Integrative Structure Validation Report

October 09, 2025 - 04:48 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	9A7P | pdb_00009a7p
PDB-Dev ID	PDBDEV_00000354
Structure Title	Integrative model of SIGA-RPOB by crosslinking MS and deep learning
Structure Authors	Stahl, K.; Brock, O.; Rappsilber, J.
Deposited on	2024-01-24

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

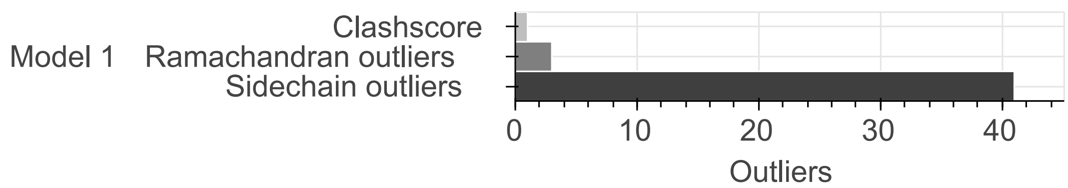
This entry consists of 1 model(s). A total of 1 dataset(s) were used to build this entry.

Name	Type	Count
Crosslinking-MS data	Experimental data	1

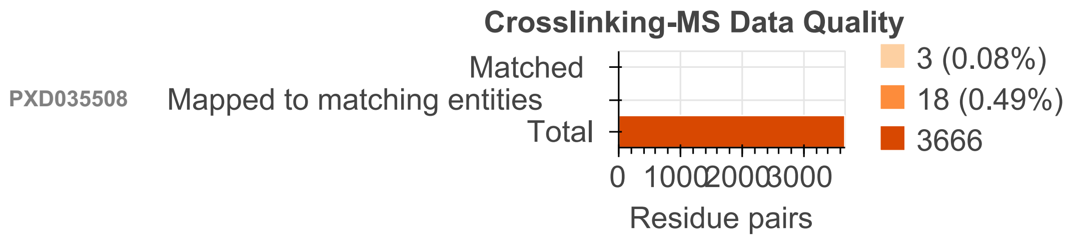
1.2. Overall quality

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.

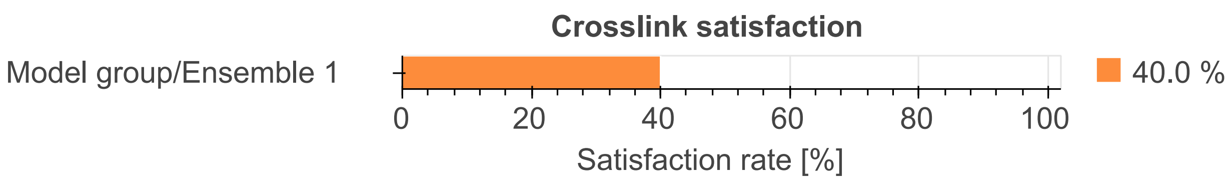
Model Quality: MolProbity Analysis



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	1	SIGA_BACSU	A	371	-	1-371	100.00 / 0.00	Atomic
		2	RPOB_BACSU	B	1193	-	1-1193	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	PRIDE	PXD035508

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	AlphaLink2	AlphaLink2	Not available	1	False	False

There is 1 software package reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	AlphaLink2	1.00	model building	https://github.com/Rappsilber-Laboratory/AlphaLink2

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).

PXD035508

Number of entities in the crosslinking-MS dataset:	810
Number of entities in the entry:	2

Matching entities:

Entity ID	Molecule name	Crosslinking-MS Entity ID	E-value	Exact match
1	SIGA_BACSU	dbseq_P06224_target	0.00	True
2	RPOB_BACSU	dbseq_P37870_target	0.00	True

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A7P	5	5 (100.00%)	3 (60.00%)
PXD035508	3666	18 (0.49%)	3 (0.08%)

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 60 bond angle outliers in this entry (0.35% of 17044 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	578	ASN	CA-CB-CG	5.91	118.51	112.60	1	1
A	69	GLU	C-N-CA	5.54	131.68	121.70	1	1
B	773	ASP	CA-CB-CG	5.39	117.99	112.60	1	1
B	720	GLN	OE1-CD-NE2	5.34	117.26	122.60	1	1
B	121	ASP	CA-CB-CG	5.25	117.85	112.60	1	1
B	1184	ASP	N-CA-C	5.04	125.10	111.00	1	1
B	1186	GLU	C-N-CA	5.03	130.76	121.70	1	1
B	578	ASN	OD1-CG-ND2	5.03	117.57	122.60	1	1
B	438	GLN	OE1-CD-NE2	4.89	117.71	122.60	1	1
B	197	ARG	NE-CZ-NH1	4.89	126.39	121.50	1	1
B	907	GLN	OE1-CD-NE2	4.76	117.84	122.60	1	1
A	59	GLN	OE1-CD-NE2	4.73	117.87	122.60	1	1
B	701	GLN	OE1-CD-NE2	4.73	117.87	122.60	1	1
B	1188	ASP	CA-CB-CG	4.71	117.31	112.60	1	1
B	838	ASP	CA-CB-CG	4.70	117.30	112.60	1	1
B	390	ASN	CA-CB-CG	4.67	117.27	112.60	1	1
B	175	ASP	CA-CB-CG	4.67	117.27	112.60	1	1
A	302	GLN	OE1-CD-NE2	4.66	117.94	122.60	1	1
B	355	GLN	OE1-CD-NE2	4.66	117.94	122.60	1	1
B	1190	VAL	CA-CB-CG1	4.65	118.31	110.40	1	1
B	538	ASN	CA-CB-CG	4.65	117.25	112.60	1	1
B	1183	ALA	N-CA-C	4.61	123.90	111.00	1	1
A	165	GLN	OE1-CD-NE2	4.59	118.01	122.60	1	1
B	50	GLN	OE1-CD-NE2	4.58	118.02	122.60	1	1
A	261	GLN	OE1-CD-NE2	4.56	118.04	122.60	1	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	469	GLN	OE1-CD-NE2	4.54	118.06	122.60	1	1
B	1187	ARG	N-CA-C	4.48	123.55	111.00	1	1
B	419	GLN	OE1-CD-NE2	4.43	118.17	122.60	1	1
B	612	ASP	CA-CB-CG	4.39	116.99	112.60	1	1
B	825	ASP	CA-CB-CG	4.37	116.97	112.60	1	1
A	19	GLN	OE1-CD-NE2	4.36	118.24	122.60	1	1
A	48	GLN	OE1-CD-NE2	4.35	118.25	122.60	1	1
A	15	GLN	OE1-CD-NE2	4.34	118.26	122.60	1	1
B	576	GLN	OE1-CD-NE2	4.33	118.27	122.60	1	1
B	51	ASP	CA-CB-CG	4.31	116.91	112.60	1	1
B	308	GLN	OE1-CD-NE2	4.30	118.30	122.60	1	1
B	280	HIS	CB-CG-CD2	4.29	125.62	131.20	1	1
A	76	GLN	OE1-CD-NE2	4.29	118.31	122.60	1	1
B	1073	GLN	OE1-CD-NE2	4.28	118.32	122.60	1	1
B	870	ARG	NH1-CZ-NH2	4.26	113.77	119.30	1	1
A	125	GLN	OE1-CD-NE2	4.25	118.35	122.60	1	1
B	359	GLN	OE1-CD-NE2	4.25	118.35	122.60	1	1
A	326	ASP	CA-CB-CG	4.24	116.84	112.60	1	1
B	606	ASN	OD1-CG-ND2	4.22	118.38	122.60	1	1
A	359	HIS	CB-CG-CD2	4.17	125.78	131.20	1	1
B	13	GLN	OE1-CD-NE2	4.14	118.46	122.60	1	1
A	5	GLN	OE1-CD-NE2	4.14	118.46	122.60	1	1
B	15	ARG	NE-CZ-NH2	4.12	122.91	119.20	1	1
B	4	GLN	OE1-CD-NE2	4.12	118.48	122.60	1	1
B	1185	VAL	N-CA-C	4.10	122.47	111.00	1	1
A	312	ASP	CA-CB-CG	4.09	116.69	112.60	1	1
B	646	GLN	OE1-CD-NE2	4.08	118.52	122.60	1	1
B	652	GLN	OE1-CD-NE2	4.06	118.54	122.60	1	1
B	447	GLN	OE1-CD-NE2	4.05	118.55	122.60	1	1
B	225	ASN	OD1-CG-ND2	4.04	118.56	122.60	1	1
B	1184	ASP	CA-C-N	4.04	124.28	116.20	1	1
B	375	ASN	OD1-CG-ND2	4.03	118.57	122.60	1	1
B	725	ASN	OD1-CG-ND2	4.03	118.57	122.60	1	1
A	67	ASN	OD1-CG-ND2	4.03	118.57	122.60	1	1
A	77	GLN	OE1-CD-NE2	4.02	118.58	122.60	1	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	1.01	25

There are 25 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)
B:1035:MET:HE3	B:1037:MET:SD	0.73	1	1
B:777:MET:HG3	B:781:LEU:HD12	0.58	1	1
A:70:THR:HG22	A:177:LYS:HZ1	0.58	1	1
B:471:MET:HE1	B:483:LYS:HE2	0.55	1	1
B:1135:LEU:HD22	B:1140:MET:CE	0.55	1	1
B:179:VAL:HG13	B:220:THR:HA	0.53	1	1
B:1135:LEU:HD22	B:1140:MET:HE1	0.52	1	1
B:791:HIS:CE1	B:1043:MET:HE1	0.51	1	1
A:70:THR:CG2	A:177:LYS:HZ1	0.50	1	1
B:957:LEU:HD11	B:1035:MET:HE1	0.50	1	1
A:293:HIS:CD2	B:1107:VAL:CG1	0.50	1	1
B:917:ARG:HH11	B:1043:MET:HE3	0.50	1	1
B:788:THR:HG21	B:916:LYS:HE2	0.50	1	1
B:781:LEU:HD11	B:954:ASP:HB3	0.48	1	1
B:978:MET:HE3	B:982:TYR:CZ	0.47	1	1
B:957:LEU:CD1	B:1035:MET:HE1	0.46	1	1
B:651:MET:CE	B:916:LYS:HE3	0.45	1	1
B:761:VAL:HG21	B:1035:MET:HE2	0.44	1	1
B:269:ASN:OD1	B:374:LYS:HE2	0.44	1	1
B:90:ALA:HB3	B:118:ILE:HD11	0.43	1	1
B:981:ARG:HH11	B:1034:ILE:HD11	0.43	1	1
A:17:LYS:HE3	A:52:TYR:CD1	0.41	1	1
B:674:VAL:HG21	B:712:LEU:HD23	0.41	1	1
B:658:VAL:HG22	B:976:MET:HE3	0.41	1	1
B:1075:PHE:CE1	B:1079:GLU:HB3	0.40	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	1560	1528	29	3

There are 3 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	70	THR	1
B	268	ALA	1
B	1185	VAL	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	1355	1259	55	41

There are 41 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	6	THR	1
A	9	THR	1
A	39	MET	1
A	288	THR	1
A	312	ASP	1
A	341	VAL	1
A	370	LEU	1
B	2	THR	1
B	48	MET	1
B	71	LEU	1
B	86	VAL	1
B	159	THR	1
B	161	THR	1
B	174	THR	1
B	189	LYS	1
B	192	VAL	1
B	255	LEU	1
B	319	LEU	1
B	353	THR	1
B	368	TYR	1

Chain	Res	Type	Models (Total)
B	391	LEU	1
B	440	THR	1
B	578	ASN	1
B	590	ASP	1
B	753	LEU	1
B	831	ARG	1
B	838	ASP	1
B	846	VAL	1
B	852	THR	1
B	861	LEU	1
B	942	GLU	1
B	1038	ILE	1
B	1044	VAL	1
B	1048	LEU	1
B	1132	ILE	1
B	1169	LEU	1
B	1180	THR	1
B	1185	VAL	1
B	1186	GLU	1
B	1189	VAL	1
B	1191	THR	1

5. Fit to Data Used for Modeling Assessment ?

5.2. Crosslinking-MS ?

5.2.1. Restraint types ?

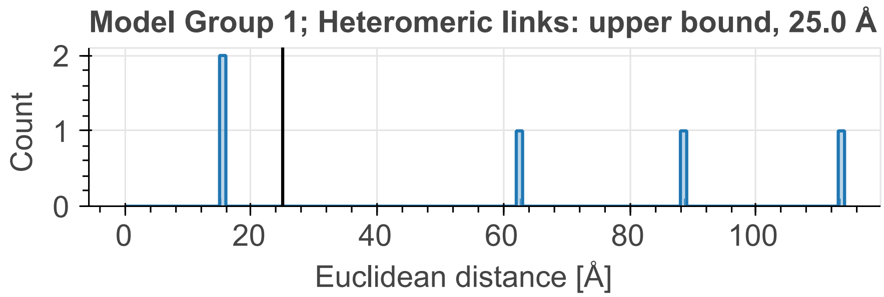
This table summarizes information about crosslinker(s) used for data generation, and how crosslinking information was translated into actual modeling restraints. Restraints assigned "by-residue" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "coarse-grained".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

There are 5 crosslinking restraints combined in 5 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
SDA	LYS	CA	LYS	CA	upper bound	25.00	5

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=5)
1	1	1	1/1	All	40.00	60.00	5
				Heteromeric links/ Intermolecular	40.00	60.00	5

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.