

Integrative Structure Validation Report ?

March 27, 2025 - 09:53 AM PDT

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

ATSAS Version 3.2.1 (r14885)

PDB ID	8ZZ4
PDB-Dev ID	PDBDEV_00000004
Structure Title	Structure of K63-linked Diubiquitin
Structure Authors	Liu Z; Gong Z; Cao Y; Ding YH; Dong MQ; Lu YB; Zhang WP; Tang C
Deposited on	2017-09-08

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

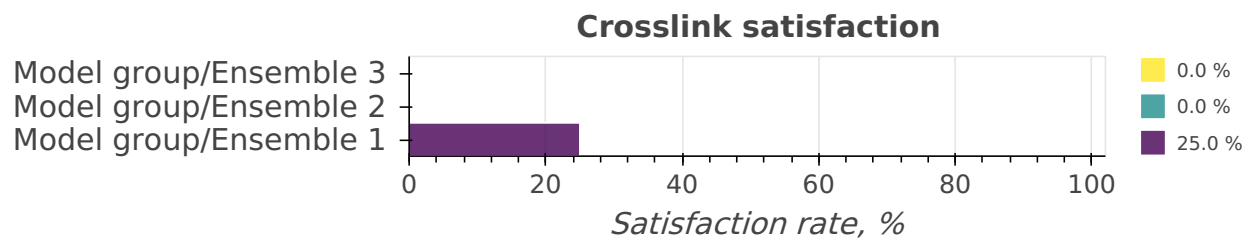
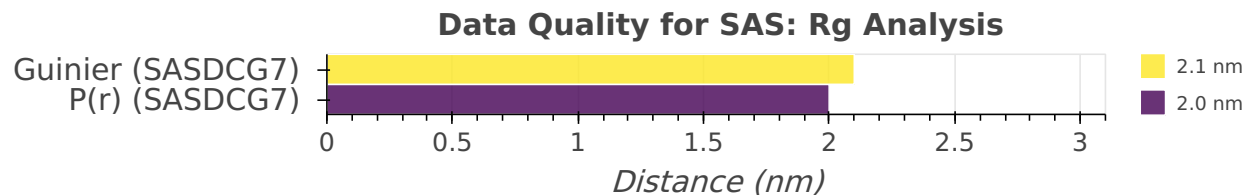
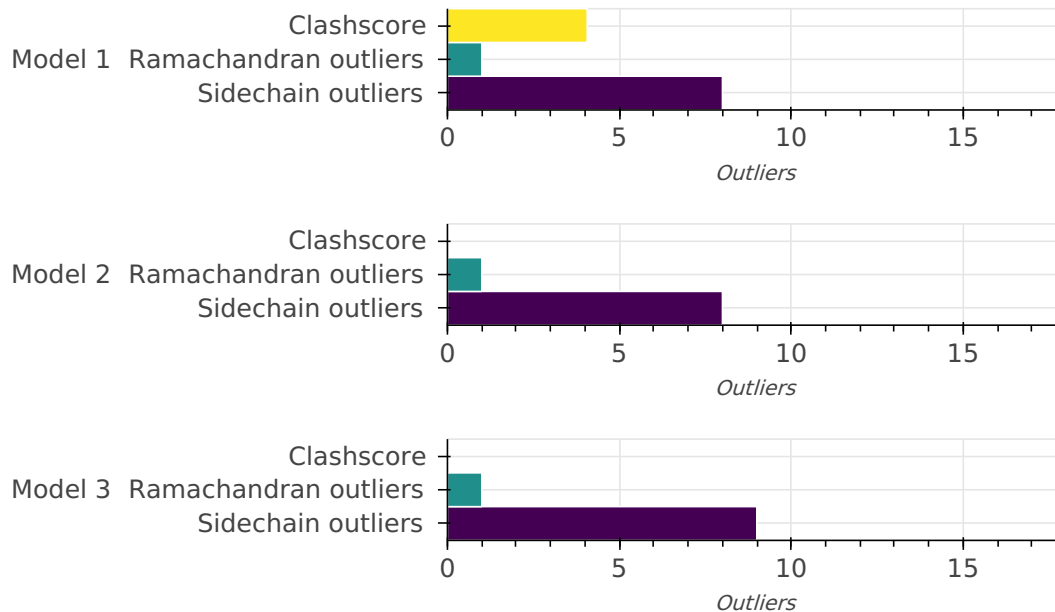
A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the ? symbol.

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 3 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	1-3	1	Ubiquitin	A	76	-	1-76	100.00 / 100.00	Atomic
				B					

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	SAS data	SASBDB	SASDCG7
2	Experimental model	PDB	1UBQ
3	Experimental model	PDB	2N2K
4	Crosslinking-MS data	Zenodo	10.5281/zenodo.1006721
5	Single molecule FRET data	Zenodo	10.5281/zenodo.1006721

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	None	None	None	True	False

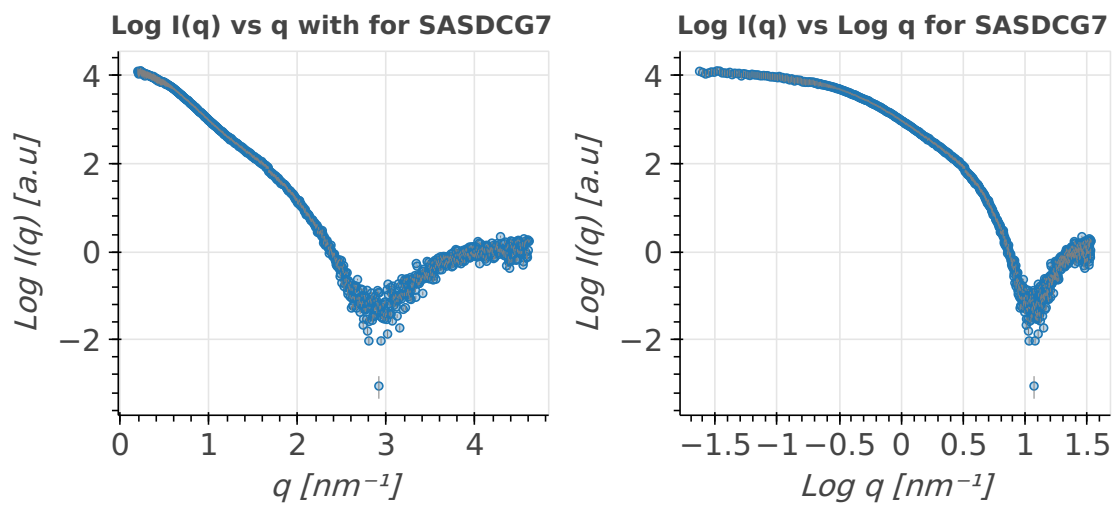
Software packages used for modeling were either not reported or not used.

Data quality ?

Scattering profile ?

SAS data used in this integrative model was obtained from 1 deposited SASBDB entry (entries).

Scattering profile for [SASDCG7](#): data from solutions of biological macromolecules are presented as both log I(q) vs q and log I(q) vs log (q) based on [SAS validation task force \(SASvtf\) recommendations](#). I(q) is the intensity (in arbitrary units) and q is the modulus of the scattering vector.



Key experimental estimates ?

Molecular weight (MW) estimates from experiments and analysis: true molecular weight can be compared to the Porod estimate from scattering profiles.

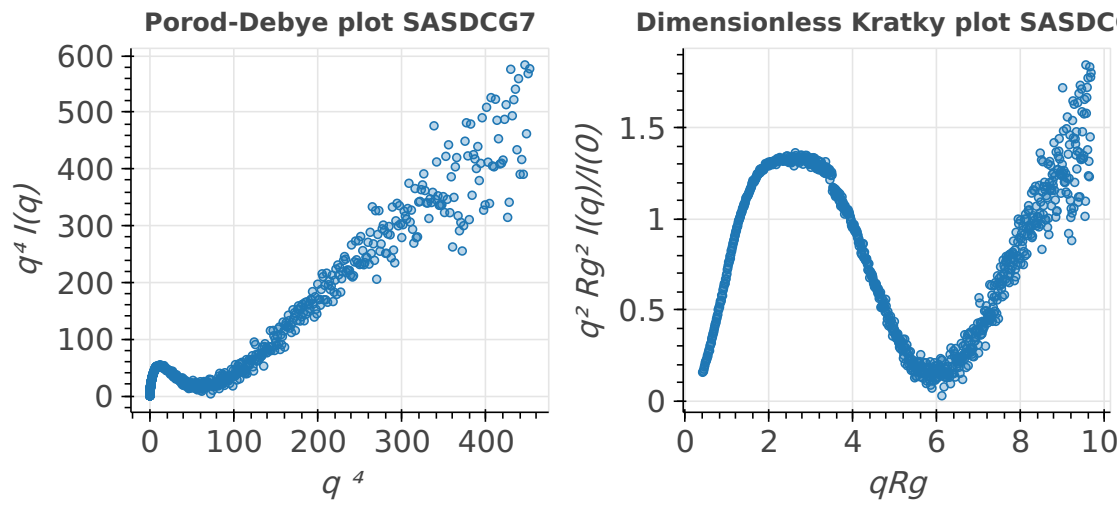
SASDB ID	Chemical composition MW	Standard MW	Porod Volume/MW
SASDCG7	13.0 kDa	Not available	1.69 nm ³ /kDa

Volume estimates from experiments and analysis: estimated volume can be compared to Porod volume obtained from scattering profiles.

SASDB ID	Estimated Volume	Porod Volume	Specific Volume	Sample Contrast	Sample Concentration
SASDCG7	Not available	22.00 nm ³	Not available	Not available	2.60 mg/mL

Flexibility analysis ?

Flexibility analysis for SASDCG7: In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas, fully unfolded domains are devoid of any discernable plateau. For details, refer to Figure 5 in Rambo and Tainer, 2011. In a Kratky plot, a parabolic shape is observed for globular (partial or fully folded) domains and a hyperbolic shape is observed for fully unfolded domains.

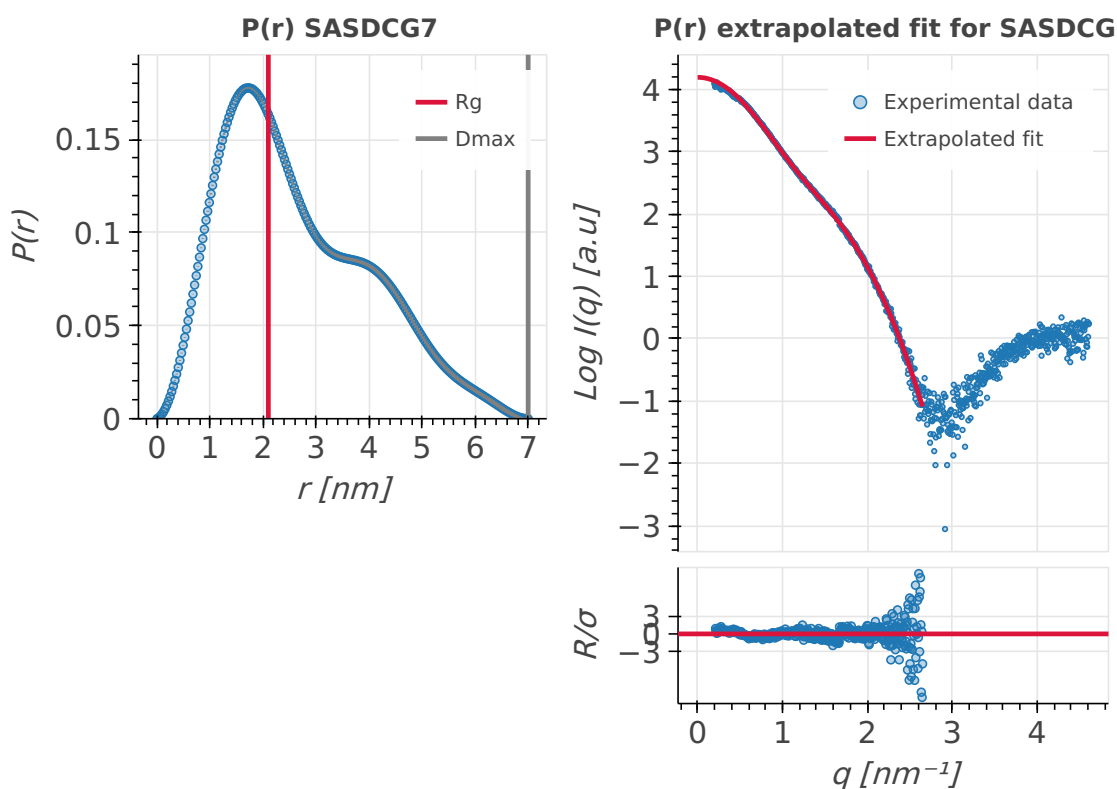


Pair-distance distribution analysis ?

P(r) analysis: $P(r)$ represents the distribution of distances between all pairs of atoms within the particle weighted by the respective electron densities. $P(r)$ is the Fourier transform of $I(s)$ (and vice versa). R_g can be estimated from integrating the $P(r)$ function. Agreement between the $P(r)$ and Guinier-determined R_g (table below) is a good measure of the self-consistency of the SAS profile. R_g is a measure for the overall size of a macromolecule; e.g. a protein with a smaller R_g is more compact than a protein with a larger R_g , provided both have the same molecular weight (MW). The point where $P(r)$ is decaying to zero is called D_{\max} and represents the maximum size of the particle.

SASDB ID	Software used	D_{\max}	D_{\max} error	R_g	R_g error
SASDCG7	GNOM 5.0	7.000 nm	Not available	2.100 nm	0.010 nm

P(r) for SASDCG7: The value of $P(r)$ should be zero beyond $r=D_{\max}$.



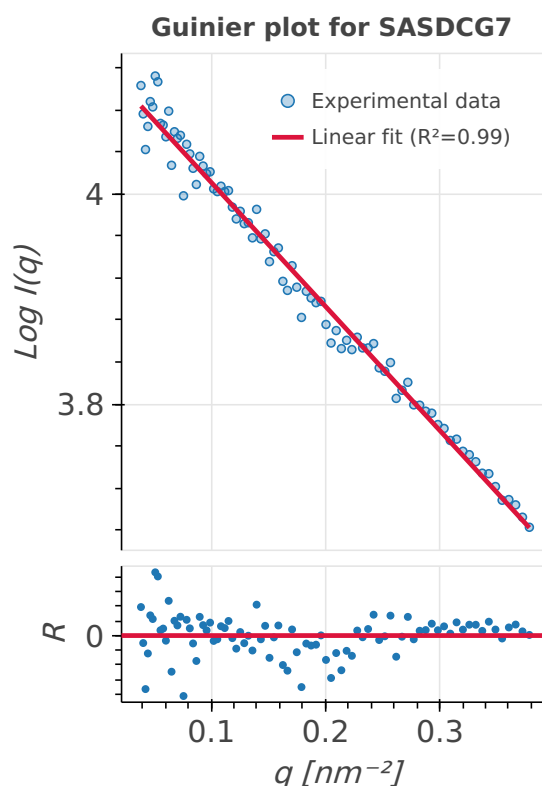
Guinier analysis ?

Guinier analysis: agreement between the $P(r)$ and Guinier-determined R_g (table below) is a good measure of the self-consistency of the SAS profile. Molecular weight estimates can also be compared to Porod and sample molecular weights for consistency.

SASDB ID	R_g	R_g error	MW	MW error
SASDCG7	2.00 nm	0.15 nm	Not available	Not available

Guinier analysis: the linearity of the Guinier plot is a sensitive indicator of the quality of the experimental SAS data; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the same size. Deviations from linearity usually point to strong interference effects, polydispersity of the samples or improper background subtraction. Residual value plot and

coefficient of determination (R^2) are measures to assess linear fit to the data. A perfect fit has an R^2 value of 1. Residual values should be equally and randomly spaced around the horizontal axis.



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.

Single molecule FRET

Validation for this section is under development.

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 no bond length outliers.

Standard geometry: angle outliers ?

There are 104 bond angle outliers in this entry (2.12% of 4902 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	54	ARG	CD-NE-CZ	8.30	136.01	124.40	3	3
B	54	ARG	CD-NE-CZ	8.25	135.95	124.40	3	3
B	14	THR	CA-CB-OG1	7.33	98.60	109.60	3	3
A	14	THR	CA-CB-OG1	7.31	98.64	109.60	1	3
B	32	ASP	CA-CB-CG	7.29	105.31	112.60	3	3
A	32	ASP	CA-CB-CG	7.24	105.36	112.60	3	3
A	68	HIS	CA-CB-CG	6.40	107.40	113.80	1	3
B	68	HIS	CA-CB-CG	6.39	107.41	113.80	1	3
B	72	ARG	CD-NE-CZ	5.77	116.33	124.40	2	3
A	72	ARG	CD-NE-CZ	5.73	116.38	124.40	3	3
A	9	THR	CA-CB-OG1	5.55	101.28	109.60	1	3
B	9	THR	CA-CB-OG1	5.50	101.35	109.60	2	3
A	55	THR	CA-CB-OG1	4.97	102.15	109.60	2	3
B	55	THR	CA-CB-OG1	4.96	102.16	109.60	1	3
A	7	THR	CA-CB-OG1	4.92	102.22	109.60	1	3
B	7	THR	CA-CB-OG1	4.92	102.22	109.60	3	3
B	18	GLU	CG-CD-OE1	4.88	129.63	118.40	2	3
A	18	GLU	CG-CD-OE1	4.87	129.60	118.40	2	3
B	51	GLU	CG-CD-OE1	4.74	107.49	118.40	1	3
A	51	GLU	CG-CD-OE1	4.73	107.53	118.40	2	3
B	3	ILE	O-C-N	4.71	130.54	123.00	1	3
A	3	ILE	O-C-N	4.67	130.47	123.00	3	3
B	58	ASP	CA-CB-CG	4.43	117.03	112.60	3	3
B	5	VAL	CA-C-O	4.41	113.30	120.80	2	3
B	72	ARG	CA-C-O	4.40	113.32	120.80	1	3
A	5	VAL	CA-C-O	4.40	113.33	120.80	3	3
A	72	ARG	CA-C-O	4.36	113.38	120.80	3	3
A	58	ASP	CA-CB-CG	4.36	116.96	112.60	3	3
B	54	ARG	NE-CZ-NH1	4.35	125.85	121.50	2	3
A	54	ARG	NE-CZ-NH1	4.27	125.77	121.50	1	3
B	25	ASN	CA-CB-CG	4.26	108.34	112.60	3	3
A	25	ASN	CA-CB-CG	4.25	108.35	112.60	3	3

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	33	LYS	CA-CB-CG	4.05	122.20	114.10	1	3
B	5	VAL	O-C-N	4.05	129.48	123.00	3	2
A	33	LYS	CA-CB-CG	4.02	122.15	114.10	1	3

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	4.07	10
2	0.00	0
3	0.00	0

There are 10 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:42:ARG:HE	A:49:GLN:NE2	0.60	1	1
B:42:ARG:HE	B:49:GLN:NE2	0.59	1	1
A:42:ARG:HE	A:49:GLN:HE21	0.53	1	1
B:42:ARG:HE	B:49:GLN:HE21	0.52	1	1
A:23:ILE:HB	A:52:ASP:HA	0.46	1	1
B:23:ILE:HB	B:52:ASP:HA	0.46	1	1
B:62:GLN:HB3	B:62:GLN:HE21	0.43	1	1
B:71:LEU:HA	B:71:LEU:HD12	0.41	1	1
B:72:ARG:O	B:73:LEU:O	0.40	1	1
B:26:VAL:HG21	B:56:LEU:HD21	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	148	144	3	1
2	148	146	1	1
3	148	145	2	1

There are 2 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
B	73	LEU	2
B	75	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	136	116	12	8
2	136	116	12	8
3	136	114	13	9

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

Chain	Res	Type	Models (Total)
A	13	ILE	3
A	15	LEU	3
A	39	ASP	3
A	71	LEU	3
B	13	ILE	3
B	15	LEU	3
B	39	ASP	3
B	71	LEU	3
A	74	ARG	1

Fit of model to data used for modeling ?

Fit of model(s) to SAS data

χ^2 goodness of fit and cormap analysis ?

Model(s) and/or fit for this entry have not been deposited.

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".

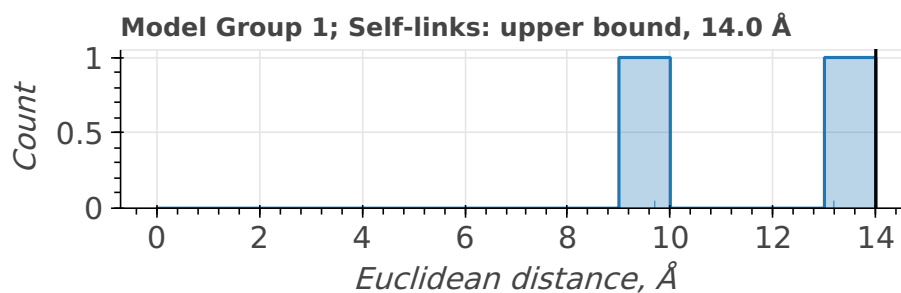
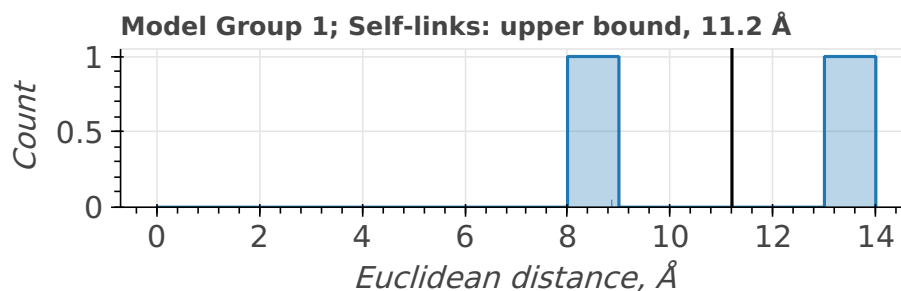
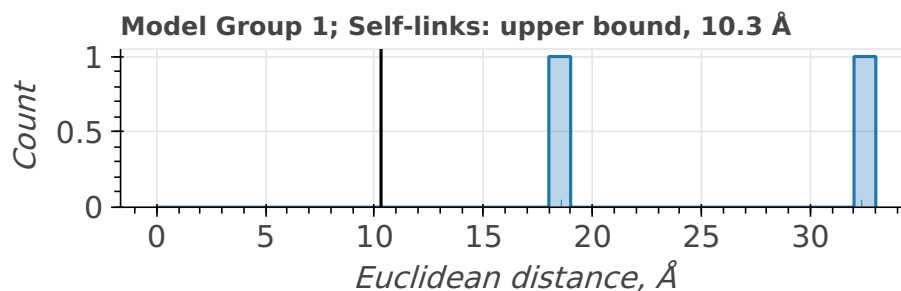
Restraint group represents a set of crosslinking restraints applied collectively in the modeling.

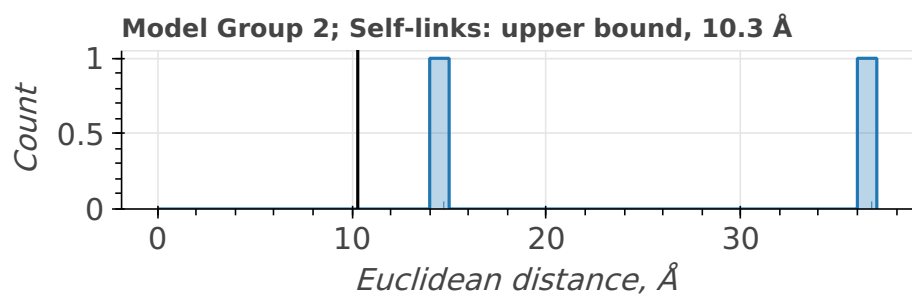
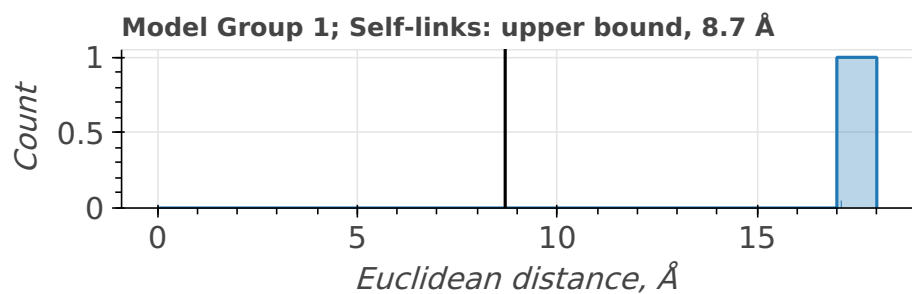
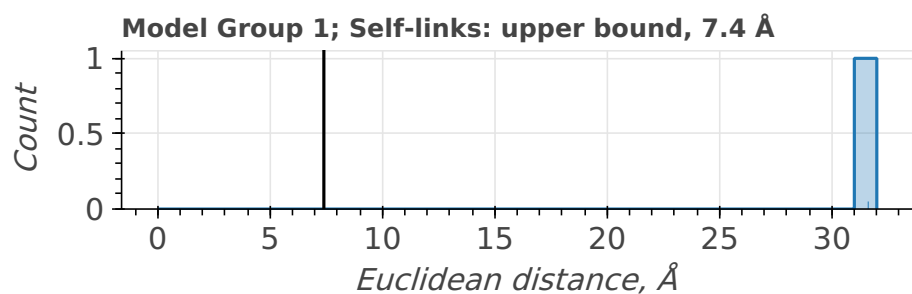
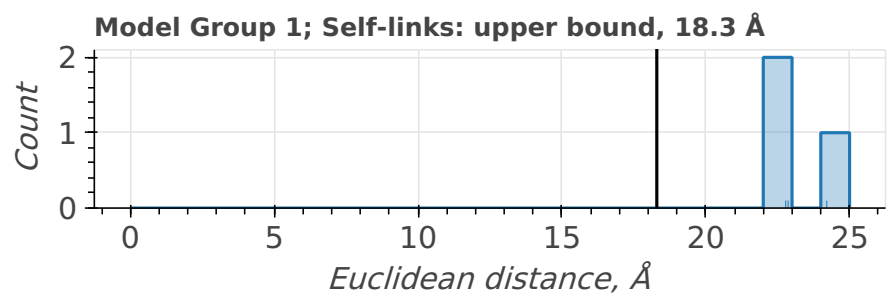
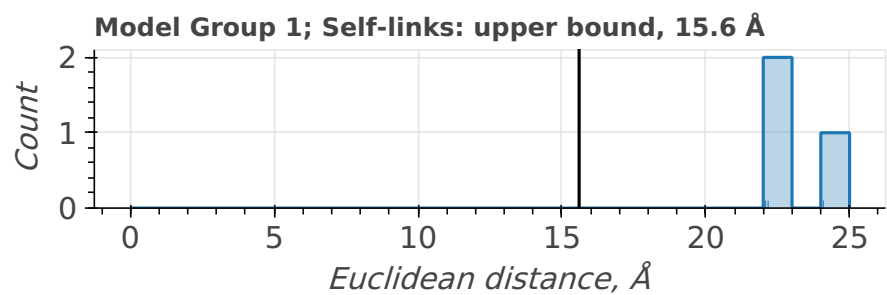
There are 14 crosslinking restraints combined in 8 restraint groups.

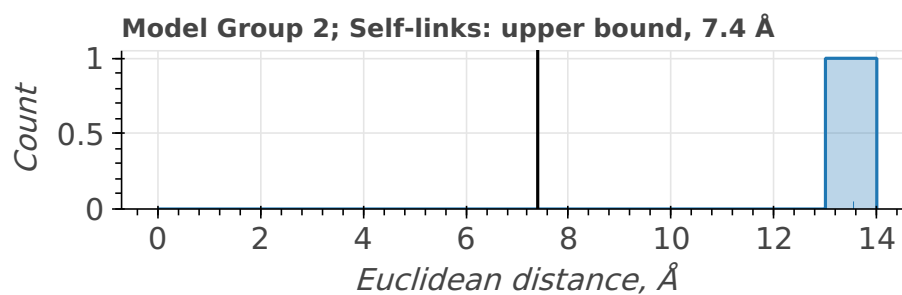
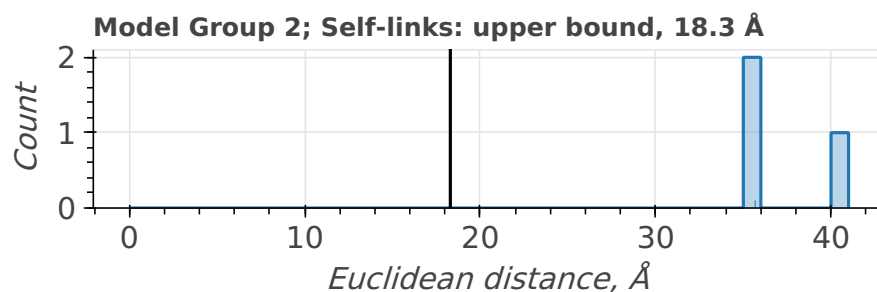
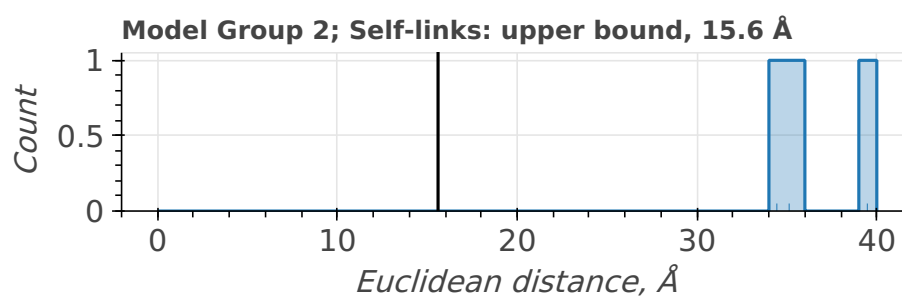
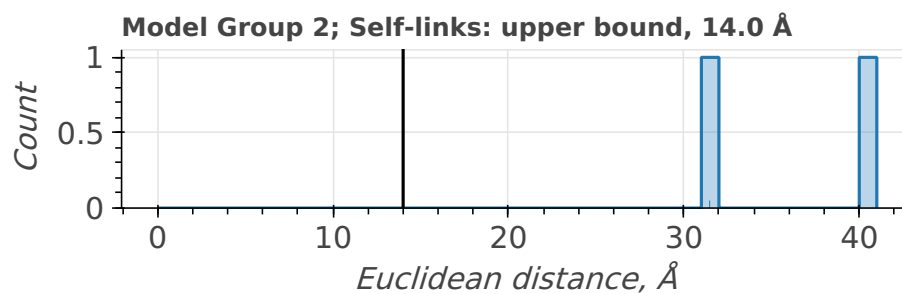
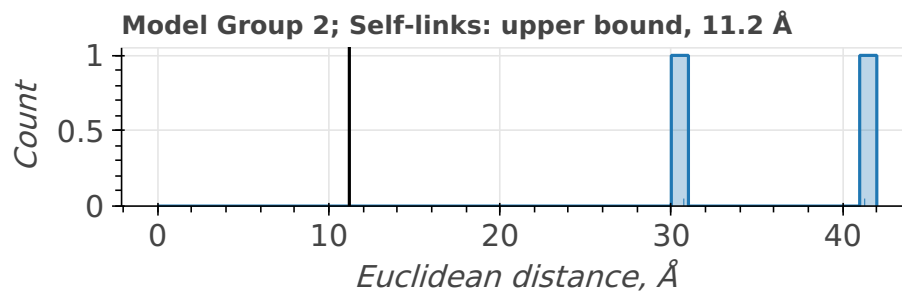
Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
EGS	LYS	CB	LYS	CB	upper bound	15.6	3
EGS	LYS	CA	LYS	CA	upper bound	18.3	3
BS3	LYS	CA	LYS	CA	upper bound	14.0	2
BS3	LYS	CB	LYS	CB	upper bound	11.2	2
BS2G	LYS	CB	LYS	CB	upper bound	7.4	1
BS2G	LYS	CA	LYS	CA	upper bound	10.3	2
DST	LYS	CA	LYS	CA	upper bound	8.7	1

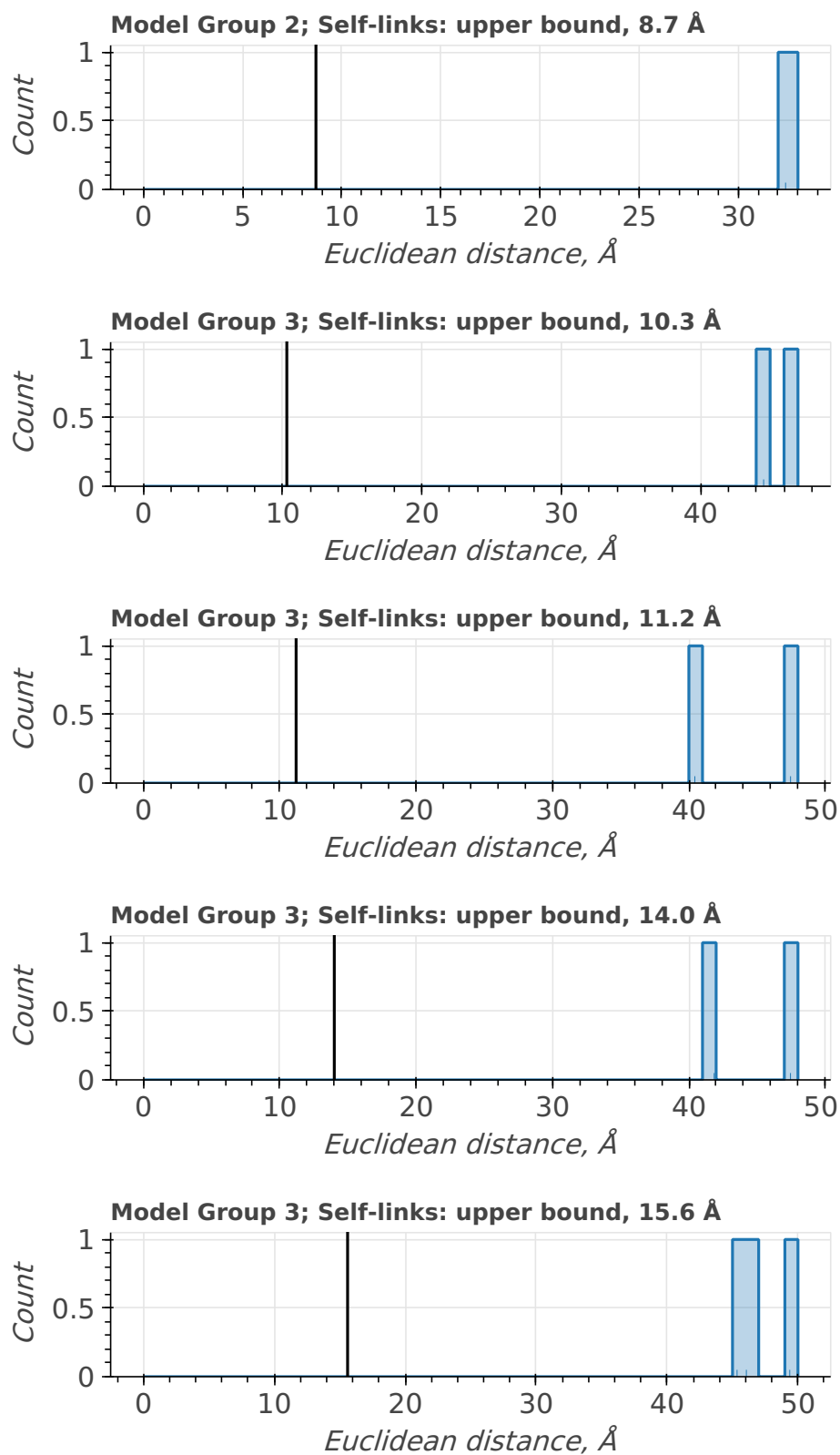
Distograms of individual restraints

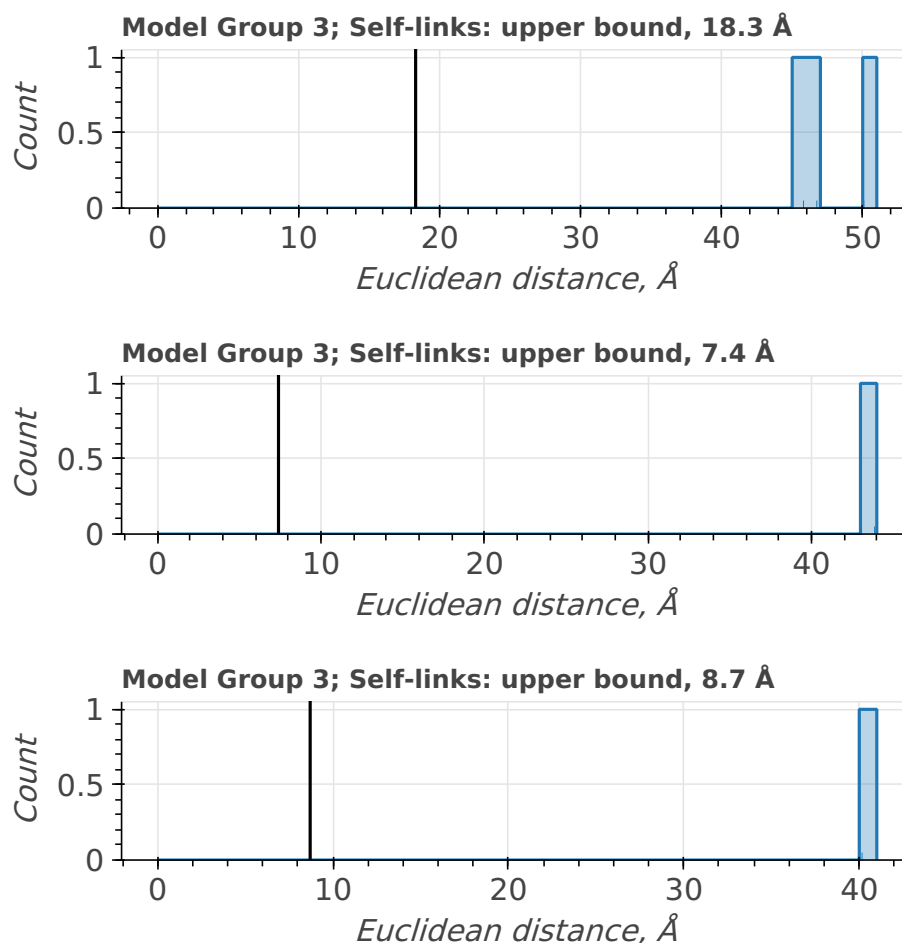
Restraints with identical thresholds are grouped into one plot. Only the best distance per restraint per model group/ensemble is plotted. Inter- and intramolecular (including self-links) restraints are also grouped into one plot. Distance for a restraint between coarse-grained beads is calculated as 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.











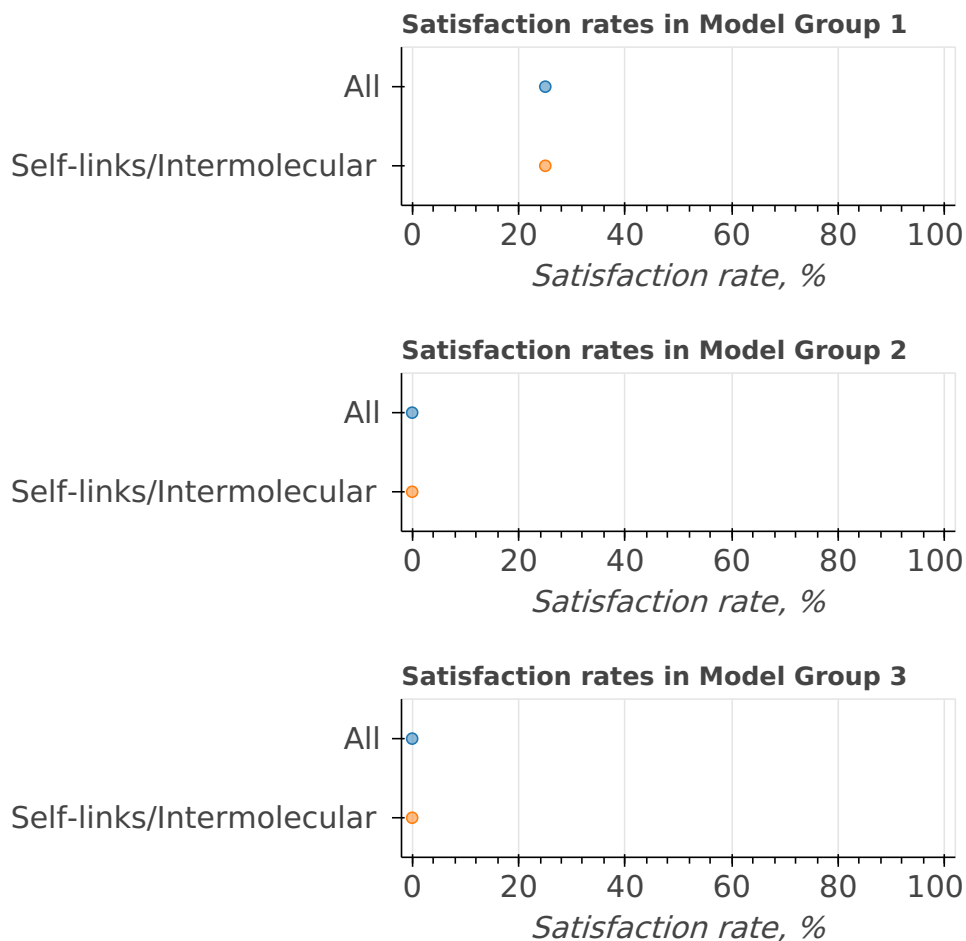
Satisfaction of restraints

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=8)
1	1	1	1/1	All	25.00	75.00	8
				Self-links/ Intermolecular	25.00	75.00	8
2	2	2	1/1	All	0.00	100.00	8
				Self-links/ Intermolecular	0.00	100.00	8
3	3	3	1/1	All	0.00	100.00	8
				Self-links/ Intermolecular	0.00	100.00	8

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.



Single molecule FRET

Validation for this section is under development.

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.