

# Integrative Structure Validation Report

October 09, 2025 - 04:41 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	9A2H   pdb_00009a2h
PDB-Dev ID	PDBDEV_00000166
Structure Title	Model of E. coli GlgA 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](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

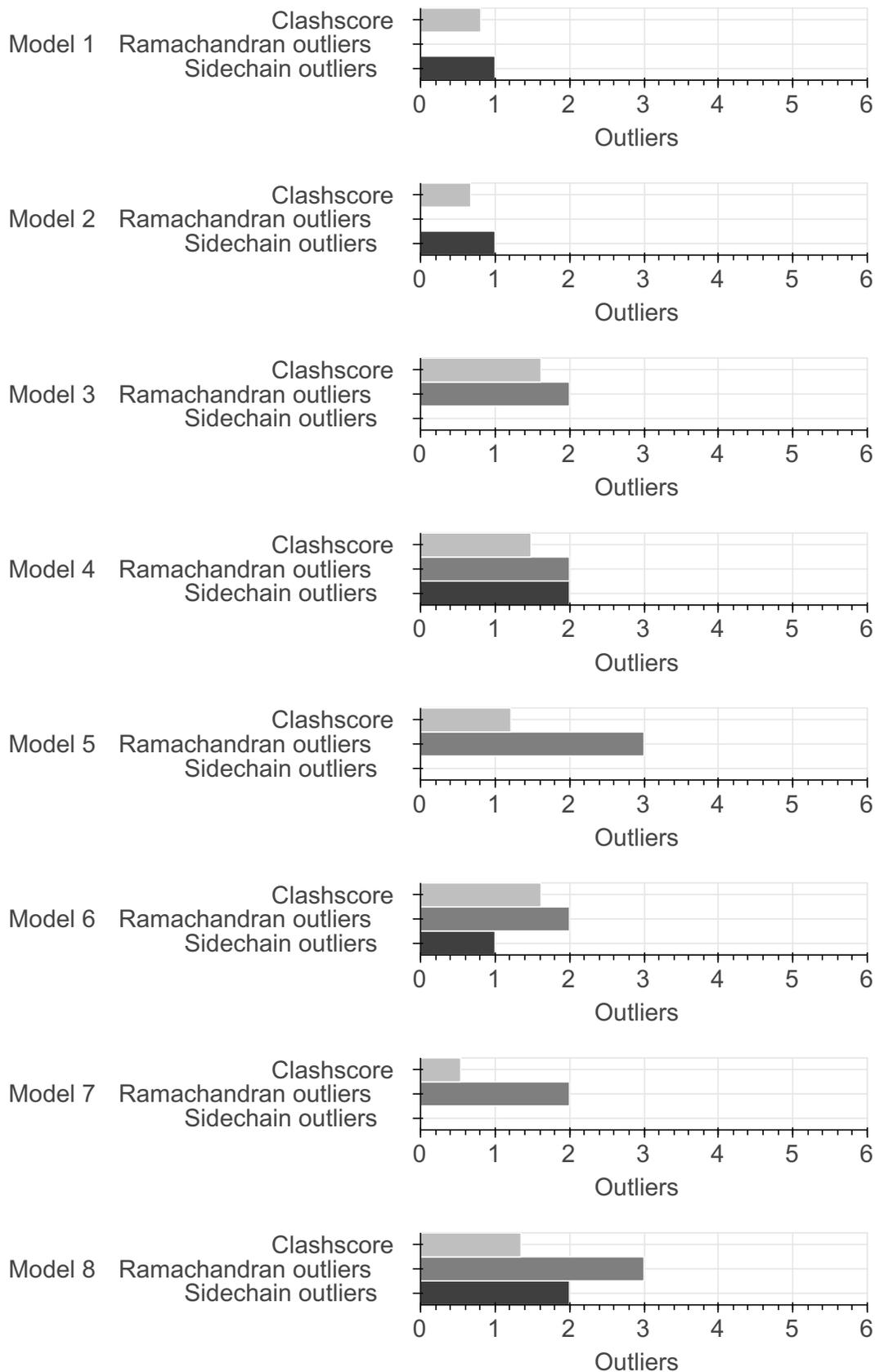
*This entry consists of 10 model(s). A total of 1 dataset(s) were used to build this entry.*

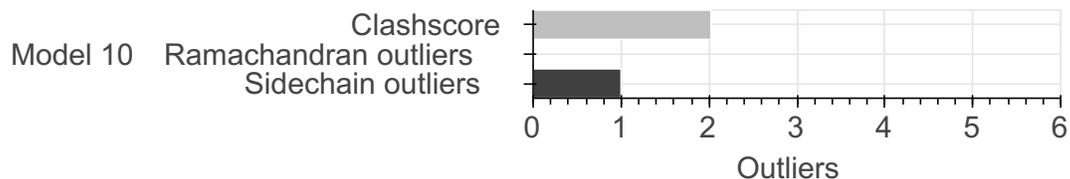
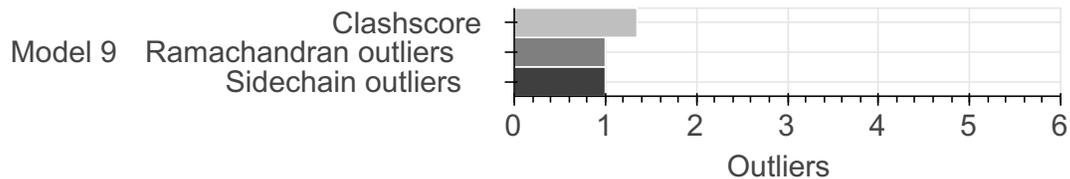
Name	Type	Count
Crosslinking-MS data	Experimental data	1

### 1.2. Overall quality

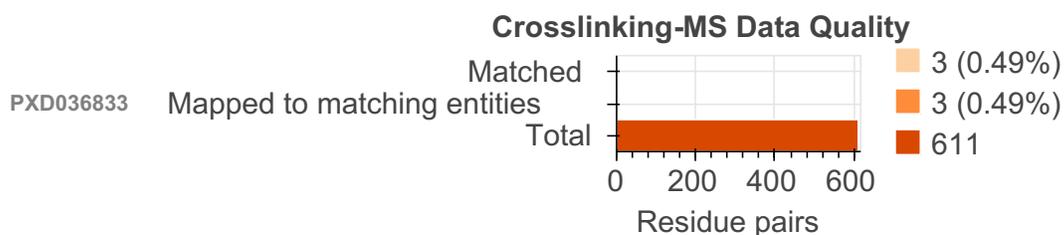
This validation report contains model quality assessments for all structures, data quality and fit to model assessments for SAS and crosslinking-MS datasets. Data quality and fit to model assessments for other datasets and model uncertainty are under development. Number of plots is limited to 256.

### Model Quality: MolProbity Analysis

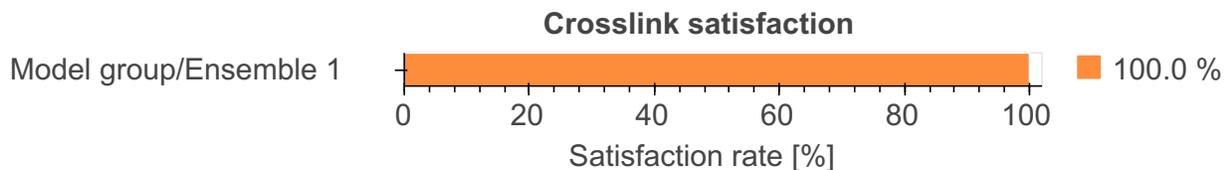




### Data Quality ?



### 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	P0A6U8	A	477	-	1-477	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	<a href="#">AlphaLink</a>	1.00	model building	<a href="https://github.com/lhatsk/AlphaLink">https://github.com/lhatsk/AlphaLink</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).

#### 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	P0A6U8	dbseq_P0A6U8_target	0.00	True

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A2H	3	3 (100.00%)	3 (100.00%)
PXD036833	611	3 (0.49%)	3 (0.49%)

## 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 1 bond length outliers in this entry (0.00% of 38350 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	90	ARG	CZ-NH2	4.17	1.28	1.33	6	1

#### Standard geometry: angle outliers

There are 122 bond angle outliers in this entry (0.23% of 52150 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	97	ASP	C-N-CA	7.39	135.00	121.70	8	1
A	90	ARG	NE-CZ-NH1	6.62	128.12	121.50	6	1
A	30	GLN	OE1-CD-NE2	6.34	116.26	122.60	1	4
A	454	GLN	OE1-CD-NE2	6.17	116.43	122.60	2	4
A	131	ASP	CA-CB-CG	6.13	118.73	112.60	9	10
A	139	HIS	CB-CG-CD2	5.79	123.67	131.20	4	3
A	246	ASN	OD1-CG-ND2	5.40	117.20	122.60	4	6
A	96	HIS	C-N-CA	5.32	131.27	121.70	8	1
A	352	GLN	OE1-CD-NE2	5.15	117.45	122.60	3	9
A	304	GLN	OE1-CD-NE2	5.11	117.49	122.60	8	2
A	397	ARG	CD-NE-CZ	5.10	131.54	124.40	5	2
A	463	GLN	OE1-CD-NE2	4.95	117.65	122.60	8	9
A	139	HIS	CB-CG-ND1	4.90	130.05	122.70	4	1
A	320	GLN	OE1-CD-NE2	4.85	117.75	122.60	9	6
A	178	GLN	OE1-CD-NE2	4.82	117.78	122.60	4	7
A	96	HIS	CB-CG-CD2	4.66	125.14	131.20	8	1
A	185	ASN	OD1-CG-ND2	4.65	117.95	122.60	9	4
A	109	ARG	NH1-CZ-NH2	4.58	113.35	119.30	8	2
A	98	THR	N-CA-C	4.53	123.70	111.00	8	1
A	300	ARG	NE-CZ-NH2	4.48	123.23	119.20	10	1
A	235	ARG	NE-CZ-NH2	4.41	115.23	119.20	7	1
A	194	GLN	OE1-CD-NE2	4.41	118.19	122.60	3	5

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	356	HIS	CB-CG-CD2	4.33	125.57	131.20	6	8
A	86	HIS	CB-CG-CD2	4.33	125.57	131.20	8	9
A	437	ARG	NE-CZ-NH1	4.26	125.76	121.50	8	3
A	136	HIS	CB-CG-CD2	4.25	125.67	131.20	3	8
A	95	TYR	C-N-CA	4.25	129.35	121.70	8	1
A	183	PHE	CA-CB-CG	4.19	117.99	113.80	4	1
A	2	GLN	OE1-CD-NE2	4.16	118.44	122.60	10	2
A	281	GLN	OE1-CD-NE2	4.14	118.46	122.60	10	5
A	470	ARG	CD-NE-CZ	4.11	130.15	124.40	1	1
A	197	PHE	CA-CB-CG	4.07	117.87	113.80	4	1
A	378	PRO	N-CA-CB	4.07	107.48	103.00	10	1
A	209	THR	CA-CB-OG1	4.00	115.60	109.60	4	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	0.81	6
2	0.67	5
3	1.62	12
4	1.48	11
5	1.21	9
6	1.62	12
7	0.54	4
8	1.35	10
9	1.35	10
10	2.02	15

There are 94 clashes. The table below contains the detailed list of all clashes based on a MolProbity analysis. Bad clashes are  $\geq 0.4$  Angstrom.

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:379:CYS:SG	A:401:LEU:HD21	0.71	10	1
A:202:LEU:HD12	A:228:MET:HE3	0.70	10	1
A:14:LEU:HD21	A:94:PRO:HG3	0.69	3	4
A:179:LEU:HD23	A:184:PHE:CE2	0.69	9	1
A:380:GLY:HA3	A:383:GLN:HE22	0.66	3	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:260:LEU:HD13	A:358:ALA:HA	0.66	5	8
A:115:TRP:CD1	A:150:ARG:HH22	0.58	6	3
A:199:LYS:HA	A:228:MET:HE1	0.58	10	1
A:11:PHE:CD1	A:12:PRO:HA	0.57	2	3
A:199:LYS:CB	A:228:MET:HE1	0.56	10	1
A:206:ASP:HA	A:240:ARG:HH21	0.55	10	4
A:16:THR:HA	A:95:TYR:CE1	0.54	2	1
A:199:LYS:HG3	A:228:MET:HE1	0.53	6	1
A:199:LYS:CA	A:228:MET:HE1	0.53	10	1
A:429:TRP:CH2	A:433:ARG:NH2	0.52	5	1
A:235:ARG:HH12	A:240:ARG:CZ	0.52	4	1
A:384:LEU:HD21	A:404:THR:HG21	0.51	5	4
A:407:ASP:CG	A:437:ARG:HH22	0.49	9	2
A:292:VAL:HG13	A:321:GLY:O	0.48	3	2
A:2:GLN:HE21	A:128:TRP:CD1	0.47	10	4
A:119:GLU:CD	A:152:ARG:HH22	0.47	8	1
A:174:MET:SD	A:184:PHE:CE2	0.47	9	1
A:115:TRP:NE1	A:150:ARG:HH22	0.47	7	2
A:309:LEU:HD21	A:375:ARG:NH2	0.46	9	2
A:215:TYR:CE1	A:219:ILE:CG2	0.46	3	5
A:270:LEU:HD21	A:456:MET:SD	0.46	6	1
A:235:ARG:HH12	A:240:ARG:NH1	0.45	4	1
A:95:TYR:CE1	A:139:HIS:NE2	0.44	1	1
A:246:ASN:ND2	A:379:CYS:SG	0.44	4	1
A:301:LEU:HD12	A:335:GLN:HG2	0.44	5	1
A:22:VAL:HG21	A:161:HIS:HE1	0.44	6	1
A:384:LEU:HD22	A:460:PHE:CG	0.43	3	1
A:13:LEU:HD22	A:50:GLY:HA3	0.43	1	2
A:90:ARG:NH2	A:109:ARG:HD2	0.43	6	1
A:146:TYR:CD1	A:179:LEU:HD13	0.43	9	1
A:379:CYS:HB2	A:401:LEU:HG	0.43	10	1
A:421:PHE:CZ	A:438:ALA:HA	0.43	7	4
A:25:ALA:HB1	A:462:TRP:CD1	0.42	6	1
A:171:ALA:HB2	A:184:PHE:CZ	0.42	6	1
A:189:LEU:HG	A:200:ALA:HB2	0.42	4	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:187:HIS:HB3	A:203:TYR:CE1	0.42	3	1
A:248:VAL:CG2	A:384:LEU:HB3	0.42	4	1
A:179:LEU:HD23	A:184:PHE:CZ	0.42	8	1
A:186:ILE:HG21	A:203:TYR:CZ	0.42	8	1
A:14:LEU:HD21	A:94:PRO:HD3	0.42	8	1
A:380:GLY:CA	A:383:GLN:HE22	0.42	9	1
A:215:TYR:CE1	A:219:ILE:HG23	0.41	5	1
A:397:ARG:HE	A:424:GLU:CD	0.41	5	1
A:171:ALA:HA	A:195:ILE:CD1	0.41	9	1
A:118:ALA:HB1	A:152:ARG:HH11	0.41	8	1
A:210:ALA:HB1	A:215:TYR:CD2	0.41	8	2
A:245:LEU:HD21	A:404:THR:HG22	0.41	3	2
A:380:GLY:HA3	A:383:GLN:NE2	0.41	3	1
A:90:ARG:HH22	A:105:ASP:CG	0.41	6	1
A:161:HIS:O	A:378:PRO:HG2	0.41	10	1
A:198:LEU:HD23	A:228:MET:HE1	0.41	3	1
A:314:LEU:HD13	A:324:LEU:HD22	0.40	5	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	475	460	15	0
2	475	461	14	0
3	475	458	15	2
4	475	461	12	2
5	475	458	14	3
6	475	458	15	2
7	475	462	11	2
8	475	452	20	3
9	475	457	17	1
10	475	460	15	0

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

Chain	Res	Type	Models (Total)
A	380	GLY	5

Chain	Res	Type	Models (Total)
A	137	ASP	4
A	96	HIS	1
A	98	THR	1
A	100	LEU	1
A	101	PHE	1
A	187	HIS	1
A	399	GLY	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	382	377	4	1
2	382	375	6	1
3	382	377	5	0
4	382	373	7	2
5	382	374	8	0
6	382	376	5	1
7	382	378	4	0
8	382	374	6	2
9	382	373	8	1
10	382	375	6	1

There are 7 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	137	ASP	2
A	209	THR	2
A	139	HIS	1
A	181	TRP	1
A	212	SER	1
A	271	GLU	1
A	463	GLN	1

## 5. Fit to Data Used for Modeling Assessment ?

## 5.2. Crosslinking-MS ?

### 5.2.1. Restraint types ?

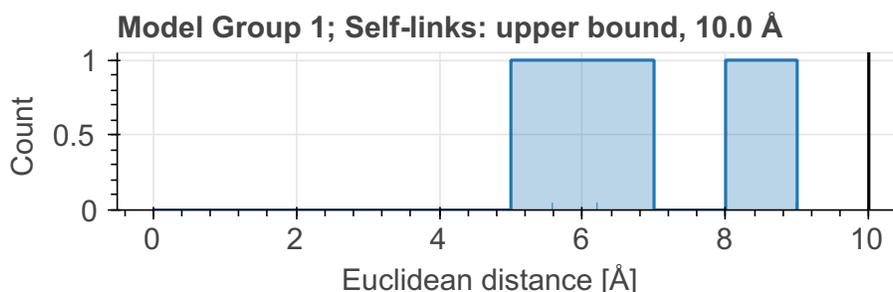
This table summarizes information about crosslinker(s) used for data generation, and how crosslinking information was translated into actual modeling restraints. Restraints assigned "by-residue" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "coarse-grained".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

There are 3 crosslinking restraints combined in 3 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
L-Photo-Leucine	ASP	CA	LEU	CA	upper bound	10.00	2
L-Photo-Leucine	LEU	CA	PRO	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.



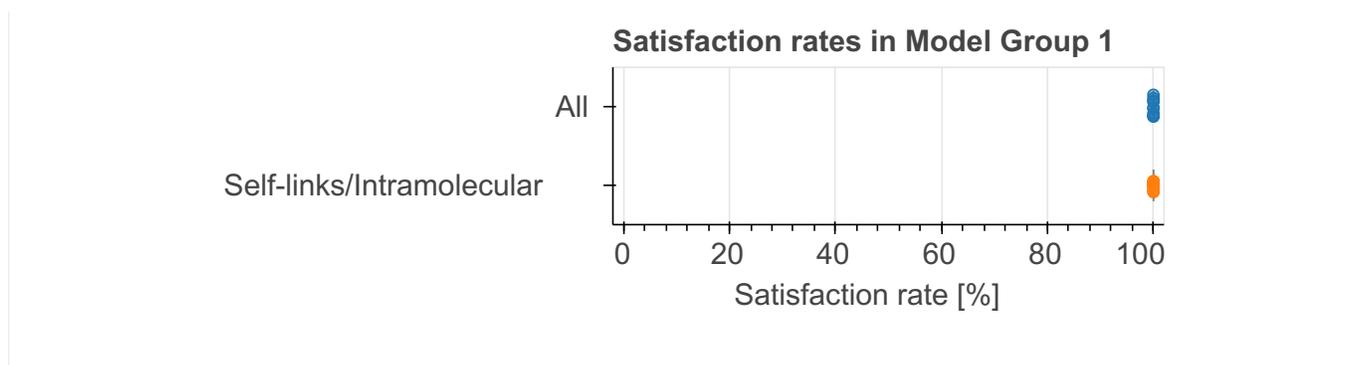
### 5.2.2. Satisfaction of restraints ?

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=3)
1	1	1	10/10	All	100.00	0.00	3
				Self-links/ Intramolecular	100.00	0.00	3

#### 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 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.*