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	9A2G | pdb_00009a2g
PDB-Dev ID	PDBDEV_00000165
Structure Title	Model of E. coli GroEL 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

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

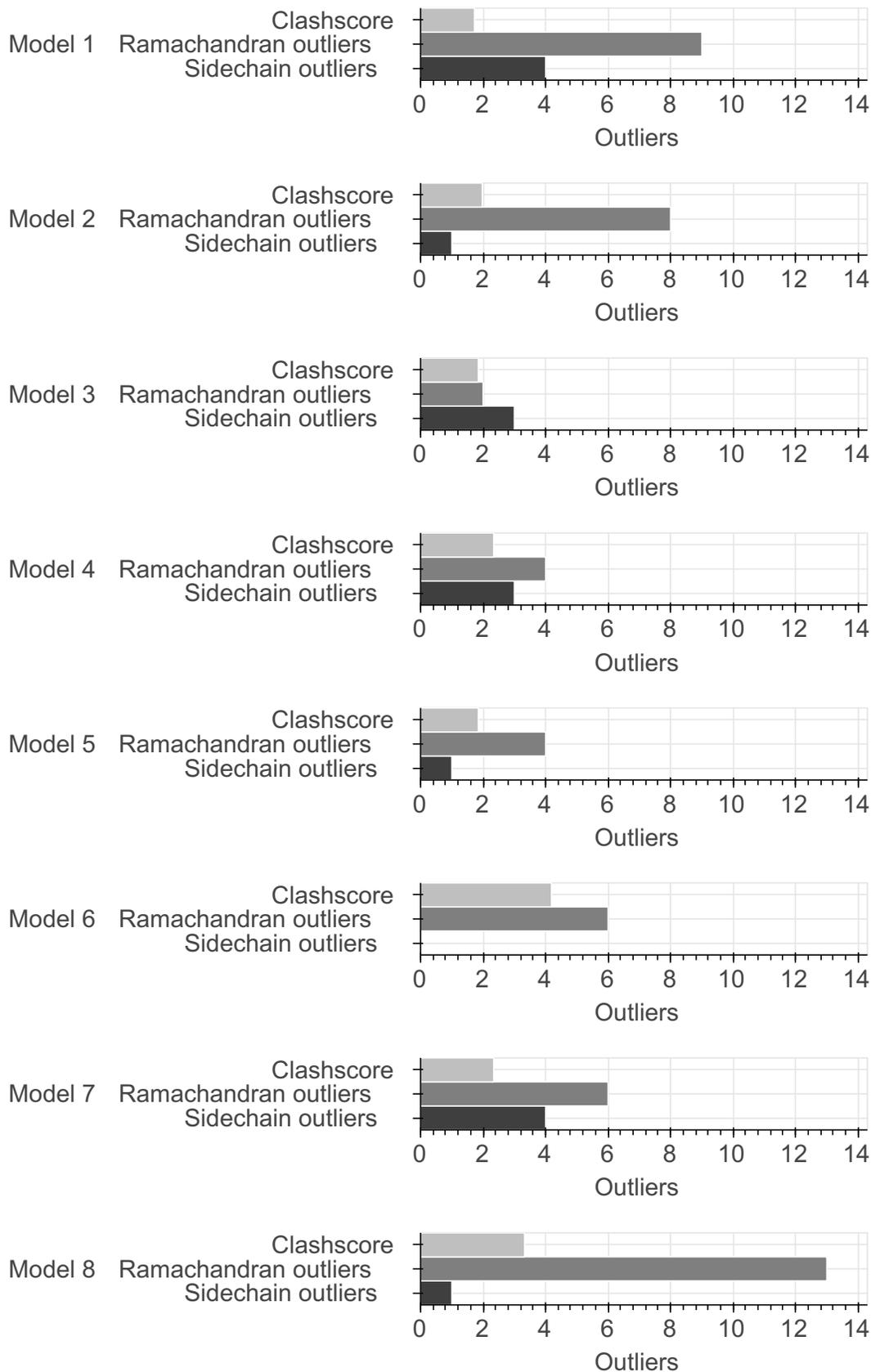*List of references used to build this report is available here.*

1. Overview ?

1.1. Summary ?

This entry consists of 10 model(s). A total of 1 dataset(s) were used to build this entry.
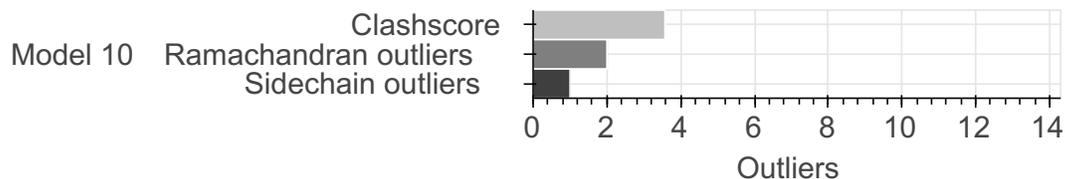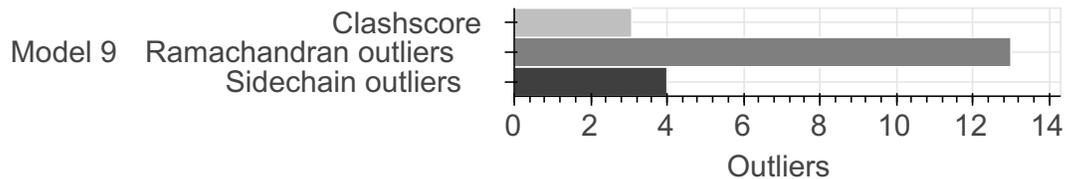
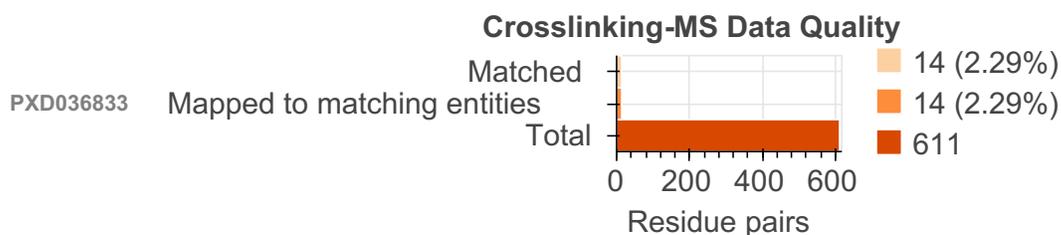| Name	Type	Count
Crosslinking-MS data	Experimental data	1

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 ⊙

Model 1

- Clashscore
- Ramachandran outliers
- Sidechain outliers

Outliers

Model 2

- Clashscore
- Ramachandran outliers
- Sidechain outliers

Outliers

Model 3

- Clashscore
- Ramachandran outliers
- Sidechain outliers

Outliers

Model 4

- Clashscore
- Ramachandran outliers
- Sidechain outliers

Outliers

Model 5

- Clashscore
- Ramachandran outliers
- Sidechain outliers

Outliers

Model 6

- Clashscore
- Ramachandran outliers
- Sidechain outliers

Outliers

Model 7

- Clashscore
- Ramachandran outliers
- Sidechain outliers

Outliers

Model 8

- Clashscore
- Ramachandran outliers
- Sidechain outliers

Outliers

Data Quality ?

Crosslinking-MS Data Quality

PXD036833

- 14 (2.29%)
- 14 (2.29%)
- 611

Fit to Data Used for Modeling ?

Crosslink satisfaction

Model group/Ensemble 1

- 50.0 %

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	P0A6F5	A	548	-	1-548	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	AlphaLink	1.00	model building	https://github.com/lhatsk/AlphaLink

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	P0A6F5	dbseq_P0A6F5_target	0.00	True

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A2G	14	14 (100.00%)	14 (100.00%)
PXD036833	611	14 (2.29%)	14 (2.29%)

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 151 bond angle outliers in this entry (0.28% of 54330 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	332	ILE	C-N-CA	8.67	137.31	121.70	1	1
A	224	ASP	CA-CB-CG	8.66	121.26	112.60	1	1
A	53	GLY	C-N-CA	8.02	136.13	121.70	9	1
A	194	GLN	OE1-CD-NE2	7.03	115.57	122.60	4	7
A	522	THR	C-N-CA	6.42	133.25	121.70	2	1
A	319	GLN	OE1-CD-NE2	6.12	116.48	122.60	8	3
A	332	ILE	CA-C-N	5.88	127.96	116.20	1	1
A	332	ILE	O-C-N	5.86	113.62	123.00	1	1
A	530	ALA	C-N-CA	5.85	132.23	121.70	8	1
A	541	MET	C-N-CA	5.75	132.04	121.70	4	1
A	547	MET	C-N-CA	5.67	131.91	121.70	9	4
A	285	ARG	NE-CZ-NH2	5.58	124.22	119.20	6	1
A	543	GLY	C-N-CA	5.56	131.71	121.70	9	5
A	72	GLN	OE1-CD-NE2	5.54	117.06	122.60	1	4
A	542	GLY	C-N-CA	5.32	131.28	121.70	2	1
A	121	ASP	CA-CB-CG	5.32	117.92	112.60	1	1
A	322	ARG	NE-CZ-NH2	5.14	123.83	119.20	2	2
A	334	ASP	CA-CB-CG	5.11	117.71	112.60	3	4
A	184	GLN	OE1-CD-NE2	5.02	117.58	122.60	1	10
A	146	GLN	OE1-CD-NE2	5.02	117.58	122.60	8	6
A	82	ASN	CA-CB-CG	4.97	107.63	112.60	7	1
A	546	GLY	C-N-CA	4.93	130.57	121.70	5	4
A	537	GLY	C-N-CA	4.86	130.45	121.70	4	1
A	366	GLN	OE1-CD-NE2	4.86	117.74	122.60	2	10
A	544	MET	C-N-CA	4.85	130.44	121.70	2	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	540	GLY	C-N-CA	4.85	130.43	121.70	4	2
A	538	MET	C-N-CA	4.82	130.38	121.70	4	1
A	224	ASP	N-CA-CB	4.82	102.31	110.50	1	1
A	280	GLY	C-N-CA	4.76	130.27	121.70	1	1
A	453	GLN	OE1-CD-NE2	4.75	117.85	122.60	7	5
A	505	GLN	OE1-CD-NE2	4.69	117.91	122.60	3	10
A	352	GLN	OE1-CD-NE2	4.66	117.94	122.60	10	10
A	541	MET	O-C-N	4.62	115.61	123.00	4	1
A	97	GLN	OE1-CD-NE2	4.60	118.00	122.60	6	1
A	256	GLY	CA-C-N	4.60	125.39	116.20	9	1
A	196	ASP	C-N-CA	4.56	129.90	121.70	6	2
A	538	MET	O-C-N	4.55	115.72	123.00	4	1
A	531	ASP	N-CA-CB	4.52	102.82	110.50	8	2
A	457	ASN	OD1-CG-ND2	4.47	118.13	122.60	7	1
A	83	ASP	CA-CB-CG	4.46	117.06	112.60	4	1
A	285	ARG	NH1-CZ-NH2	4.46	113.50	119.30	6	1
A	546	GLY	O-C-N	4.41	115.95	123.00	5	1
A	545	GLY	C-N-CA	4.40	129.62	121.70	1	2
A	41	ASP	C-N-CA	4.37	129.56	121.70	9	1
A	460	GLU	CB-CG-CD	4.35	105.21	112.60	4	2
A	368	ARG	CD-NE-CZ	4.35	130.49	124.40	1	1
A	43	SER	C-N-CA	4.34	129.51	121.70	9	1
A	348	GLN	OE1-CD-NE2	4.32	118.28	122.60	7	5
A	36	ARG	NH1-CZ-NH2	4.31	113.70	119.30	7	1
A	153	ASN	OD1-CG-ND2	4.26	118.34	122.60	1	1
A	542	GLY	O-C-N	4.26	116.19	123.00	2	1
A	413	ALA	C-N-CA	4.26	129.36	121.70	9	2
A	334	ASP	C-CA-CB	4.23	118.13	110.10	4	1
A	542	GLY	CA-C-N	4.21	124.63	116.20	2	1
A	522	THR	O-C-N	4.18	116.31	123.00	2	1
A	80	LYS	CG-CD-CE	4.14	101.78	111.30	1	1
A	37	ASN	OD1-CG-ND2	4.14	118.46	122.60	10	1
A	54	VAL	N-CA-CB	4.13	104.48	111.50	9	1
A	401	HIS	CB-CG-CD2	4.13	125.83	131.20	10	1
A	291	ASP	CA-CB-CG	4.12	116.72	112.60	8	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	432	GLN	OE1-CD-NE2	4.11	118.49	122.60	7	5
A	321	LYS	CG-CD-CE	4.10	120.73	111.30	7	1
A	202	PRO	CA-N-CD	4.08	106.29	112.00	7	1
A	206	ASN	OD1-CG-ND2	4.06	118.54	122.60	6	1
A	68	ASN	OD1-CG-ND2	4.04	118.56	122.60	5	1
A	229	ASN	OD1-CG-ND2	4.03	118.57	122.60	9	1
A	42	LYS	CG-CD-CE	4.02	102.05	111.30	8	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	1.72	14
2	1.97	16
3	1.85	15
4	2.34	19
5	1.85	15
6	4.19	34
7	2.34	19
8	3.33	27
9	3.08	25
10	3.57	29

There are 213 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)
A:40:LEU:HD12	A:57:ALA:HB2	0.99	8	1
A:16:MET:HB3	A:514:MET:HE3	0.75	7	3
A:20:VAL:HG21	A:514:MET:HE1	0.74	5	2
A:193:MET:HE1	A:291:ASP:HB3	0.72	8	2
A:40:LEU:HD11	A:56:VAL:O	0.71	10	1
A:29:VAL:CG2	A:90:THR:HG21	0.71	7	1
A:199:TYR:CE2	A:202:PRO:HD3	0.69	7	1
A:233:MET:CE	A:263:VAL:HG21	0.68	1	1
A:17:LEU:HD21	A:104:LEU:CD1	0.67	6	2
A:40:LEU:HD12	A:56:VAL:HA	0.67	7	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:73:MET:HE3	A:513:LEU:HD23	0.66	10	2
A:168:LYS:CD	A:189:VAL:HG21	0.65	8	2
A:198:GLY:HA3	A:277:LYS:H	0.65	10	1
A:31:LEU:HD12	A:90:THR:HG22	0.65	4	1
A:233:MET:HE2	A:263:VAL:HG21	0.64	1	1
A:248:LEU:HD22	A:323:VAL:HG11	0.64	10	6
A:194:GLN:HE21	A:330:THR:HG21	0.63	6	1
A:174:VAL:HG23	A:370:ALA:CB	0.63	8	1
A:77:VAL:HG21	A:510:VAL:HB	0.63	4	2
A:200:LEU:HD21	A:213:VAL:CG2	0.63	8	1
A:54:VAL:HG22	A:89:THR:HG21	0.62	3	2
A:16:MET:SD	A:514:MET:HE3	0.61	5	1
A:233:MET:HE1	A:249:ILE:CD1	0.61	6	2
A:194:GLN:NE2	A:330:THR:HG21	0.60	6	1
A:54:VAL:HG22	A:89:THR:HB	0.60	6	1
A:332:ILE:HG23	A:333:ILE:HG13	0.60	1	1
A:30:THR:HG22	A:36:ARG:O	0.58	10	1
A:26:ALA:HB1	A:40:LEU:HD21	0.58	8	1
A:168:LYS:HD3	A:189:VAL:HG21	0.58	8	3
A:168:LYS:HD2	A:189:VAL:HG21	0.57	8	2
A:17:LEU:HD21	A:104:LEU:HD12	0.57	6	2
A:26:ALA:HB1	A:40:LEU:HD23	0.57	7	1
A:27:VAL:HA	A:38:VAL:HG21	0.57	4	1
A:263:VAL:HG22	A:267:MET:HE2	0.57	6	1
A:8:PHE:CD1	A:519:CYS:HA	0.57	9	1
A:73:MET:CE	A:513:LEU:HD23	0.57	10	1
A:40:LEU:H	A:57:ALA:HB2	0.56	4	1
A:194:GLN:HE21	A:329:THR:HG21	0.56	4	1
A:299:THR:HG22	A:307:MET:HE1	0.55	2	1
A:58:ARG:HA	A:75:LYS:HE3	0.55	6	4
A:26:ALA:HB1	A:40:LEU:CD2	0.55	8	1
A:299:THR:CG2	A:307:MET:HE1	0.55	3	3
A:200:LEU:HD11	A:213:VAL:CG2	0.55	10	1
A:266:THR:HG22	A:273:VAL:H	0.54	4	1
A:23:LEU:HD11	A:57:ALA:HA	0.54	5	2

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:31:LEU:HD21	A:90:THR:HG21	0.54	9	1
A:6:VAL:CG1	A:520:MET:HE3	0.54	4	1
A:39:VAL:HG13	A:47:PRO:HB2	0.54	10	1
A:233:MET:SD	A:309:LEU:HD21	0.54	3	2
A:200:LEU:HD11	A:213:VAL:HG23	0.54	10	1
A:200:LEU:HD21	A:213:VAL:HG23	0.53	8	2
A:230:ILE:HD12	A:233:MET:HE2	0.53	8	1
A:39:VAL:HG13	A:47:PRO:CB	0.53	10	1
A:248:LEU:HD13	A:325:ILE:HD11	0.53	10	2
A:254:VAL:HG11	A:259:LEU:CB	0.53	6	1
A:233:MET:HE1	A:249:ILE:HD11	0.52	6	1
A:225:LYS:HE3	A:303:GLU:HB2	0.52	1	1
A:7:LYS:HE2	A:66:PHE:CE2	0.52	5	1
A:233:MET:HE1	A:249:ILE:HD13	0.52	8	1
A:160:LYS:HE2	A:164:GLU:OE2	0.52	1	1
A:38:VAL:HG21	A:56:VAL:HG21	0.52	3	1
A:199:TYR:CD2	A:254:VAL:HG11	0.51	9	1
A:54:VAL:HG21	A:87:ASP:OD1	0.51	4	1
A:30:THR:CG2	A:38:VAL:HG23	0.51	10	1
A:204:PHE:CE1	A:264:VAL:CG2	0.51	9	1
A:23:LEU:HD13	A:60:ILE:HG21	0.50	2	1
A:199:TYR:CZ	A:202:PRO:HG3	0.50	7	1
A:30:THR:HG21	A:38:VAL:HG23	0.50	4	1
A:279:PRO:HG2	A:288:MET:HG2	0.50	4	1
A:6:VAL:HG11	A:520:MET:HE3	0.50	4	1
A:193:MET:HE1	A:291:ASP:CB	0.50	8	1
A:77:VAL:HG12	A:92:ALA:HB1	0.49	9	3
A:262:LEU:HD22	A:273:VAL:HG11	0.49	8	1
A:193:MET:CB	A:295:LEU:HD22	0.49	6	1
A:221:LEU:HD22	A:233:MET:HE1	0.49	3	4
A:200:LEU:HD21	A:213:VAL:HG21	0.49	8	1
A:332:ILE:HG23	A:333:ILE:CG1	0.48	1	1
A:85:ALA:HB3	A:499:VAL:HG22	0.48	9	1
A:107:VAL:HG12	A:524:LEU:HD11	0.48	6	1
A:40:LEU:HB2	A:57:ALA:H	0.48	9	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:172:GLU:HB3	A:373:ALA:HB1	0.48	1	1
A:207:LYS:HE3	A:214:GLU:OE1	0.48	2	4
A:20:VAL:CG2	A:514:MET:HE1	0.48	5	1
A:3:ALA:CB	A:524:LEU:HD12	0.47	10	3
A:332:ILE:HG23	A:333:ILE:CD1	0.47	1	1
A:305:ILE:HD11	A:307:MET:HE2	0.47	2	1
A:169:VAL:HG23	A:173:GLY:HA3	0.47	6	1
A:254:VAL:HG11	A:259:LEU:HB2	0.47	6	1
A:27:VAL:CG1	A:90:THR:HG23	0.47	6	1
A:230:ILE:HG21	A:261:THR:HG21	0.47	8	1
A:204:PHE:CE1	A:264:VAL:HG21	0.46	9	1
A:52:ASP:HA	A:398:ASP:OD2	0.46	7	1
A:3:ALA:HB1	A:524:LEU:HD12	0.46	10	1
A:39:VAL:CG1	A:47:PRO:CB	0.46	10	1
A:256:GLY:HA2	A:259:LEU:H	0.46	9	1
A:353:ILE:HG23	A:362:ARG:HG3	0.46	4	2
A:128:VAL:HG22	A:501:ARG:HG3	0.46	6	1
A:190:VAL:HG13	A:331:THR:HG21	0.46	8	1
A:23:LEU:HD13	A:60:ILE:HD12	0.46	8	1
A:197:ARG:HD3	A:326:ASN:O	0.46	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	546	504	33	9
2	546	516	22	8
3	546	524	20	2
4	546	520	22	4
5	546	525	17	4
6	546	519	21	6
7	546	516	24	6
8	546	515	18	13
9	546	513	20	13
10	546	526	18	2

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

Chain	Res	Type	Models (Total)
A	414	GLY	7
A	529	ALA	4
A	544	MET	4
A	547	MET	4
A	531	ASP	3
A	5	ASP	2
A	57	ALA	2
A	59	GLU	2
A	255	GLU	2
A	536	GLY	2
A	538	MET	2
A	539	GLY	2
A	541	MET	2
A	7	LYS	1
A	8	PHE	1
A	31	LEU	1
A	43	SER	1
A	44	PHE	1
A	52	ASP	1
A	60	ILE	1
A	85	ALA	1
A	87	ASP	1
A	193	MET	1
A	196	ASP	1
A	197	ARG	1
A	199	TYR	1
A	201	SER	1
A	226	LYS	1
A	256	GLY	1
A	280	GLY	1
A	281	PHE	1
A	334	ASP	1
A	374	GLY	1
A	413	ALA	1
A	519	CYS	1

Chain	Res	Type	Models (Total)
A	521	VAL	1
A	528	ASP	1
A	534	ALA	1
A	537	GLY	1
A	542	GLY	1
A	543	GLY	1
A	546	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	415	405	6	4
2	415	409	5	1
3	415	410	2	3
4	415	406	6	3
5	415	410	4	1
6	415	407	8	0
7	415	405	6	4
8	415	408	6	1
9	415	403	8	4
10	415	407	7	1

There are 15 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	531	ASP	3
A	30	THR	2
A	31	LEU	2
A	48	THR	2
A	50	THR	2
A	541	MET	2
A	43	SER	1
A	52	ASP	1
A	55	SER	1
A	193	MET	1
A	199	TYR	1

Chain	Res	Type	Models (Total)
A	254	VAL	1
A	333	ILE	1
A	334	ASP	1
A	389	MET	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 14 crosslinking restraints combined in 14 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
L-Photo-Leucine	GLU	CA	LEU	CA	upper bound	10.00	6
L-Photo-Leucine	LEU	CA	LEU	CA	upper bound	10.00	1
L-Photo-Leucine	ILE	CA	LEU	CA	upper bound	10.00	1
L-Photo-Leucine	ASP	CA	LEU	CA	upper bound	10.00	2
L-Photo-Leucine	LEU	CA	VAL	CA	upper bound	10.00	1
L-Photo-Leucine	ALA	CA	LEU	CA	upper bound	10.00	1
L-Photo-Leucine	LEU	CA	SER	CA	upper bound	10.00	1
L-Photo-Leucine	LEU	CA	MET	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.

Model Group 1; Self-links: upper bound, 10.0 Å
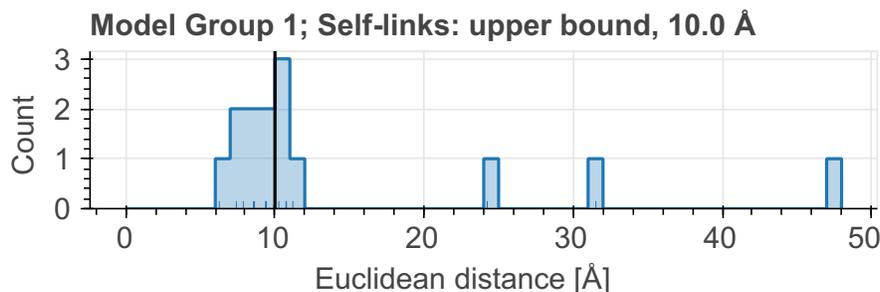
5.2.2. Satisfaction of restraints ⊙?

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=14)
1	1	1	10/10	All	50.00	50.00	14
				Self-links/ Intramolecular	50.00	50.00	14

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.

Satisfaction rates in Model Group 1

6. Fit to Data Used for Validation Assessment ⊙?

Validation for this section is under development.