

Integrative Structure Validation Report ?

February 18, 2025 - 08:30 AM PST

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

Integrative Modeling Validation Version 2.0

Python-IHM Version 1.8

MolProbity Version 4.5.2

PDB ID	9A0W
PDB-Dev ID	PDBDEV_00000068
Structure Title	Plasmid replication initiator protein TrfA complexed with double stranded DNA
Structure Authors	Wegrzyn K; Zabrocka E; Bury K; Tomiczek B; Wieczor M; Czub J; Uciechowska U; Moreno-del Alamo M; Walkow U; Grochowina I; Dutkiewicz R; Bujnicki JM; Giraldo R; Konieczny I
Deposited on	2020-11-25

This is a PDB-IHM IM Structure Validation Report for a publicly released PDB-IHM entry.

We welcome your comments at helpdesk@pdb-ihm.org

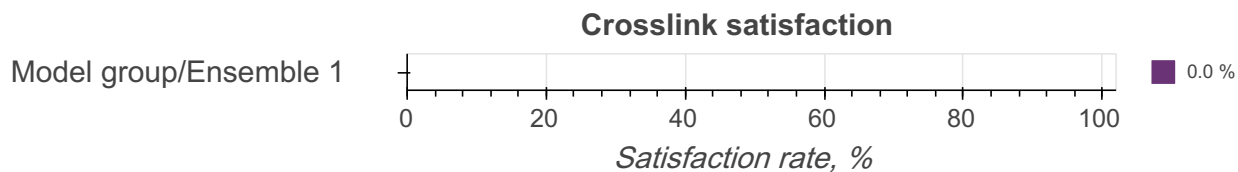
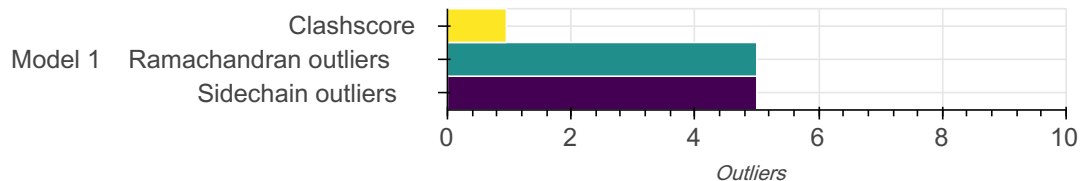
A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the ? symbol.

List of references used to build this report is available [here](#).

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



Ensemble information ?

This entry consists of 0 distinct ensemble(s).

Summary ?

This entry consists of 1 model(s). A total of 5 datasets were used to build this entry.

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	TrfA33	A	285	-	1-92, 93-99, 100-285	100.00 / 97.54	Atomic
		2	DNA (26-MER)	B	26	-	1-26	100.00 / 100.00	Atomic
		3	DNA (26-MER)	C	26	-	1-26	100.00 / 100.00	Atomic

Datasets used for modeling ?

There are 5 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	Crosslinking-MS data	PRIDE	PXD013286

ID	Dataset type	Database name	Data access code
2	De Novo model	Not available	Not available
3	Comparative model	Not available	10.1002/pro.68
4	Comparative model	Not available	10.1002/pro.68
5	Comparative model	Not available	10.1002/pro.68

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	None	Rosetta ab initio modeling	None	None	False	False
2	1	None	Modeller modeling using crosslinks	None	None	False	False

There are 6 software packages reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	Rosetta	Not available	model building	https://www.rosettacommons.org/
2	Modeller	Not available	homology modeling	https://salilab.org/modeller/
3	Gromacs	Not available	model building	http://www.gromacs.org
4	PROSESS	Not available	model evaluation	https://www.prosess.ca
5	Molprobitry	Not available	model evaluation	http://molprobitry.biochem.duke.edu
6	mCSM	Not available	data analysis	http://biosig.unimelb.edu.au/mcsm/

Data quality ?

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.

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.

Standard geometry: bond outliers ?

There are 68 bond length outliers in this entry (1.90% of 3570 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	273	HIS	CB-CG	8.30	1.61	1.50	1	1
A	165	HIS	CG-ND1	7.83	1.29	1.38	1	1
C	23	DG	O3'-P	6.57	1.51	1.61	1	1
B	25	DG	O3'-P	6.45	1.51	1.61	1	1
A	165	HIS	CE1-NE2	6.44	1.39	1.32	1	1
A	205	THR	CB-OG1	6.31	1.33	1.43	1	1
A	285	ARG	CD-NE	6.14	1.54	1.46	1	1
A	104	VAL	C-N	5.89	1.41	1.33	1	1
A	157	GLY	N-CA	5.81	1.54	1.45	1	1
A	120	PRO	C-N	5.71	1.41	1.33	1	1
B	10	DC	P-O5'	5.60	1.71	1.60	1	1
B	5	DC	C5'-C4'	5.54	1.63	1.52	1	1
A	220	ARG	CZ-NH2	5.44	1.26	1.33	1	1
C	12	DC	C5'-C4'	5.36	1.63	1.52	1	1
A	57	THR	C-N	5.28	1.40	1.33	1	1
A	220	ARG	CD-NE	5.22	1.53	1.46	1	1
A	134	TRP	NE1-CE2	5.14	1.43	1.37	1	1
A	97	ARG	CZ-NH2	5.11	1.26	1.33	1	1
A	241	ARG	CZ-NH1	5.07	1.25	1.32	1	1
A	162	ARG	CD-NE	5.02	1.53	1.46	1	1
C	21	DA	N7-C5	5.01	1.49	1.39	1	1
A	268	SER	CA-C	4.99	1.63	1.52	1	1
A	99	ASP	C-N	4.90	1.26	1.33	1	1
A	192	ASP	C-N	4.74	1.40	1.33	1	1
A	28	HIS	ND1-CE1	4.73	1.37	1.32	1	1
A	285	ARG	N-CA	4.65	1.55	1.46	1	1
A	97	ARG	CZ-NH1	4.58	1.26	1.32	1	1
A	51	ARG	CD-NE	4.57	1.52	1.46	1	1
A	1	MET	CA-CB	4.56	1.62	1.53	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	165	HIS	ND1-CE1	4.55	1.37	1.32	1	1
A	21	LYS	CA-CB	4.55	1.62	1.53	1	1
A	222	MET	N-CA	4.49	1.54	1.46	1	1
B	10	DC	O3'-P	4.48	1.54	1.61	1	1
A	241	ARG	CD-NE	4.46	1.52	1.46	1	1
B	2	DA	N3-C4	4.46	1.43	1.34	1	1
A	167	GLU	N-CA	4.43	1.54	1.46	1	1
A	120	PRO	N-CD	4.42	1.41	1.47	1	1
A	250	ARG	N-CA	4.40	1.54	1.46	1	1
A	13	LEU	C-N	4.40	1.39	1.33	1	1
A	273	HIS	C-N	4.40	1.39	1.33	1	1
A	203	HIS	CB-CG	4.40	1.56	1.50	1	1
A	15	LYS	C-N	4.37	1.39	1.33	1	1
A	107	GLN	N-CA	4.35	1.38	1.46	1	1
B	21	DC	C4-N4	4.34	1.25	1.34	1	1
C	18	DT	C4'-O4'	4.31	1.36	1.45	1	1
A	134	TRP	CZ2-CH2	4.31	1.45	1.37	1	1
A	186	ARG	CD-NE	4.28	1.52	1.46	1	1
C	11	DT	C4'-O4'	4.28	1.36	1.45	1	1
A	191	ILE	CA-C	4.26	1.61	1.52	1	1
A	125	PHE	C-N	4.22	1.39	1.33	1	1
A	230	ARG	CA-C	4.20	1.61	1.52	1	1
A	137	ASN	N-CA	4.15	1.54	1.46	1	1
A	259	GLY	N-CA	4.14	1.52	1.45	1	1
B	3	DT	C5-C6	4.14	1.42	1.34	1	1
A	150	ARG	NE-CZ	4.13	1.28	1.33	1	1
A	97	ARG	CD-NE	4.13	1.52	1.46	1	1
A	255	ARG	C-N	4.12	1.39	1.33	1	1
A	99	ASP	C-O	4.12	1.15	1.23	1	1
A	209	TRP	NE1-CE2	4.11	1.42	1.37	1	1
A	5	LYS	CA-CB	4.10	1.61	1.53	1	1
A	282	HIS	ND1-CE1	4.09	1.36	1.32	1	1
A	150	ARG	CD-NE	4.09	1.52	1.46	1	1
A	281	VAL	C-N	4.08	1.39	1.33	1	1
C	15	DG	C2-N2	4.06	1.26	1.34	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	24	ALA	CA-C	4.02	1.61	1.52	1	1
A	185	SER	N-CA	4.02	1.53	1.46	1	1
A	217	PRO	CA-C	4.01	1.61	1.52	1	1
A	232	PRO	CA-CB	4.01	1.61	1.53	1	1

Standard geometry: angle outliers ?

There are 295 bond angle outliers in this entry (5.85% of 5040 assessed bonds). A summary is provided below. The output is limited to 100 rows.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	42	ASP	CA-CB-CG	10.40	123.00	112.60	1	1
A	47	PRO	N-CA-CB	9.51	113.46	103.00	1	1
A	229	HIS	CB-CG-CD2	9.32	119.09	131.20	1	1
A	279	ASP	CA-CB-CG	9.31	121.91	112.60	1	1
A	165	HIS	ND1-CG-CD2	8.99	115.09	106.10	1	1
C	22	DG	C5'-C4'-C3'	8.72	101.82	114.90	1	1
C	4	DC	C5'-C4'-O4'	8.59	122.29	109.40	1	1
C	13	DA	O5'-C5'-C4'	8.01	122.82	110.80	1	1
A	82	HIS	CB-CG-CD2	7.82	121.04	131.20	1	1
A	59	ARG	NE-CZ-NH2	7.66	112.30	119.20	1	1
A	190	LEU	C-N-CA	7.62	135.42	121.70	1	1
C	23	DG	O4'-C1'-N9	7.55	119.73	108.40	1	1
A	175	PHE	CA-CB-CG	7.27	106.53	113.80	1	1
C	6	DG	O4'-C1'-N9	7.13	119.10	108.40	1	1
C	15	DG	C3'-O3'-P	7.13	130.90	120.20	1	1
A	212	TYR	C-N-CA	7.05	134.40	121.70	1	1
B	17	DG	C4'-C3'-O3'	6.97	99.54	110.00	1	1
C	7	DC	O2-C2-N3	6.95	111.48	121.90	1	1
A	174	ARG	NE-CZ-NH1	6.93	128.43	121.50	1	1
B	24	DA	C5'-C4'-O4'	6.90	119.75	109.40	1	1
B	7	DT	O5'-C5'-C4'	6.87	121.10	110.80	1	1
A	67	ASN	OD1-CG-ND2	6.76	115.84	122.60	1	1
B	8	DG	O4'-C1'-N9	6.73	118.49	108.40	1	1
A	220	ARG	NH1-CZ-NH2	6.69	110.60	119.30	1	1
B	19	DG	O3'-P-O5'	6.69	93.97	104.00	1	1
B	16	DA	N6-C6-N1	6.66	109.01	119.00	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
C	1	DA	C5-C6-N1	6.63	127.54	117.60	1	1
C	2	DC	N3-C4-N4	6.60	108.00	117.90	1	1
C	6	DG	O6-C6-N1	6.58	110.14	120.00	1	1
B	8	DG	O4'-C1'-C2'	6.56	96.57	106.40	1	1
B	3	DT	C4'-C3'-O3'	6.54	119.82	110.00	1	1
C	26	DC	C3'-C2'-C1'	6.49	111.33	101.60	1	1
A	32	GLN	OE1-CD-NE2	6.45	116.15	122.60	1	1
C	19	DC	C3'-C2'-C1'	6.44	111.26	101.60	1	1
A	16	GLN	OE1-CD-NE2	6.44	116.16	122.60	1	1
B	14	DT	O4'-C1'-N1	6.42	118.04	108.40	1	1
B	9	DA	O5'-C5'-C4'	6.41	120.42	110.80	1	1
B	21	DC	O4'-C1'-N1	6.35	117.93	108.40	1	1
B	1	DG	C4'-O4'-C1'	6.35	100.17	109.70	1	1
C	23	DG	N9-C8-N7	6.31	122.96	113.50	1	1
C	13	DA	C5-C6-N1	6.30	127.06	117.60	1	1
A	186	ARG	CD-NE-CZ	6.22	133.11	124.40	1	1
A	266	ARG	NE-CZ-NH1	6.22	127.72	121.50	1	1
A	226	PHE	CA-CB-CG	6.19	119.99	113.80	1	1
C	20	DA	O5'-C5'-C4'	6.16	120.04	110.80	1	1
C	1	DA	C4'-C3'-C2'	6.15	93.17	102.40	1	1
A	165	HIS	CB-CG-CD2	6.10	123.27	131.20	1	1
C	15	DG	O4'-C1'-C2'	6.06	97.31	106.40	1	1
A	232	PRO	N-CD-CG	6.05	112.27	103.20	1	1
C	10	DC	C3'-O3'-P	6.04	129.27	120.20	1	1
A	240	PHE	CA-CB-CG	6.04	107.76	113.80	1	1
A	41	PHE	N-CA-CB	6.04	120.77	110.50	1	1
A	205	THR	CA-CB-OG1	6.03	118.64	109.60	1	1
C	1	DA	O4'-C1'-C2'	6.02	97.38	106.40	1	1
A	34	LYS	O-C-N	6.01	113.39	123.00	1	1
A	144	ALA	C-CA-CB	5.99	119.49	110.50	1	1
C	16	DT	O3'-C3'-C2'	5.97	102.54	111.50	1	1
A	182	LYS	CA-C-N	5.96	128.11	116.20	1	1
A	203	HIS	CB-CG-CD2	5.95	123.46	131.20	1	1
A	184	THR	C-N-CA	5.89	132.30	121.70	1	1
A	150	ARG	NE-CZ-NH1	5.88	127.38	121.50	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
C	26	DC	C4-C5-C6	5.85	108.82	117.60	1	1
A	52	ALA	C-N-CA	5.82	132.17	121.70	1	1
C	6	DG	O4'-C1'-C2'	5.80	97.70	106.40	1	1
B	15	DG	N2-C2-N3	5.79	111.01	119.70	1	1
C	1	DA	N6-C6-N1	5.78	110.33	119.00	1	1
A	273	HIS	CB-CG-CD2	5.77	123.70	131.20	1	1
A	91	TYR	CG-CD2-CE2	5.70	112.66	121.20	1	1
A	229	HIS	CB-CG-ND1	5.67	131.21	122.70	1	1
A	15	LYS	C-CA-CB	5.67	120.88	110.10	1	1
A	247	ASP	CA-CB-CG	5.64	118.24	112.60	1	1
C	8	DC	O4'-C1'-N1	5.60	116.80	108.40	1	1
C	1	DA	C6-C5-C4	5.58	108.53	116.90	1	1
A	71	PRO	N-CA-CB	5.57	109.12	103.00	1	1
A	6	ARG	CD-NE-CZ	5.55	132.18	124.40	1	1
A	114	ARG	NE-CZ-NH1	5.55	115.95	121.50	1	1
B	8	DG	N1-C2-N2	5.55	124.62	116.30	1	1
C	1	DA	O3'-P-O5'	5.54	95.68	104.00	1	1
B	26	DT	C7-C5-C6	5.54	115.69	124.00	1	1
B	15	DG	N1-C2-N2	5.53	124.60	116.30	1	1
C	23	DG	C3'-C2'-C1'	5.53	93.31	101.60	1	1
A	55	ASN	C-CA-CB	5.52	120.58	110.10	1	1
C	9	DC	P-O5'-C5'	5.51	128.27	120.00	1	1
A	66	ARG	NE-CZ-NH1	5.45	116.05	121.50	1	1
A	192	ASP	CA-CB-CG	5.45	118.05	112.60	1	1
A	177	VAL	CA-C-N	5.45	127.10	116.20	1	1
A	220	ARG	NE-CZ-NH1	5.36	126.86	121.50	1	1
B	11	DA	C2'-C1'-N9	5.35	121.52	113.50	1	1
C	26	DC	O5'-C5'-C4'	5.34	118.81	110.80	1	1
C	5	DT	C5'-C4'-C3'	5.31	106.93	114.90	1	1
B	22	DA	O3'-P-O5'	5.29	111.93	104.00	1	1
A	34	LYS	C-N-CA	5.28	131.21	121.70	1	1
A	221	ARG	NE-CZ-NH2	5.27	114.45	119.20	1	1
A	277	ASN	CA-CB-CG	5.27	107.33	112.60	1	1
B	21	DC	O2-C2-N3	5.25	114.03	121.90	1	1
A	225	TYR	N-CA-CB	5.24	101.59	110.50	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	255	ARG	CD-NE-CZ	5.23	131.72	124.40	1	1
B	21	DC	O3'-C3'-C2'	5.22	103.67	111.50	1	1
B	15	DG	O5'-C5'-C4'	5.21	118.61	110.80	1	1
A	72	ARG	CD-NE-CZ	5.19	131.67	124.40	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	0.95	6

There are 6 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)
A:171:LEU:C	A:171:LEU:HD13	0.53	1	1
A:204:TYR:CE1	A:208:ILE:HG21	0.47	1	1
B:21:DC:H2"	B:22:DA:C8	0.46	1	1
B:16:DA:H1'	B:17:DG:C8	0.42	1	1
A:4:LYS:HE2	A:134:TRP:CD2	0.41	1	1
A:235:LEU:H	A:235:LEU:HD12	0.41	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	283	253	25	5

There are 5 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	6	ARG	1
A	24	ALA	1
A	117	ILE	1
A	135	SER	1
A	279	ASP	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	253	230	18	5

There are 5 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	171	LEU	1
A	208	ILE	1
A	224	ASP	1
A	258	VAL	1
A	278	ASP	1

Fit of model to data used for modeling ?

Fit of model(s) to crosslinking-MS data

Restraint types

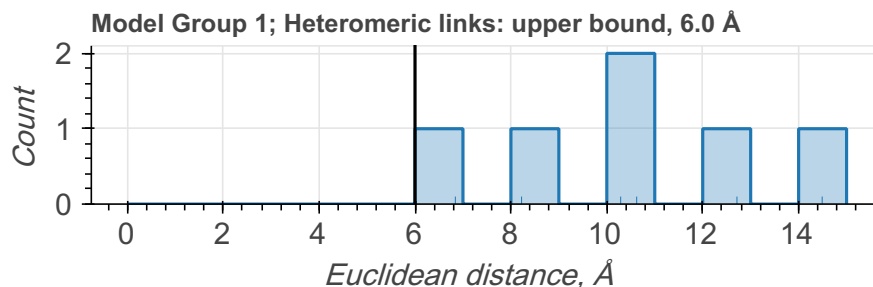
Restraint types are summarized in the table below. Restraints assigned "*by-residue*" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "*coarse-grained*". *Restraint group* represents a set of crosslinking restraints applied collectively in the modeling.

There are 6 crosslinking restraints combined in 6 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
BrdU	ASN	OD1	DT	C5	upper bound	6.0	1
BrdU	DT	C5	LYS	NZ	upper bound	6.0	2
BrdU	DT	C5	ILE	CD1	upper bound	6.0	1
BrdU	DT	C5	PRO	CD	upper bound	6.0	1
BrdU	ARG	NH2	DT	C5	upper bound	6.0	1

Distograms of individual restraints

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.



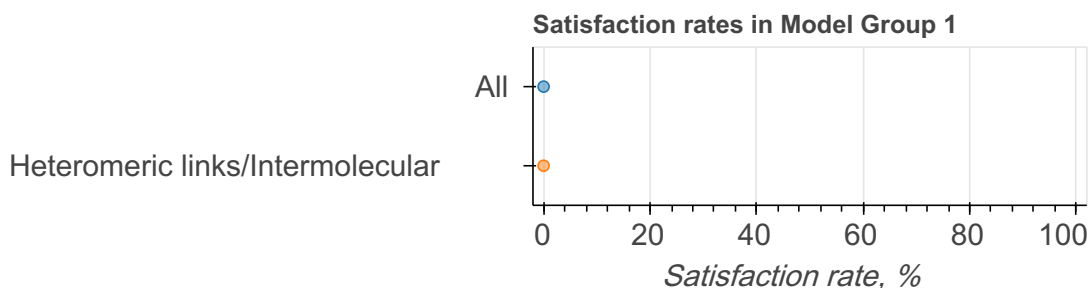
Satisfaction of restraints

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=6)
1	1	1	1/1	All	0.00	100.00	6
				Heteromeric links/Intermolecular	0.00	100.00	6

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.



Fit of model to data used for validation ?

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.