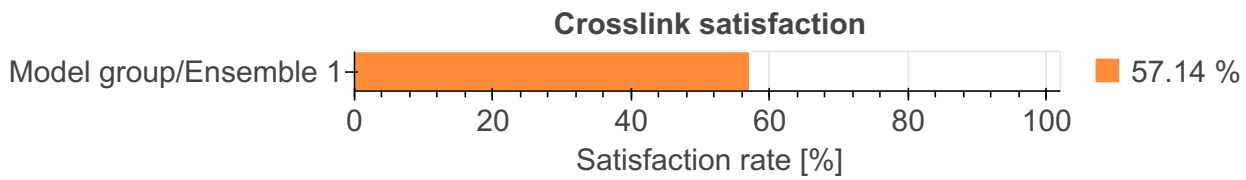
1.2. Overall quality

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	1-7	1	Biotin carboxylase	A	447	-	1-447	100.00 / 100.00	Atomic
				B					
				C					
				D					

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	MASSIVE	MSV000100391
2	Experimental model	PDB	pdb_00009zvm

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	Rigid body docking	Rigid body docking	Initial rigid body docking with crosslinking restraints	600	False	False
2	1	Semi-flexible refinement	Simulated annealing	Refinement with bulk water	120	False	False

There is 1 software package reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	HADDOCK	2.40	model building	https://www.bonvinlab.org/software/haddock2.4/

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 no bond angle outliers.

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	6.05	168
2	6.34	176
3	5.77	160
4	5.95	165
5	5.95	165
6	6.20	172
7	6.02	167

There are 1173 clashes. The table below contains the detailed list of all clashes based on a MolProbity analysis. Bad clashes are >= 0.4 Angstrom. The output is limited to 100 rows.

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
D:168:ALA:HB2	D:177:LEU:HD12	0.88	6	4
D:157:THR:HB	D:193:GLY:HA3	0.83	3	3
D:156:ALA:HB1	D:185:ALA:HB2	0.81	2	2

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
D:185:ALA:HA	D:189:PHE:HB3	0.81	5	2
A:341:PRO:HG3	A:418:TYR:HB2	0.80	4	3
B:274:LEU:HD22	B:284:MET:HG2	0.79	2	1
B:237:GLU:HG2	B:293:GLU:HG2	0.76	4	1
B:60:ILE:HB	B:61:PRO:HD3	0.73	5	3
A:54:ASN:HA	A:58:LEU:HB2	0.73	1	7
D:54:ASN:HA	D:58:LEU:HB2	0.73	4	7
C:438:ILE:HB	C:439:PRO:HD3	0.72	3	7
C:395:ARG:HH12	D:20:GLU:HB3	0.72	2	7
B:61:PRO:HG2	C:319:ARG:HG2	0.71	3	2
D:154:ILE:HG13	D:196:LEU:HD23	0.71	1	2
A:395:ARG:HH12	B:20:GLU:HB3	0.70	5	7
C:54:ASN:HA	C:58:LEU:HB2	0.70	6	7
D:245:THR:HG22	D:246:PRO:HD2	0.70	4	4
D:338:ALA:HA	D:414:THR:HA	0.69	1	1
A:438:ILE:HB	A:439:PRO:HD3	0.69	6	7
A:2:ILE:HG22	A:74:MET:HB2	0.69	6	2
D:154:ILE:HD11	D:194:VAL:HG13	0.68	2	3
D:438:ILE:HB	D:439:PRO:HD3	0.68	5	7
B:374:VAL:HG11	D:247:GLU:HG2	0.68	5	3
D:166:ARG:HB3	D:177:LEU:HD12	0.67	1	2
B:209:GLN:HE22	B:268:VAL:HG13	0.67	3	3
D:142:VAL:HB	D:175:VAL:HG23	0.67	3	2
B:60:ILE:HG23	B:88:PHE:HD1	0.67	5	2
A:343:LYS:HB2	A:346:ARG:HG2	0.66	5	2
D:102:PRO:HA	D:266:ILE:HG13	0.66	6	4
B:438:ILE:HB	B:439:PRO:HD3	0.66	6	7
B:34:LYS:HB3	D:245:THR:HG23	0.66	1	3
D:145:ALA:HB1	D:152:VAL:HG11	0.66	1	1
D:207:GLU:HB3	D:272:GLU:HG3	0.66	7	2
D:168:ALA:HB1	D:173:GLU:HB3	0.65	2	1
D:166:ARG:HB3	D:177:LEU:HD13	0.65	6	1
C:252:MET:HE3	C:281:PHE:HB3	0.65	2	1
C:60:ILE:HB	C:61:PRO:HD3	0.64	5	7
C:294:HIS:HB2	C:305:LEU:HD12	0.63	3	6

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
D:174:PHE:HZ	D:196:LEU:HD22	0.63	2	3
D:156:ALA:HA	D:194:VAL:HG12	0.63	7	2
B:32:ILE:HG22	D:246:PRO:HG2	0.63	3	1
A:295:PRO:HB2	A:387:ILE:HG12	0.63	1	7
C:340:ASP:HB3	C:413:PRO:HB2	0.62	3	3
D:138:THR:HG22	D:141:GLY:HA3	0.62	5	4
D:227:SER:HB2	D:433:VAL:HG12	0.62	4	4
B:54:ASN:HA	B:58:LEU:HB3	0.62	3	2
B:102:PRO:HA	B:266:ILE:HG13	0.62	5	6
D:273:PHE:HA	D:283:PHE:HA	0.61	6	7
B:32:ILE:HG23	B:376:PRO:HB3	0.61	5	1
A:362:PHE:HB3	A:389:TRP:HB3	0.61	7	7
A:74:MET:HE1	A:310:ILE:HG23	0.61	5	3
D:198:LYS:HG3	D:200:VAL:H	0.61	6	1
D:295:PRO:HB2	D:387:ILE:HG12	0.61	4	7
B:87:LEU:HG	C:215:PHE:CZ	0.61	3	1
A:60:ILE:HB	A:61:PRO:HD3	0.61	2	7
D:175:VAL:HG22	D:178:LEU:HB3	0.61	1	2
D:168:ALA:HB3	D:174:PHE:HB2	0.60	3	3
C:226:CYS:HB3	C:234:LYS:HG2	0.60	1	6
B:295:PRO:HB2	B:387:ILE:HG12	0.60	7	7
D:27:ALA:HB1	D:38:HIS:HD2	0.60	2	7
D:60:ILE:HB	D:61:PRO:HD3	0.60	2	7
B:207:GLU:HB3	B:272:GLU:HG3	0.60	6	7
B:58:LEU:O	B:59:LEU:HD22	0.60	2	1
D:176:LYS:HD3	D:177:LEU:HD22	0.60	1	1
A:227:SER:HB2	A:433:VAL:HG12	0.60	5	7
D:102:PRO:HG2	D:288:THR:HB	0.60	5	1
D:227:SER:HB2	D:433:VAL:HG22	0.60	1	1
B:362:PHE:HB3	B:389:TRP:HB3	0.59	2	7
D:3:LEU:HD22	D:67:ALA:HB2	0.59	1	7
B:1:LYS:HB3	B:72:CYS:HA	0.59	6	2
A:3:LEU:HD22	A:67:ALA:HB2	0.58	3	7
C:48:ILE:HG23	C:62:ASN:HB3	0.58	5	4
D:271:VAL:HG22	D:286:MET:HE2	0.58	5	3

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
C:27:ALA:HB1	C:38:HIS:HD2	0.58	2	7
D:168:ALA:HB1	D:173:GLU:HG3	0.58	4	2
B:349:PRO:HG2	B:377:PRO:HB3	0.58	7	5
A:27:ALA:HB1	A:38:HIS:HD2	0.58	7	7
C:4:VAL:HG21	C:11:ALA:HA	0.57	6	4
A:102:PRO:HG2	A:288:THR:HB	0.57	7	4
D:362:PHE:HB3	D:389:TRP:HB3	0.57	2	7
D:265:TYR:HE1	D:288:THR:HG22	0.57	6	6
B:46:VAL:HG11	B:66:ALA:HB1	0.57	5	1
A:209:GLN:HE21	A:268:VAL:HG22	0.56	5	4
C:317:LYS:HG2	C:318:LEU:H	0.56	5	1
D:143:ARG:O	D:144:LEU:HG	0.56	1	1
C:340:ASP:HB2	C:342:PHE:CE1	0.56	7	2
A:304:ASP:HB3	A:307:GLU:HB3	0.56	7	3
C:225:ASP:HB3	C:238:GLU:HG3	0.56	4	2
B:166:ARG:HB3	B:177:LEU:HD23	0.56	4	5
D:175:VAL:HG23	D:179:GLN:HE22	0.56	6	1
B:64:LEU:HD11	B:91:MET:HB3	0.56	2	1
C:227:SER:HB2	C:433:VAL:HG12	0.55	4	6
B:61:PRO:HG3	C:317:LYS:HB3	0.55	5	1
B:59:LEU:HD12	B:61:PRO:HD2	0.55	4	1
A:367:SER:HB3	A:384:GLY:HA3	0.55	1	2
C:367:SER:HB3	C:384:GLY:HA3	0.55	4	1
D:227:SER:O	D:435:THR:HA	0.55	2	7
D:173:GLU:HB3	D:177:LEU:HD23	0.55	1	2
C:207:GLU:HB3	C:272:GLU:HG3	0.55	4	6
A:337:ASN:HB3	A:382:LEU:HA	0.54	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	1780	1600	149	31
2	1780	1607	150	23
3	1780	1600	157	23
4	1780	1610	140	30

Model ID	Analysed	Favored	Allowed	Outliers
5	1780	1596	159	25
6	1780	1610	151	19
7	1780	1616	147	17

There are 42 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	77	PRO	7
A	81	PHE	7
A	282	TYR	7
A	283	PHE	7
A	431	GLY	7
B	77	PRO	7
C	77	PRO	7
C	81	PHE	7
D	77	PRO	7
B	216	GLY	6
B	283	PHE	6
C	283	PHE	6
D	81	PHE	6
D	194	VAL	6
D	192	ASP	5
D	200	VAL	5
D	216	GLY	5
A	279	GLY	4
B	315	GLY	4
B	369	VAL	4
B	370	TYR	4
D	157	THR	4
D	174	PHE	4
D	190	GLY	4
B	59	LEU	3
D	283	PHE	3
A	344	GLY	2
A	369	VAL	2
A	370	TYR	2

Chain	Res	Type	Models (Total)
B	60	ILE	2
B	62	ASN	2
B	81	PHE	2
C	216	GLY	2
C	315	GLY	2
C	369	VAL	2
C	370	TYR	2
A	340	ASP	1
B	31	THR	1
B	51	ALA	1
D	170	GLU	1
D	189	PHE	1
D	201	GLN	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	1480	1234	137	109
2	1480	1233	134	113
3	1480	1238	132	110
4	1480	1242	135	103
5	1480	1236	129	115
6	1480	1238	127	115
7	1480	1246	126	108

There are 234 unique sidechain outliers. Detailed list of outliers are tabulated below. The output is limited to 100 rows.

Chain	Res	Type	Models (Total)
A	10	ILE	7
A	154	ILE	7
A	196	LEU	7
A	320	TYR	7
A	393	ARG	7
B	127	THR	7
B	154	ILE	7
B	196	LEU	7

Chain	Res	Type	Models (Total)
B	317	LYS	7
B	320	TYR	7
B	342	PHE	7
B	414	THR	7
B	443	GLU	7
C	10	ILE	7
C	154	ILE	7
C	196	LEU	7
C	234	LYS	7
C	296	VAL	7
C	381	SER	7
C	414	THR	7
C	415	THR	7
C	443	GLU	7
D	135	LEU	7
D	140	GLU	7
D	167	LEU	7
D	196	LEU	7
D	293	GLU	7
D	296	VAL	7
D	320	TYR	7
D	327	LEU	7
D	443	GLU	7
D	447	GLU	7
A	327	LEU	6
B	9	GLU	6
B	10	ILE	6
B	59	LEU	6
B	81	PHE	6
B	293	GLU	6
B	299	MET	6
B	415	THR	6
C	33	ASP	6
C	244	LEU	6
C	321	LYS	6

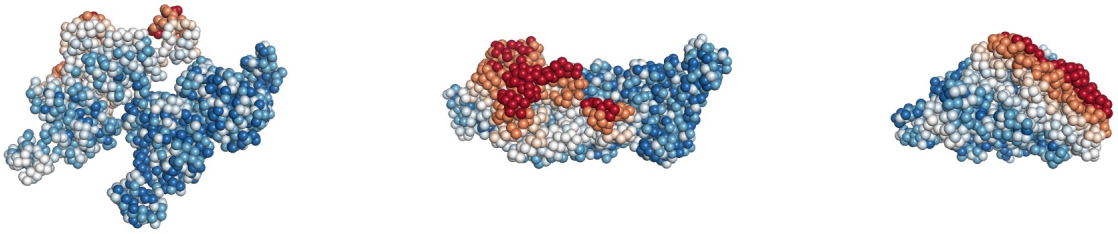
Chain	Res	Type	Models (Total)
C	323	GLU	6
C	328	ARG	6
D	9	GLU	6
D	23	ILE	6
D	81	PHE	6
D	127	THR	6
D	154	ILE	6
D	292	VAL	6
D	398	GLU	6
A	252	MET	5
A	272	GLU	5
A	299	MET	5
A	414	THR	5
B	139	GLU	5
B	167	LEU	5
B	272	GLU	5
B	296	VAL	5
B	304	ASP	5
B	406	ASP	5
B	434	ASP	5
C	299	MET	5
C	342	PHE	5
C	393	ARG	5
D	169	ASN	5
D	179	GLN	5
D	202	ASN	5
D	233	GLN	5
D	299	MET	5
D	321	LYS	5
D	414	THR	5
A	33	ASP	4
A	84	GLU	4
A	268	VAL	4
A	271	VAL	4
A	415	THR	4

Chain	Res	Type	Models (Total)
B	6	ASN	4
B	23	ILE	4
B	33	ASP	4
B	155	LYS	4
B	180	GLN	4
B	316	GLU	4
C	271	VAL	4
C	317	LYS	4
D	10	ILE	4
D	33	ASP	4
D	139	GLU	4
D	174	PHE	4
D	245	THR	4
D	289	ARG	4
D	335	ARG	4
D	415	THR	4
A	23	ILE	3
A	93	ARG	3
A	205	HIS	3
A	230	ARG	3
A	296	VAL	3
A	321	LYS	3

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: 7/7).



5. Fit to Data Used for Modeling Assessment

5.2. Crosslinking-MS

5.2.1. Restraint types

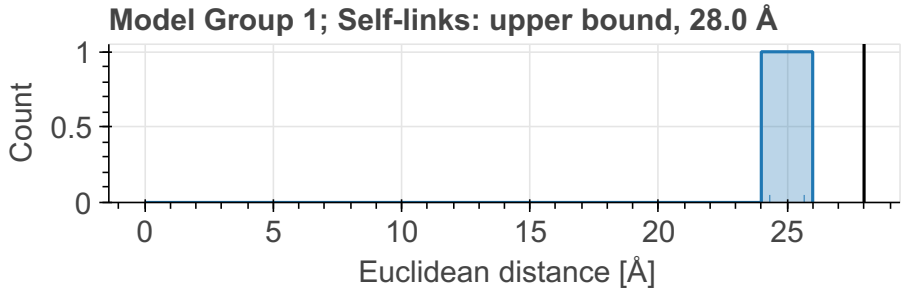
This table summarizes information about crosslinker(s) used for data generation, and how crosslinking information was translated into actual modeling restraints. Restraints assigned "by-residue" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "coarse-grained". Restraint group represents a set of crosslinking restraints applied collectively in the modeling.

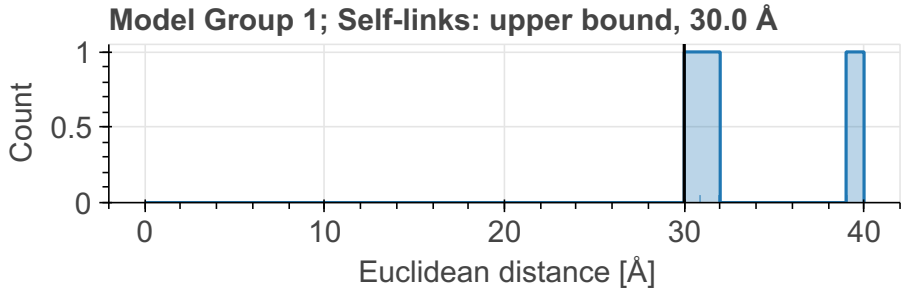
There are 7 crosslinking restraints combined in 7 restraint groups.

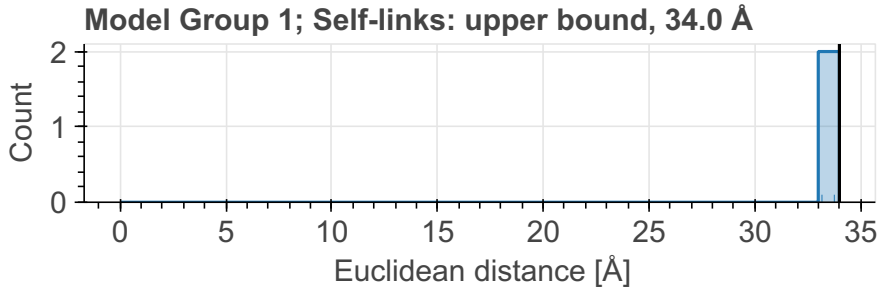
Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
DSBU	LYS	CA	LYS	CA	upper bound	28.00	2
DSBU	LYS	CA	LYS	CA	upper bound	30.00	3
DSBU	LYS	CA	LYS	CA	upper bound	34.00	2

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.







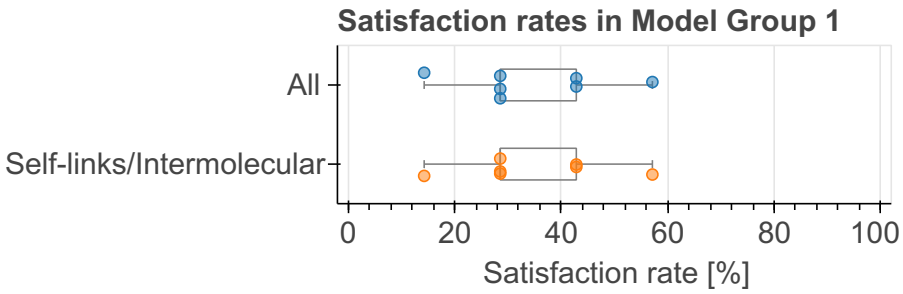
5.2.2. Satisfaction of restraints

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=7)
1	1	1	7/7	All	57.14	42.86	7
				Self-links/ Intermolecular	57.14	42.86	7

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.