Integrative Structure Validation Report

February 18, 2025 - 08:45 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	9A8C
PDB-Dev ID	PDBDEV_00000377
Structure Title	Escherichia coli sigma70 in apo form
Structure Authors	Joron, K.; Zamel, J.; Kalisman, N.; Lerner, E.
Deposited on	2024-03-12

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

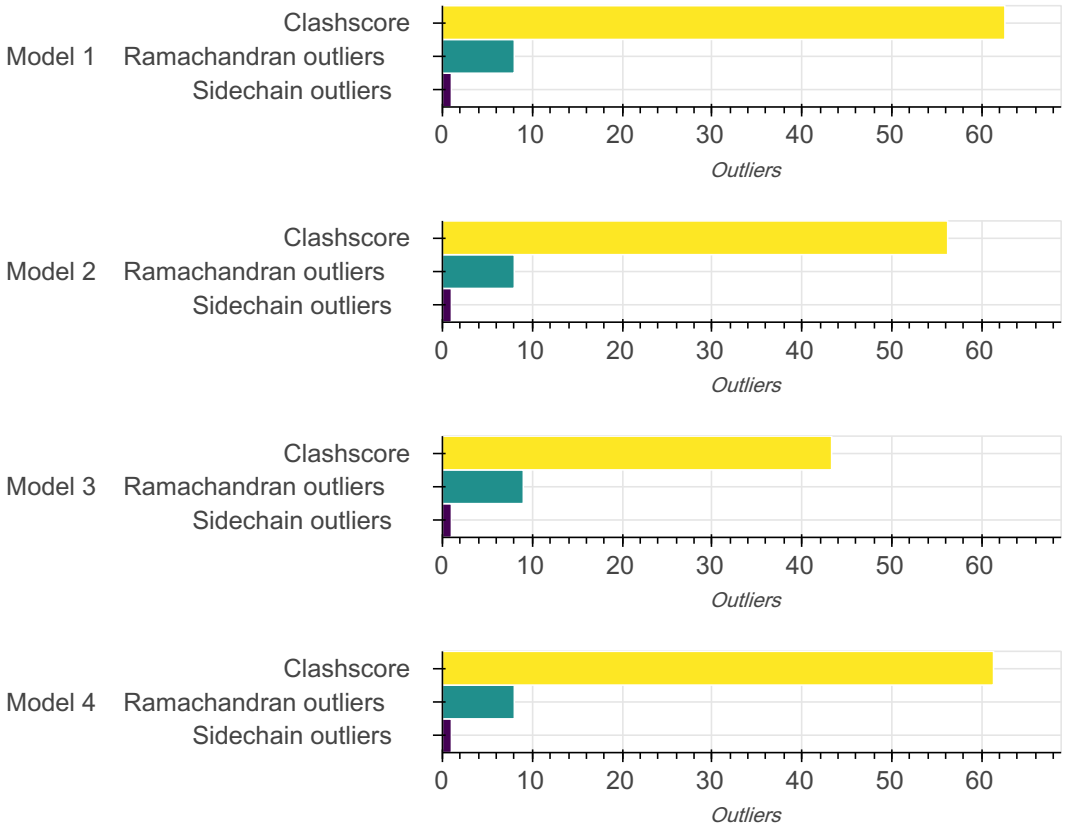
A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the ? symbol.

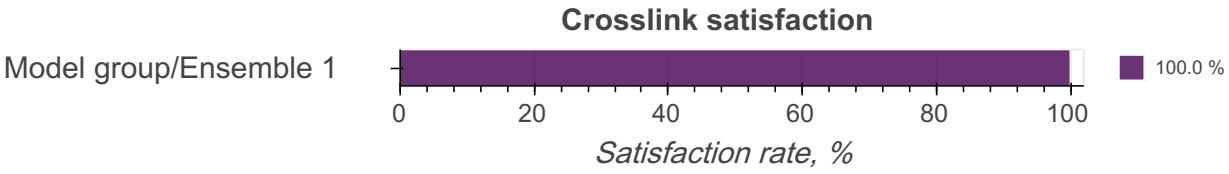
List of references used to build this report is available here.

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 4 model(s). A total of 2 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	1-4	1	RNA polymerase sigma factor RpoD	A	536	-	1-536	100.00 / 100.00	Atomic

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	PRIDE	PXD037183
2	Experimental model	PDB	6P1K

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	Docking	None	None	None	False	False
2	1	Docking	None	None	None	False	False
3	1	Docking	None	None	None	False	False

There are 2 software packages reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	PatchDock	1.30	Docking	https://bioinfo3d.cs.tau.ac.il/PatchDock/patchdock.html
2	ChimeraX	1.2.5	Structure visualization and analysis	https://www.cgl.ucsf.edu/chimerax/

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 10 bond length outliers in this entry (0.06% of 17498 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	453	LEU	C-N	68.68	2.44	1.34	1	1
A	369	GLN	C-N	38.91	0.78	1.33	4	4
A	59	GLU	C-N	28.98	1.73	1.33	3	1
A	483	ARG	CD-NE	4.07	1.52	1.46	2	4

Standard geometry: angle outliers

There are 103 bond angle outliers in this entry (0.44% of 23555 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	59	GLU	CA-C-N	51.27	13.66	116.20	3	1
A	369	GLN	CA-C-N	45.39	25.42	116.20	4	4
A	453	LEU	CA-C-N	42.95	52.47	116.90	1	1
A	453	LEU	O-C-N	33.80	68.92	123.00	1	1
A	59	GLU	C-N-CA	32.84	62.59	121.70	3	1
A	453	LEU	C-N-CD	25.29	21.30	125.00	1	1
A	369	GLN	C-N-CA	23.12	80.08	121.70	4	4
A	369	GLN	O-C-N	9.64	107.57	123.00	4	4
A	59	GLU	O-C-N	7.80	110.52	123.00	3	1
A	19	ASP	CA-CB-CG	6.67	119.27	112.60	4	4
A	536	ASP	CA-CB-CG	6.14	118.74	112.60	3	4
A	104	VAL	N-CA-CB	5.43	102.27	111.50	4	4
A	402	THR	CA-C-N	5.38	124.96	116.90	4	4
A	116	ASP	CA-CB-CG	5.35	117.95	112.60	4	4
A	23	MET	CG-SD-CE	5.34	89.15	100.90	2	4
A	468	HIS	ND1-CG-CD2	4.80	110.90	106.10	2	4

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	493	ASP	CA-CB-CG	4.77	117.37	112.60	1	4
A	453	LEU	C-N-CA	4.67	99.24	122.60	1	1
A	185	VAL	CA-C-N	4.61	123.81	116.90	4	4
A	261	HIS	ND1-CG-CD2	4.48	110.58	106.10	4	4
A	488	ILE	C-N-CA	4.47	129.74	121.70	2	4
A	103	HIS	ND1-CG-CD2	4.45	110.55	106.10	4	4
A	535	ASP	CA-CB-CG	4.45	117.05	112.60	4	4
A	224	ASN	CA-CB-CG	4.42	108.18	112.60	4	4
A	378	HIS	ND1-CG-CD2	4.29	110.39	106.10	4	4
A	220	MET	CA-C-N	4.28	123.32	116.90	4	4
A	254	HIS	ND1-CG-CD2	4.26	110.36	106.10	4	4
A	104	VAL	CA-CB-CG1	4.21	117.56	110.40	4	4
A	524	PRO	C-N-CA	4.16	129.19	121.70	3	4
A	247	LYS	CA-C-N	4.12	123.07	116.90	4	4
A	165	HIS	ND1-CG-CD2	4.04	110.14	106.10	4	4

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	62.59	538
2	56.22	483
3	43.31	372
4	61.32	527

There are 1920 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:46:ILE:CG2	A:183:ARG:HA	1.66	4	1
A:57:VAL:HG22	A:179:PHE:CD2	1.66	2	1
A:46:ILE:HG21	A:183:ARG:CA	1.65	4	1
A:208:ARG:CD	A:501:LYS:HG2	1.64	2	2
A:32:GLU:HG3	A:353:TYR:CE1	1.63	4	1
A:423:ILE:HA	A:526:ARG:CD	1.62	1	1
A:208:ARG:HD3	A:501:LYS:CG	1.61	2	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:208:ARG:HH21	A:501:LYS:CD	1.60	1	1
A:423:ILE:HA	A:526:ARG:CG	1.60	1	1
A:57:VAL:HG22	A:179:PHE:CE2	1.60	2	2
A:56:SER:CB	A:196:MET:HE1	1.60	1	1
A:57:VAL:CG1	A:179:PHE:HB3	1.58	3	2
A:373:ILE:HB	A:524:PRO:CG	1.57	3	1
A:24:TYR:HB3	A:388:ARG:CZ	1.57	4	1
A:187:LYS:CE	A:412:MET:HG2	1.57	4	4
A:200:MET:CE	A:504:ASP:HB2	1.56	2	2
A:56:SER:HA	A:196:MET:CE	1.53	1	2
A:27:GLU:CB	A:359:ARG:NH2	1.52	3	1
A:50:ILE:HD13	A:179:PHE:CD2	1.52	4	1
A:56:SER:CA	A:196:MET:HE1	1.50	1	2
A:57:VAL:CG1	A:179:PHE:CB	1.50	3	2
A:31:VAL:CG2	A:353:TYR:N	1.49	1	1
A:30:THR:CG2	A:355:THR:HG21	1.49	1	1
A:451:LEU:CD1	A:455:LEU:HD23	1.48	1	1
A:423:ILE:CD1	A:526:ARG:HG2	1.48	3	1
A:423:ILE:HD11	A:526:ARG:CG	1.48	3	1
A:378:HIS:CE1	A:516:LYS:CE	1.47	1	1
A:24:TYR:CB	A:388:ARG:CZ	1.47	4	1
A:296:ARG:HH22	A:507:ARG:CG	1.47	1	1
A:416:LYS:CE	A:477:ARG:NH2	1.47	1	1
A:296:ARG:NH1	A:512:GLN:NE2	1.47	3	1
A:378:HIS:CE1	A:516:LYS:HE3	1.46	1	1
A:193:VAL:CG1	A:477:ARG:NH1	1.46	2	1
A:31:VAL:CG2	A:352:THR:HB	1.46	2	1
A:41:ASP:OD1	A:345:ARG:CD	1.45	4	1
A:300:LYS:CE	A:508:GLU:CG	1.45	1	1
A:57:VAL:CG2	A:179:PHE:CD2	1.44	2	2
A:41:ASP:OD2	A:345:ARG:CG	1.43	4	1
A:422:LYS:CG	A:526:ARG:HH11	1.43	1	1
A:185:VAL:HG21	A:411:LEU:CB	1.42	4	4
A:57:VAL:CG2	A:179:PHE:CE2	1.42	1	2
A:423:ILE:CA	A:526:ARG:HD2	1.42	1	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:51:ASN:HD21	A:180:LYS:CG	1.42	4	1
A:200:MET:HE2	A:504:ASP:CB	1.42	3	2
A:300:LYS:CE	A:508:GLU:HG3	1.41	1	1
A:208:ARG:HE	A:501:LYS:C	1.41	2	1
A:44:LYS:HD2	A:294:LYS:CE	1.41	2	1
A:41:ASP:CG	A:345:ARG:CD	1.41	4	1
A:36:ARG:C	A:93:ALA:HA	1.41	4	1
A:300:LYS:HE2	A:508:GLU:CG	1.41	1	1
A:27:GLU:HB2	A:359:ARG:NH2	1.41	3	1
A:208:ARG:NH2	A:501:LYS:CD	1.40	1	1
A:36:ARG:O	A:93:ALA:CA	1.40	4	1
A:24:TYR:HB3	A:388:ARG:NE	1.40	4	1
A:32:GLU:HG3	A:353:TYR:CZ	1.40	4	1
A:185:VAL:CG2	A:411:LEU:HD22	1.40	4	4
A:59:GLU:CG	A:280:GLN:O	1.39	4	1
A:51:ASN:ND2	A:180:LYS:HG2	1.39	4	1
A:200:MET:CE	A:504:ASP:CB	1.39	2	2
A:378:HIS:HE1	A:496:LEU:CD1	1.39	4	1
A:451:LEU:HD13	A:455:LEU:CD2	1.39	1	1
A:36:ARG:HG3	A:93:ALA:C	1.39	4	1
A:33:LEU:HG	A:349:LYS:NZ	1.38	2	2
A:30:THR:N	A:352:THR:CG2	1.38	1	2
A:34:LEU:HD12	A:349:LYS:NZ	1.37	3	1
A:32:GLU:CB	A:351:SER:OG	1.37	1	1
A:56:SER:O	A:63:ALA:HB2	1.37	1	2
A:185:VAL:HG22	A:411:LEU:CD2	1.37	4	4
A:42:ILE:HD13	A:342:PHE:CE2	1.37	3	2
A:31:VAL:CG2	A:352:THR:C	1.37	1	1
A:31:VAL:HG22	A:353:TYR:CA	1.36	1	1
A:36:ARG:CG	A:93:ALA:C	1.36	4	1
A:56:SER:CB	A:196:MET:CE	1.36	1	1
A:204:ARG:CD	A:501:LYS:HG2	1.36	1	1
A:27:GLU:CG	A:359:ARG:NH2	1.36	3	1
A:57:VAL:HG13	A:179:PHE:CG	1.35	3	2
A:57:VAL:CG2	A:179:PHE:HD2	1.35	2	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
A:372:THR:OG1	A:524:PRO:CB	1.35	1	1
A:56:SER:CA	A:196:MET:CE	1.35	1	1
A:194:ASN:HB3	A:516:LYS:NZ	1.35	4	1
A:187:LYS:HE2	A:412:MET:CG	1.34	4	4
A:300:LYS:NZ	A:508:GLU:HG3	1.34	1	1
A:42:ILE:HG21	A:342:PHE:CE2	1.34	3	1
A:30:THR:HB	A:355:THR:CB	1.34	1	1
A:45:ARG:NH2	A:339:VAL:HG13	1.34	3	3
A:45:ARG:NH2	A:339:VAL:CG1	1.34	3	3
A:293:ALA:CB	A:512:GLN:HE22	1.34	2	1
A:361:ALA:CB	A:377:VAL:HG11	1.34	4	4
A:422:LYS:CE	A:473:GLY:HA3	1.33	1	1
A:423:ILE:N	A:526:ARG:HD2	1.33	1	1
A:416:LYS:NZ	A:477:ARG:HH22	1.33	1	1
A:296:ARG:NH2	A:507:ARG:HG2	1.33	1	1
A:296:ARG:CZ	A:512:GLN:HE22	1.33	3	1
A:200:MET:HE1	A:504:ASP:CB	1.32	2	1
A:57:VAL:C	A:61:PRO:HD3	1.32	4	1
A:33:LEU:CA	A:349:LYS:HD3	1.32	3	1
A:208:ARG:NE	A:501:LYS:CB	1.32	2	2
A:28:MET:O	A:352:THR:HG21	1.31	1	1
A:293:ALA:HB2	A:512:GLN:NE2	1.31	2	1
A:53:VAL:O	A:60:TYR:CB	1.30	4	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	530	480	42	8
2	530	480	42	8
3	532	480	43	9
4	530	480	42	8

There are 9 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	19	ASP	4

Chain	Res	Type	Models (Total)
A	77	GLU	4
A	129	ALA	4
A	135	ILE	4
A	136	ASP	4
A	439	ASP	4
A	489	ASP	4
A	491	ASN	4
A	60	TYR	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	473	455	17	1
2	473	455	17	1
3	473	455	17	1
4	473	455	17	1

There are 1 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	422	LYS	4

Fit of model to data used for modeling

Fit of model(s) to crosslinking-MS data

Restraint types

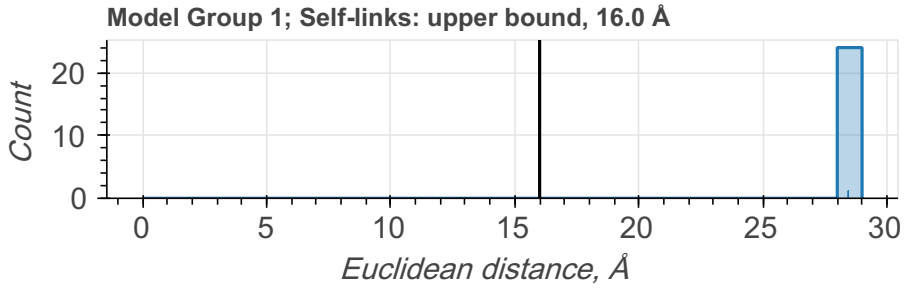
Restraint types are summarized in the table below. Restraints assigned "by-residue" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "coarse-grained". Restraint group represents a set of crosslinking restraints applied collectively in the modeling.

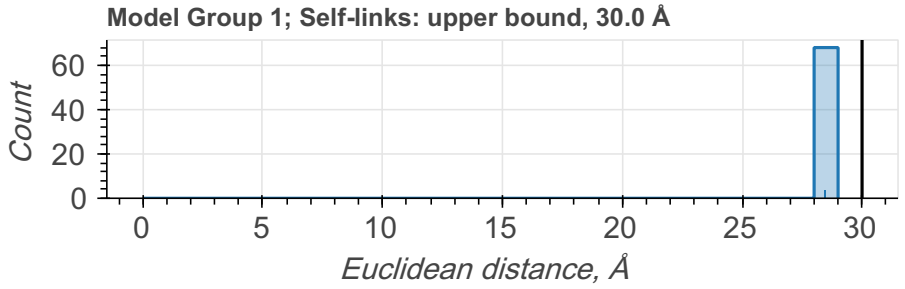
There are 92 crosslinking restraints combined in 1 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
BS3	LYS	CA	LYS	CA	upper bound	30.0	68
BS3	LYS	CA	LYS	CA	upper bound	16.0	24

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.





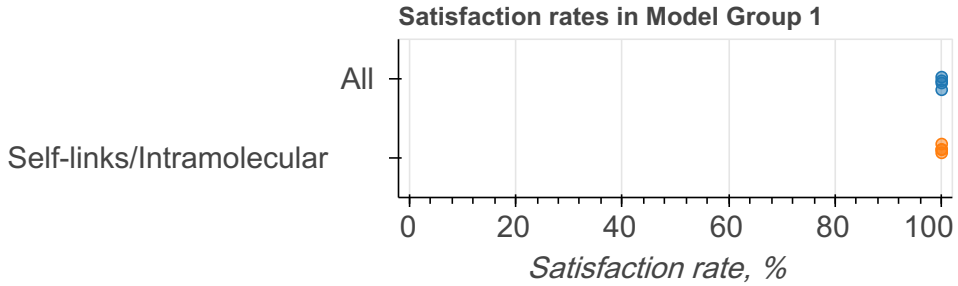
Satisfaction of restraints

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=1)
1	1	1	4/4	All	100.00	0.00	1
				Self-links/ Intramolecular	100.00	0.00	1

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 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.