Integrative Structure Validation Report ?

October 09, 2025 - 04:37 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
ATSAS Version 3.2.1 (r14885)

PDB ID	8ZZ4	pdb_00008zz4
PDB-Dev ID	PDBDEV_00000004
Structure Title	Structure of K63-linked Diubiquitin
Structure Authors	Liu, Z.; Gong, Z.; Cao, Y.; Ding, Y.H.; Dong, M.Q.; Lu, Y.B.; Zhang, W.P.; Tang, C.
Deposited on	2017-09-08

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 3 model(s). A total of 5 dataset(s) were used to build this entry.

Name	Type	Count
Crosslinking-MS data	Experimental data	1
SAS data	Experimental data	1

| Name	Type	Count
Single molecule FRET data	Experimental data	1
Experimental model	Starting model	2

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.
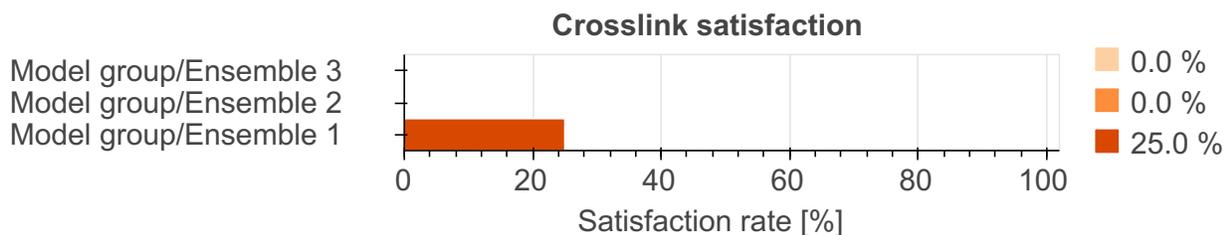
Model Quality: MolProbity Analysis ?

Data Quality ?

Data Quality for SAS: Rg Analysis

Fit to Data Used for Modeling ?

Crosslink satisfaction
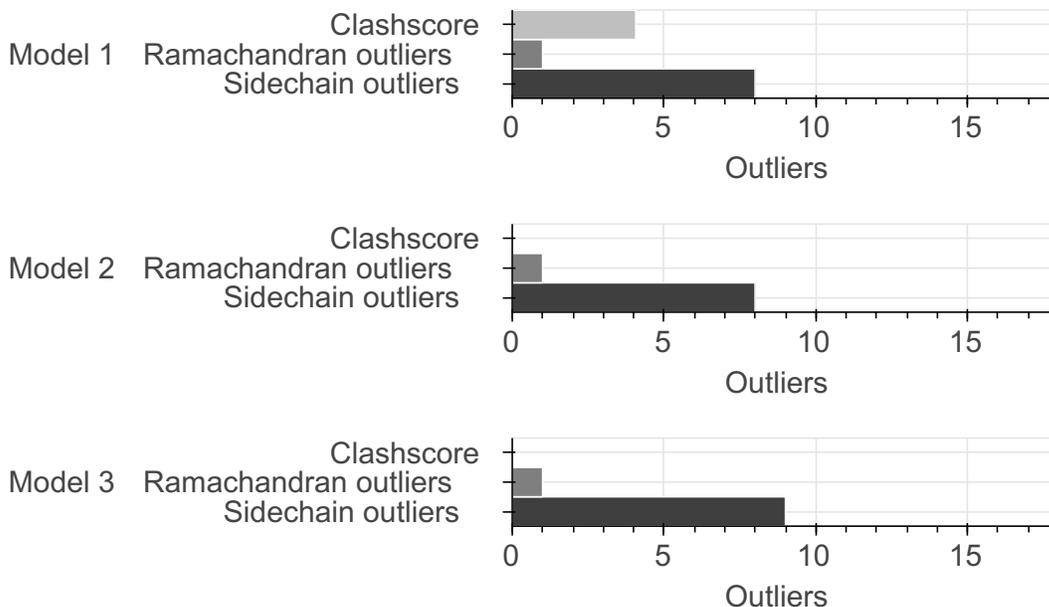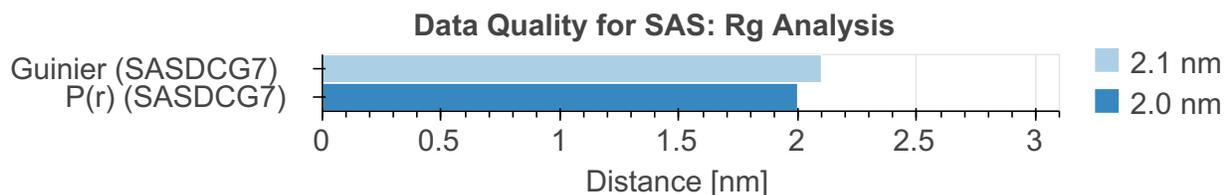
2. Model Details ⑦

2.1. Ensemble information ⑦

This entry consists of 0 distinct ensemble(s).

2.2. Representation ⑦

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					

2.3. Datasets used for modeling ⑦

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	pdb_00001ubq
3	Experimental model	PDB	pdb_00002n2k
4	Crosslinking-MS data	Zenodo	10.5281/zenodo.1006721
5	Single molecule FRET data	Zenodo	10.5281/zenodo.1006721

2.4. Methodology and software ⑦

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	Not available	Not available	Not available	Not available	True	False

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

3. Data quality ⑦

3.1. SAS ⑦

3.1.1. Scattering profile ⑦

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.

Log I(q) vs q with for SASDCG7

Log I(q) vs Log q for SASDCG7
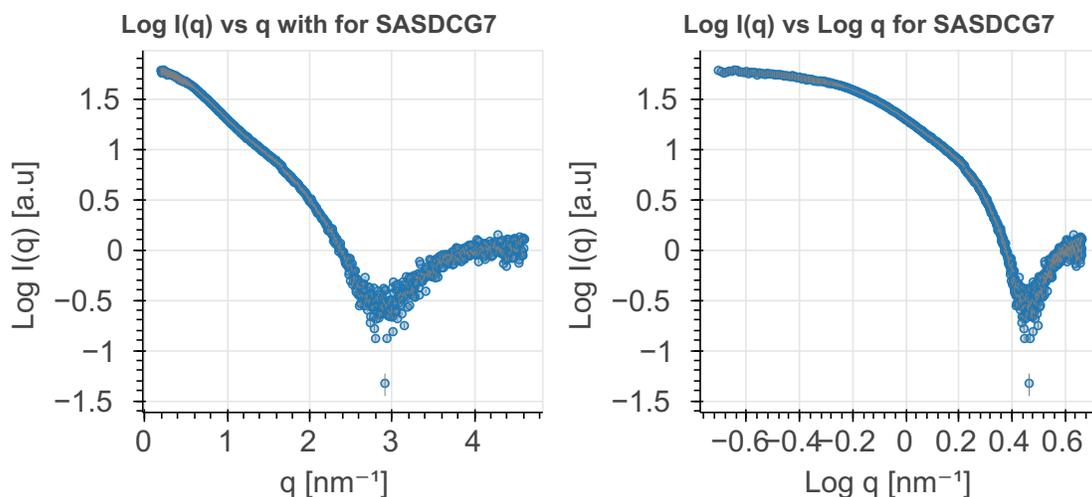
3.1.2. Key experimental estimates ❓

<u>Molecular weight (MW) estimates from experiments and analysis</u>: Theoretical MW can be compared to SAS-derived values using the forward scatter (*I(0)*) and the known concentration and partial specific volume of the scattering particle, or as estimated from the Porod volume and partial specific volume (Trewhella et al., 2017, Trewhella et al., 2023).

| SASDB ID	Chemical composition MW	Standard MW	Porod Volume/MW
SASDCG7	13.0 kDa	Not available	1.69 nm³/kDa

<u>Volume estimates from experiments and analysis</u>: 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 nm³	Not available	Not available	2.60 mg/mL

3.1.3. Flexibility analysis ❓

In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas flexible-modular, fully unfolded domains or extended/stiff rodshaped domains lack a discernible plateau (Rambo and Tainer 2013). A bell-shaped Kratky plot ($q^2I(q)$ *vs. q*) with a well-defined maximum is observed for compact/folded structures. For partially flexible/modular or extended structures the Kratky plot can show multiple maxima and/or an increase in intensity at higher q-values depending on the degree of flexibility and extension. Fully intrinsically disordered structures yield a Kratky plot that systematically increases with increasing q values and will be near linear for highly extended molecules. The dimensionless Kratky plot ($(qR_g)^2I(q)$ *vs. qR_g*) is useful for quantifying differences in shape and foldedness among scattering objects of different sizes (Trewhella et. al., 2023).

<u>Flexibility analysis for SASDCG7</u>

Porod-Debye plot SASDCG7

Dimensionless Kratky plot SASDCG7
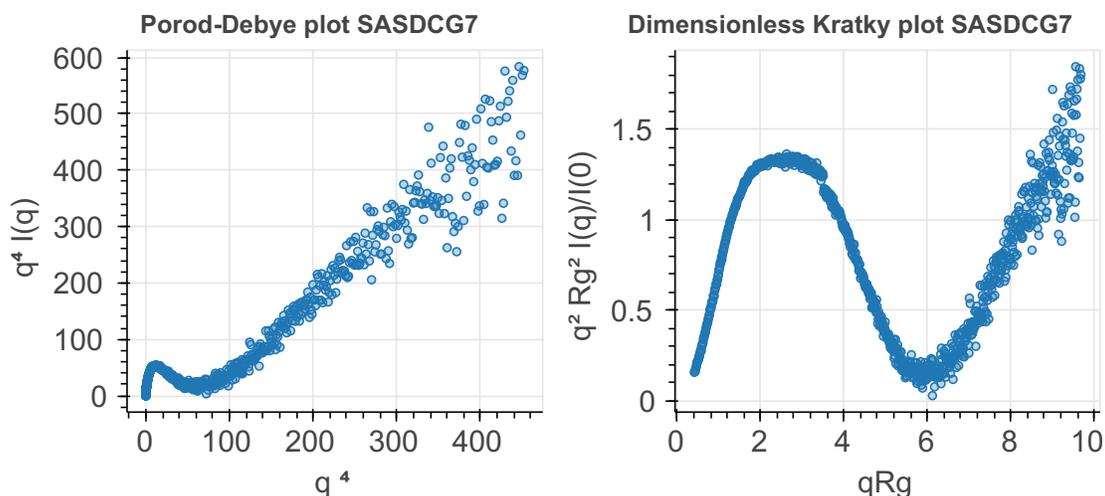
3.1.4. Pair-distance distribution analysis ❓

The the atom-pair distance distribution function (PDDF) or *P(r)* represents the distribution of distances between all pairs of atoms within the particle weighted by the respective scattering contrasts (Moore, 1980). The second moment of *P(r)* yields the radius of gyration (Rg), which is a measure of the overall size and shape of a macromolecule (i.e. the spatial distribution of volume elements). 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.

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}.

P(r) SASDCG7

P(r) extrapolated fit for SASDCG7
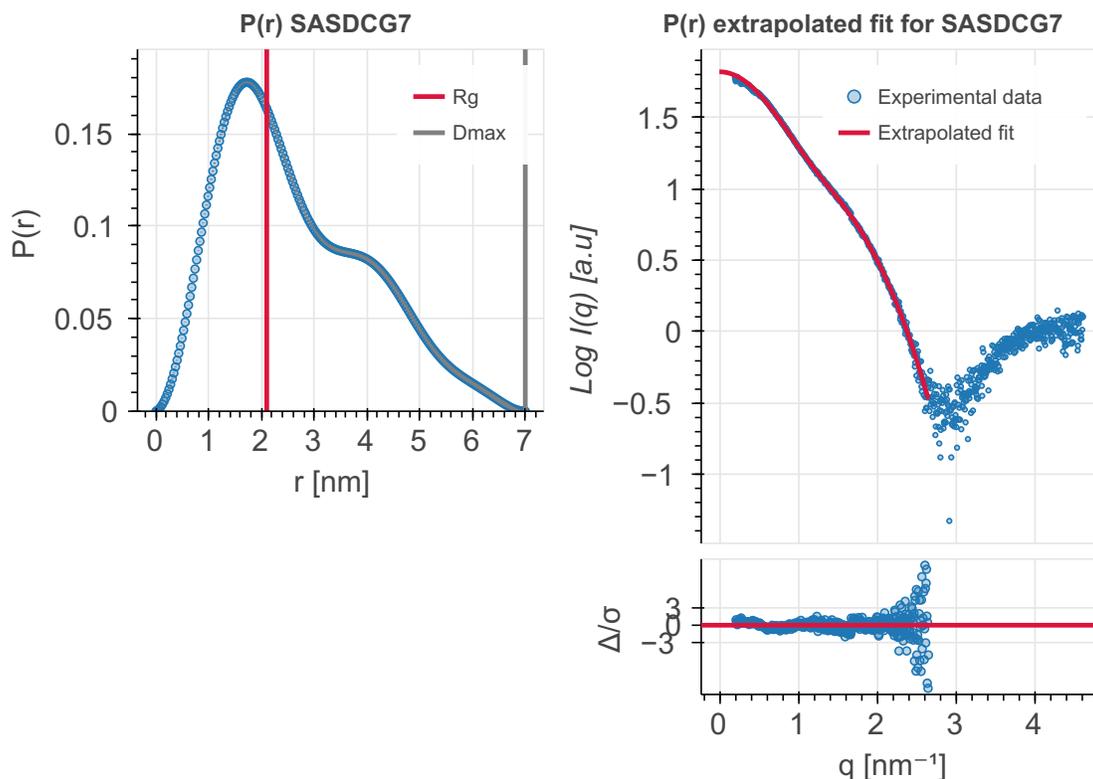
3.1.5. Guinier analysis ⍰

The linearity of the Guinier plot ($\ln(q)$ *vs.* q^2) at very-low angle ($qRg < 1.3$) is a sensitive indicator of the quality of the sample in relation to its homogeneity; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the same size. Deviations from linearity can point to strong interference effects from particle attraction or repulsion, polydispersity of the samples, or improper background subtraction (Feigin et al., 2013). Residual difference plots and Pearson correlation coefficient determination (R^2) are measures to assess quality of the linear fit to the Guinier region. A perfect fit has an R^2 value of 1. Residual values should be equally and randomly spaced around the horizontal axis with no evident systematic upward or downward curvature.

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

Guinier plot for SASDCG7
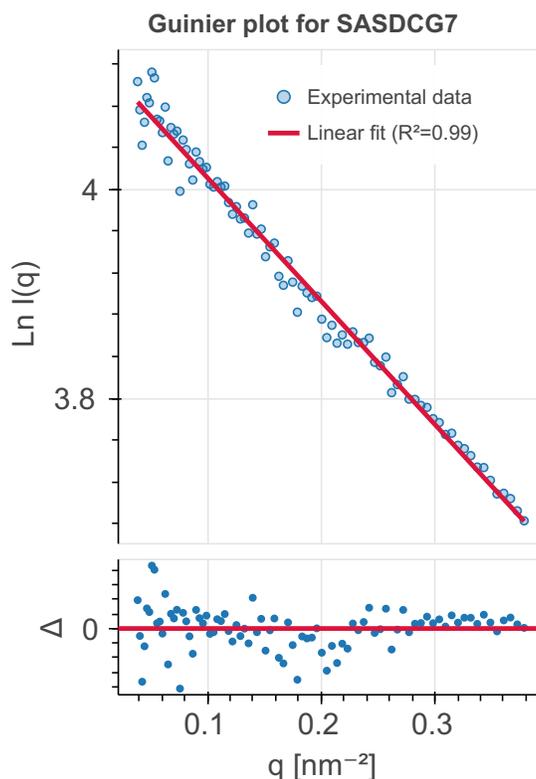
3.2. Crosslinking-MS

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).

Crosslinking-MS dataset is not available in the PRIDE Crosslinking database.

3.4. Single molecule FRET ⍰

Validation for this section is under development.

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 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

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
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
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

Model ID	Analysed	Favored	Allowed	Outliers
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

5. Fit to Data Used for Modeling Assessment ?

5.1. SAS ?

Model and fits displayed below were obtained from SASBDB.

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

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 8 restraint groups.

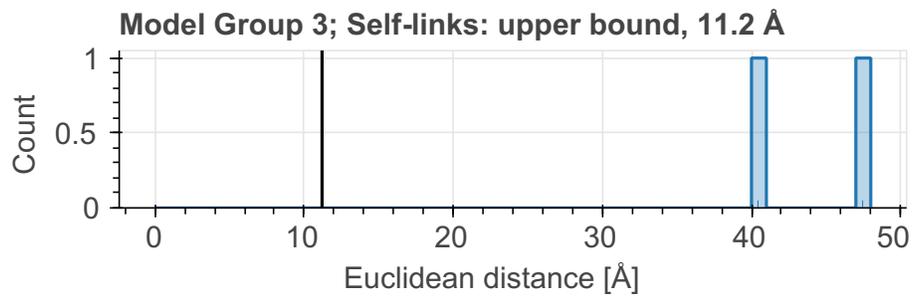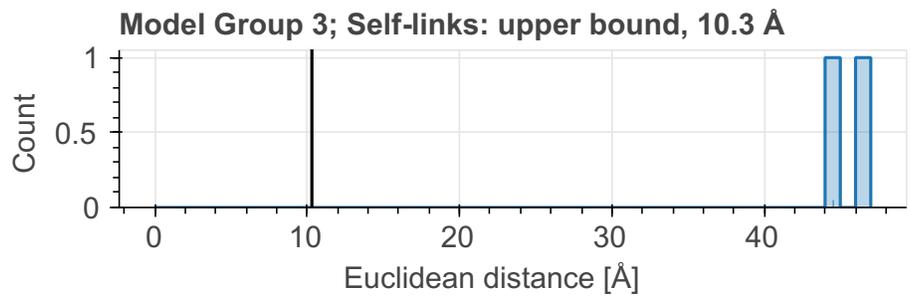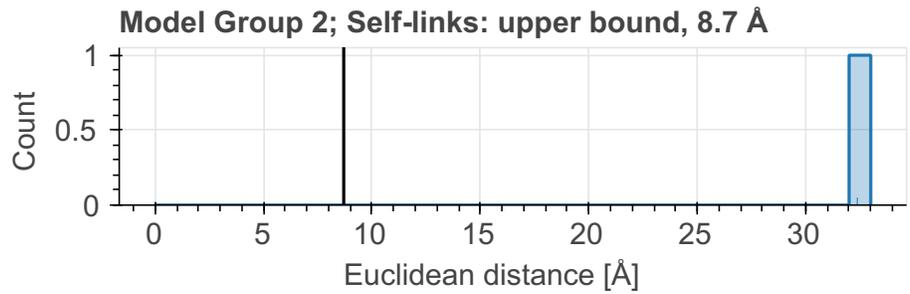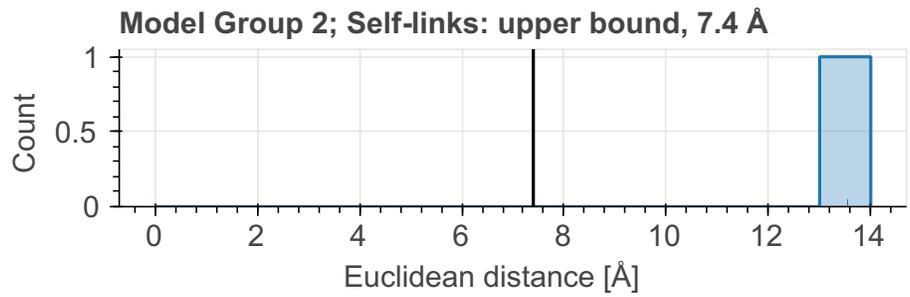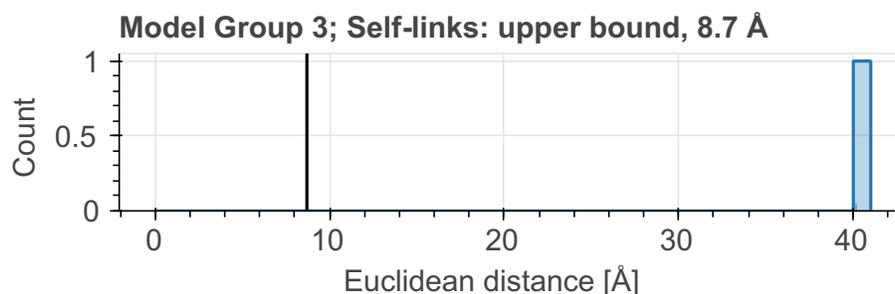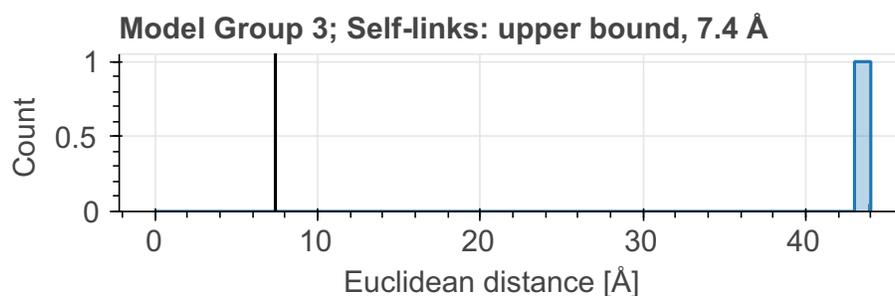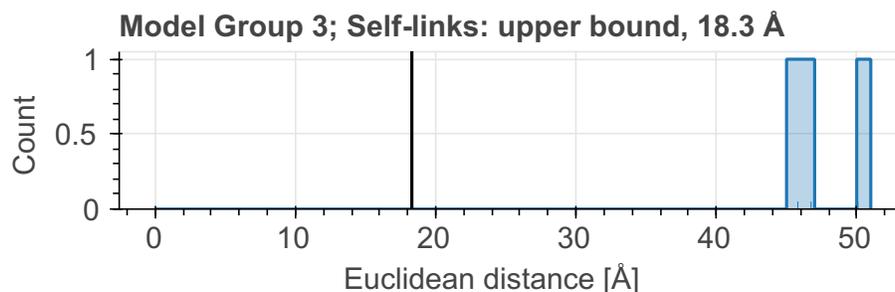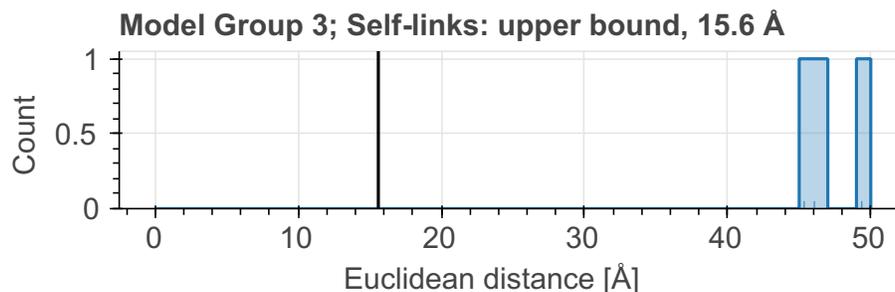| Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
EGS	LYS	CB	LYS	CB	upper bound	15.60	3
EGS	LYS	CA	LYS	CA	upper bound	18.30	3
BS3	LYS	CA	LYS	CA	upper bound	14.00	2
BS3	LYS	CB	LYS	CB	upper bound	11.20	2
BS2G	LYS	CB	LYS	CB	upper bound	7.40	1
BS2G	LYS	CA	LYS	CA	upper bound	10.30	2
DST	LYS	CA	LYS	CA	upper bound	8.70	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, 10.3 Å

Model Group 1; Self-links: upper bound, 11.2 Å

Model Group 1; Self-links: upper bound, 14.0 Å

Model Group 1; Self-links: upper bound, 15.6 Å

Model Group 1; Self-links: upper bound, 18.3 Å

Model Group 1; Self-links: upper bound, 7.4 Å

Model Group 1; Self-links: upper bound, 8.7 Å

Model Group 2; Self-links: upper bound, 10.3 Å

Count / Euclidean distance [Å]

Model Group 2; Self-links: upper bound, 11.2 Å

Count / Euclidean distance [Å]

Model Group 2; Self-links: upper bound, 14.0 Å

Count / Euclidean distance [Å]

Model Group 2; Self-links: upper bound, 15.6 Å

Count / Euclidean distance [Å]

Model Group 2; Self-links: upper bound, 18.3 Å

Count / Euclidean distance [Å]

Model Group 2; Self-links: upper bound, 7.4 Å

Model Group 2; Self-links: upper bound, 8.7 Å

Model Group 3; Self-links: upper bound, 10.3 Å

Model Group 3; Self-links: upper bound, 11.2 Å

Model Group 3; Self-links: upper bound, 14.0 Å

Model Group 3; Self-links: upper bound, 15.6 Å

Model Group 3; Self-links: upper bound, 18.3 Å

Model Group 3; Self-links: upper bound, 7.4 Å

Model Group 3; Self-links: upper bound, 8.7 Å

5.2.2. Satisfaction of restraints ⑦

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.

Satisfaction rates in Model Group 1

Satisfaction rates in Model Group 2

Satisfaction rates in Model Group 3
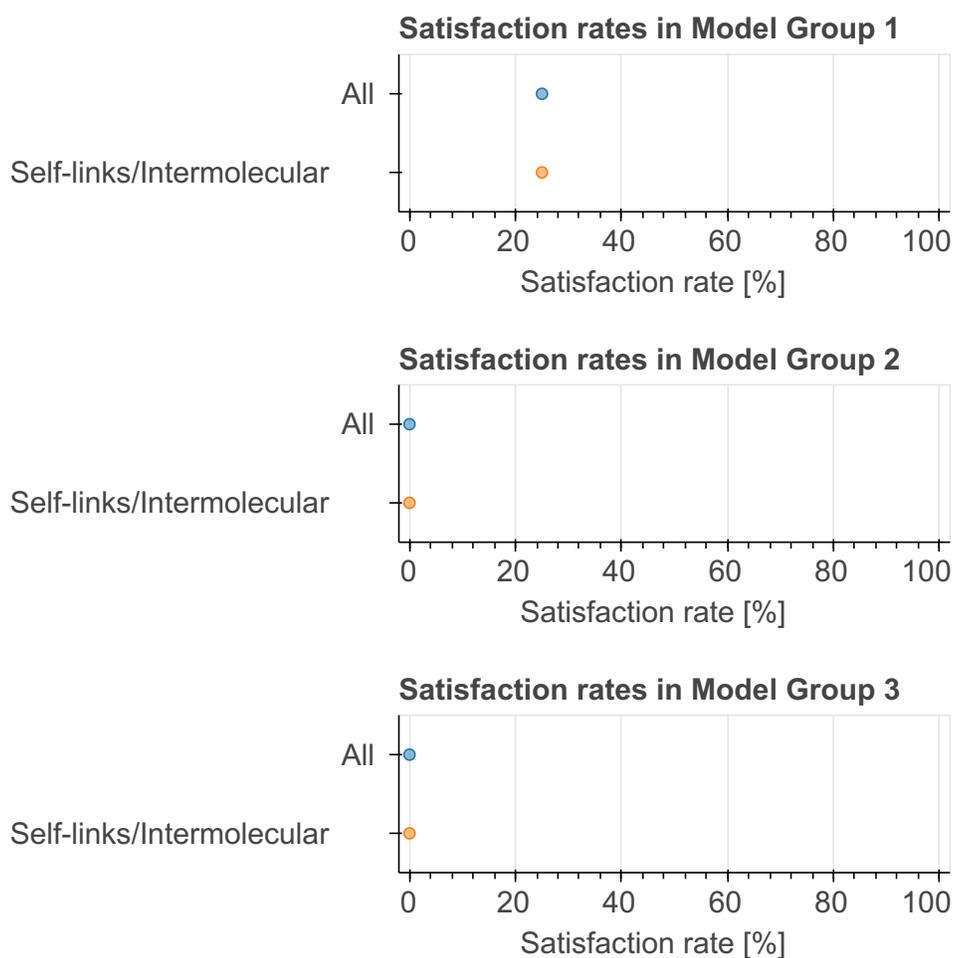
5.4. Single molecule FRET ?

Validation for this section is under development.

6. Fit to Data Used for Validation Assessment ⊙

Validation for this section is under development.