

Integrative Structure Validation Report

October 09, 2025 - 04:49 PM PDT

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

IHMValidation Version 3.0

Python-IHM Version 2.5

pyHMMER Version 0.11.1

EMDB validation analysis Version 0.0.1.dev127

ChimeraX Version 1.9

Chimera Version 1.19

MapQ Version 1.8.1

PDB ID	9A8W pdb_00009a8w
Structure Title	Integrative structure of human SNAPc-DNA
Structure Authors	Shah, S.Z.; Perry, T.N.; Graziadei, A.; Cecatiello, V.; Kaliyappan, T.; Misiaszek, A.D.; Mueller, C.W.; Ramsay, E.P.; Vannini, A.
Deposited on	2024-10-23

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 8 dataset(s) were used to build this entry.

Name	Type	Count
3DEM volume	Experimental data	3
Crosslinking-MS data	Experimental data	1
Experimental model	Starting model	3
De Novo model	Starting model	1

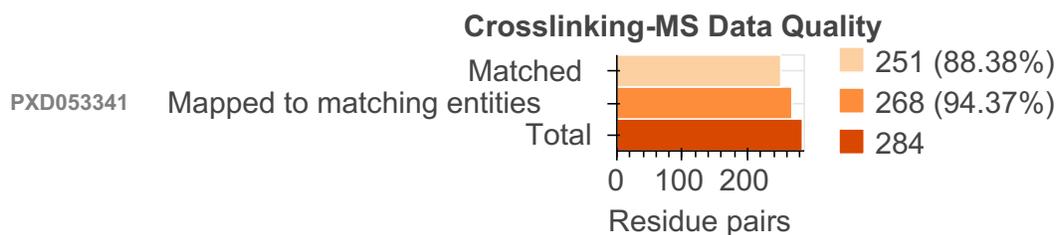
1.2. Overall quality ?

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.

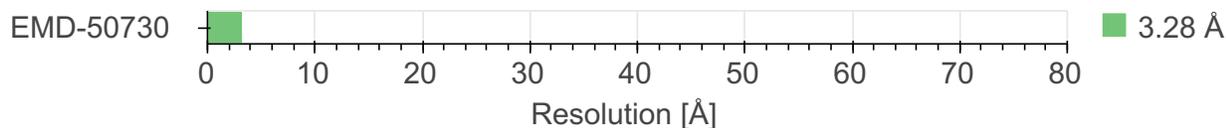
Model Quality: Excluded Volume Analysis ?



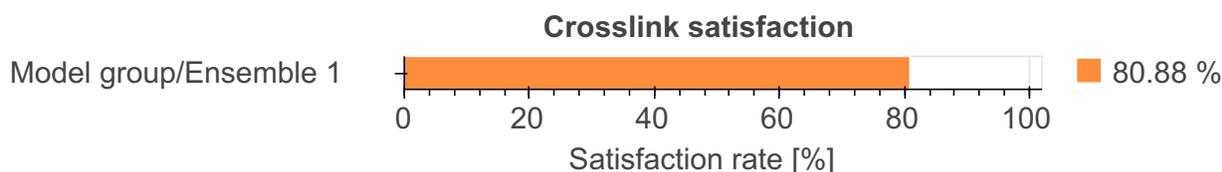
Data Quality ?



3DEM resolution



Fit to Data Used for Modeling ?



2. Model Details ?

2.1. Ensemble information ?

This entry consists of 1 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	snRNA-activating protein complex subunit 2	A	334	29-87, 97-160, 201-271, 304-334	1-28, 88-96, 161-200, 272-303	100.00 / 67.37	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
		2	snRNA-activating protein complex subunit 3	B	411	27-411	1-26	100.00 / 93.67	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
		3	snRNA-activating protein complex subunit 5	C	98	3-52	1-2, 53-98	100.00 / 51.02	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
		4	snRNA-activating protein complex subunit 1	D	368	1-141, 162-234	142-161, 235-368	100.00 / 58.15	Multiscale: Coarse-grained: 1 - 20 residue(s) per bead
		5	snRNA-activating protein complex subunit 4	E	1519	122-163, 182-416, 424-541, 706-799, 806-859, 1304-1434	41-121, 164-181, 417-423, 542-705, 800-805, 860-1303, 1435-1519	97.37 / 45.57	Multiscale: Coarse-grained: 1 - 20 residue(s) per bead
		6	T strand	F	20	1-20	-	100.00 / 100.00	Coarse-grained: 1 residue(s) per bead
		7	NT strand	G	20	1-20	-	100.00 / 100.00	Coarse-grained: 1 residue(s) per bead

2.3. Datasets used for modeling

There are 8 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
5	Crosslinking-MS data	PRIDE	PXD053341

ID	Dataset type	Database name	Data access code
6	3DEM volume	EMDB	EMD-50730
7	3DEM volume	Zenodo	10.5281/zenodo.14746890
8	De Novo model	MODEL ARCHIVE	ma-pgtjz
9	Experimental model	PDB	pdb_00009fso
10	Experimental model	PDB	pdb_00007zx8
11	Experimental model	PDB	pdb_00007xur
12	3DEM volume	Zenodo	10.5281/zenodo.14746890

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	Sampling	Replica exchange monte carlo	Not available	320000	False	True

There are 4 software packages reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	IMP PMI module	2.20.0	integrative model building	https://integrativemodeling.org
2	Integrative Modeling Platform (IMP)	2.20.0	integrative model building	https://integrativemodeling.org
3	AlphaPulldown	0.30.7	structure prediction	https://github.com/KosinskiLab/AlphaPulldown
4	AlphaFold2	2.3.2	structure prediction	https://github.com/google-deepmind/alphafold

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).

[PXD053341](#)

Number of entities in the crosslinking-MS dataset: 244

Number of entities in the entry: 7

Matching entities:

Entity ID	Molecule name	Crosslinking-MS Entity ID	E-value	Exact match
1	snRNA-activating protein complex subunit 2	dbseq_SNAPC2_target	0.00	True
2	snRNA-activating protein complex subunit 3	dbseq_SNAPC3_target	0.00	True
3	snRNA-activating protein complex subunit 5	dbseq_SNAPC5_target	0.00	True
4	snRNA-activating protein complex subunit 1	dbseq_SNAPC1_target	0.00	True
5	snRNA-activating protein complex subunit 4	dbseq_STREP-SNAPC4-His_target	0.00	True
6	T strand	Not available	Not available	Not available
7	NT strand	Not available	Not available	Not available

Residue pairs stats:

Source	Total	In matched entities	Total matched
9A8W	251	251 (100.00%)	251 (100.00%)
PXD053341	284	268 (94.37%)	251 (88.38%)

3.3. 3DEM

This section describes quality of the 3DEM datasets

[EMD-50730](#)

3.3.1. Experimental information

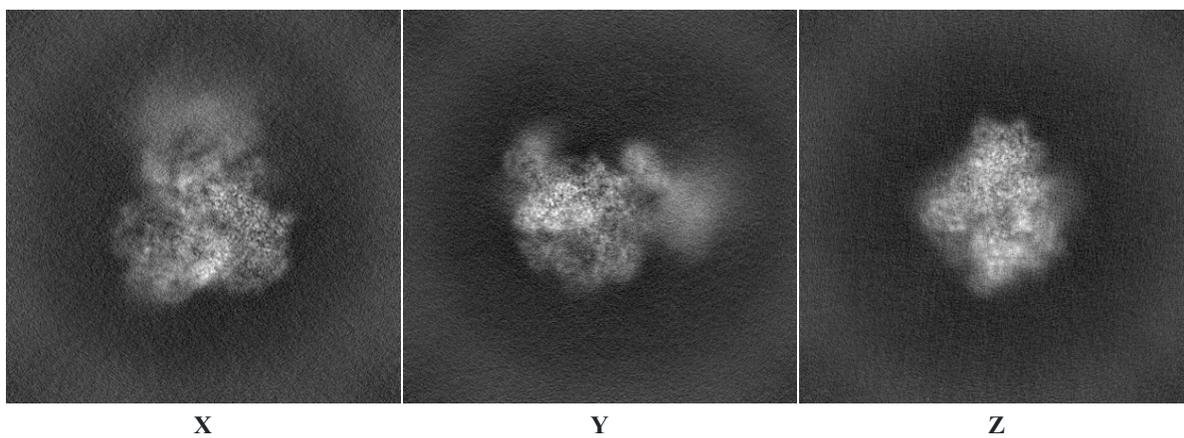
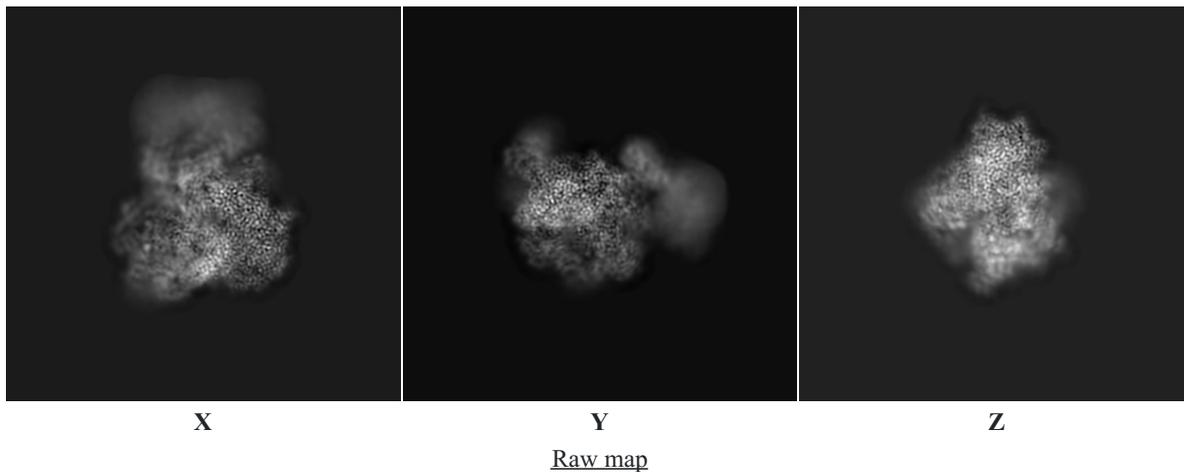
EM reconstruction method:	SINGLE PARTICLE
Resolution:	3.28 Å
Recommended level:	0.150
Estimated volume:	1016.23 nm ³
Specimen preparation:	Preparation ID 1 Vitrification
Map-only validation report:	wwPDB validation report

3.3.2. Map visualisation

This section contains visualisations of the EMDB entry EMD-50730. These allow visual inspection of the internal detail of the map and identification of artifacts. Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

3.3.2.1. Orthogonal projections

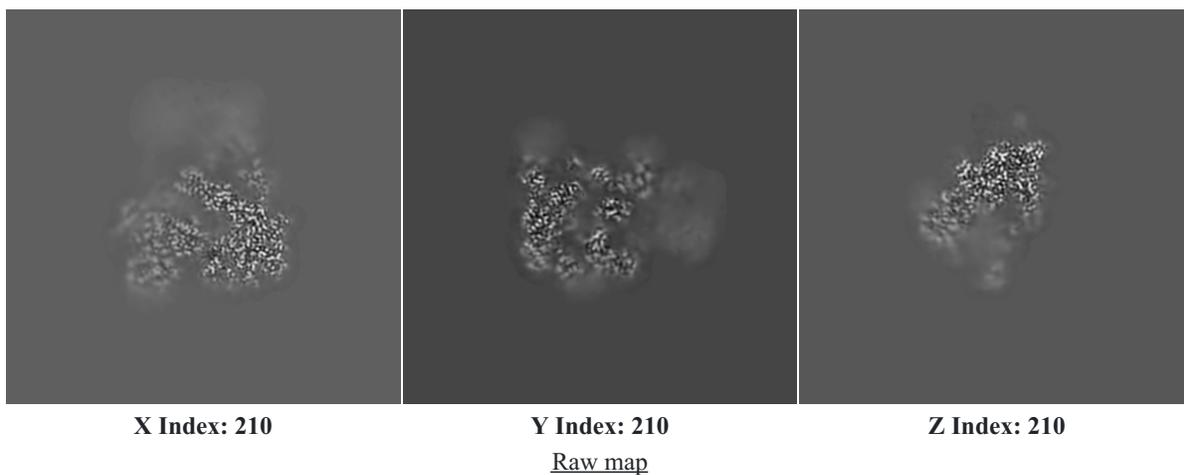
[Primary map](#)

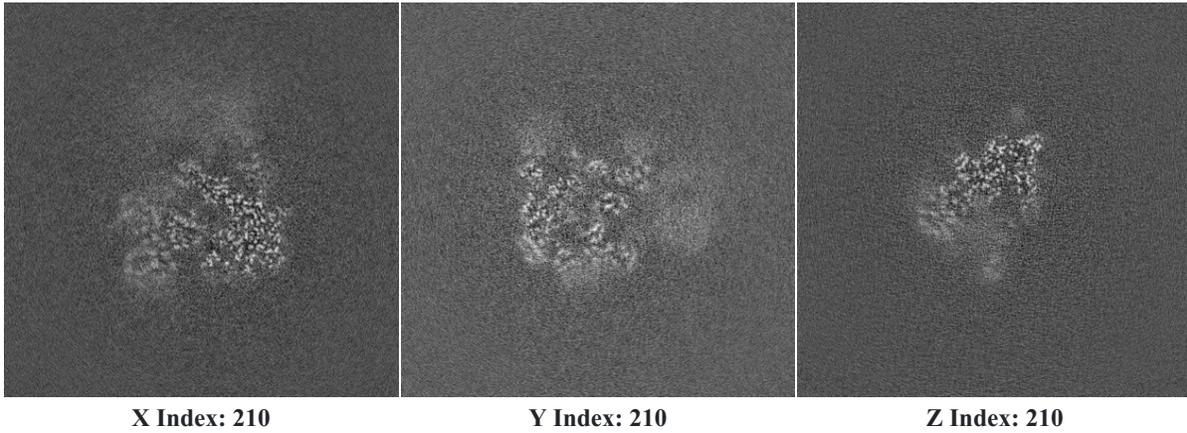


The images above show the map projected in three orthogonal directions.

3.3.2.2. Central slices ?

Primary map

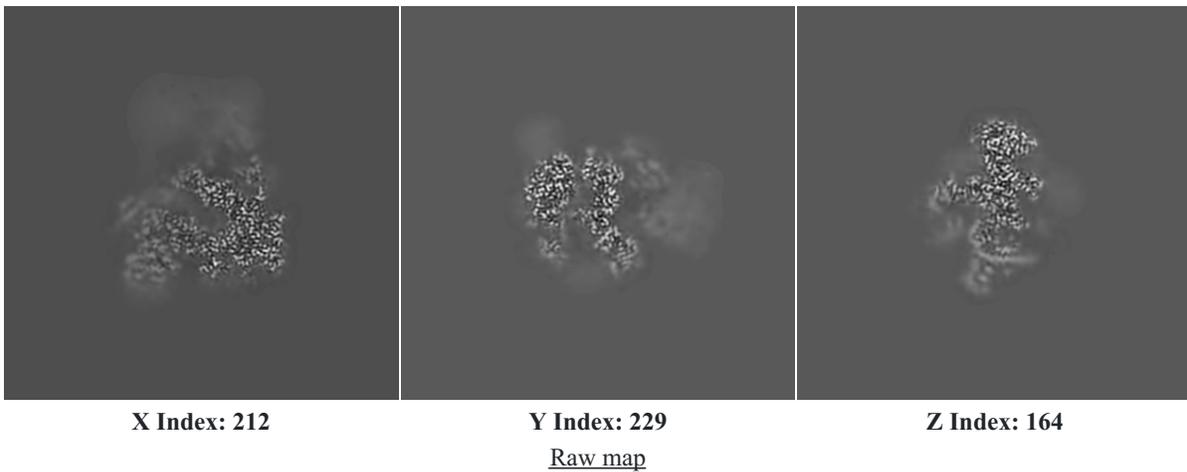




The images above show central slices of the map in three orthogonal directions.

3.3.2.3. Largest variance slices ?

Primary map

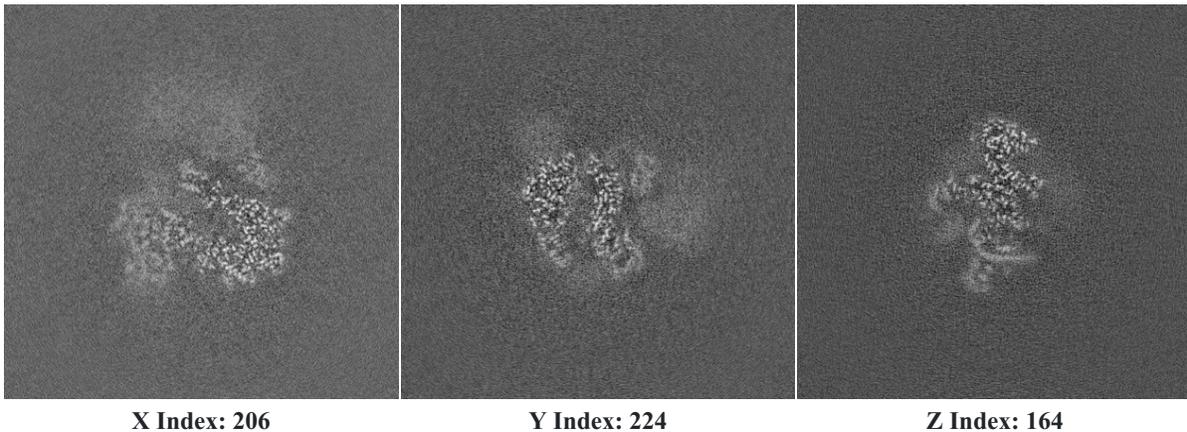


X Index: 212

Y Index: 229

Z Index: 164

Raw map



X Index: 206

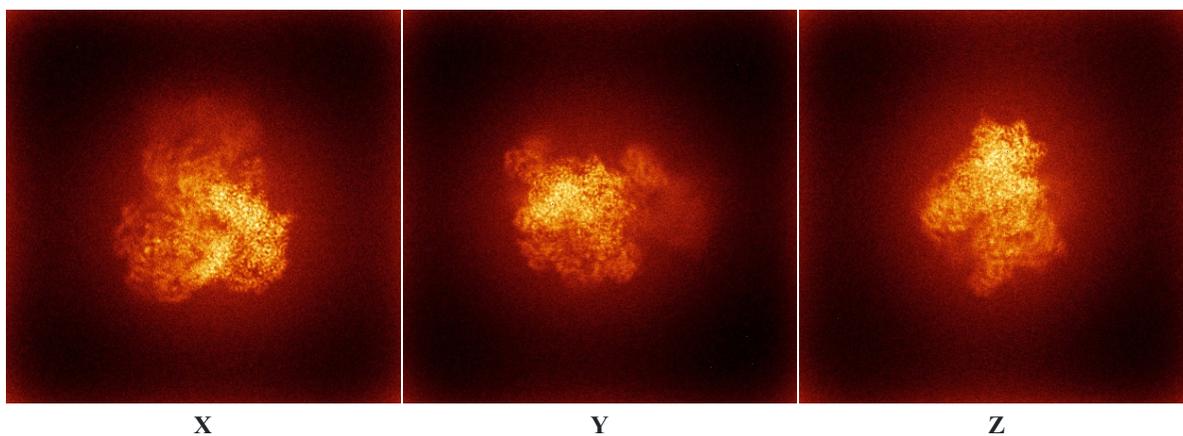
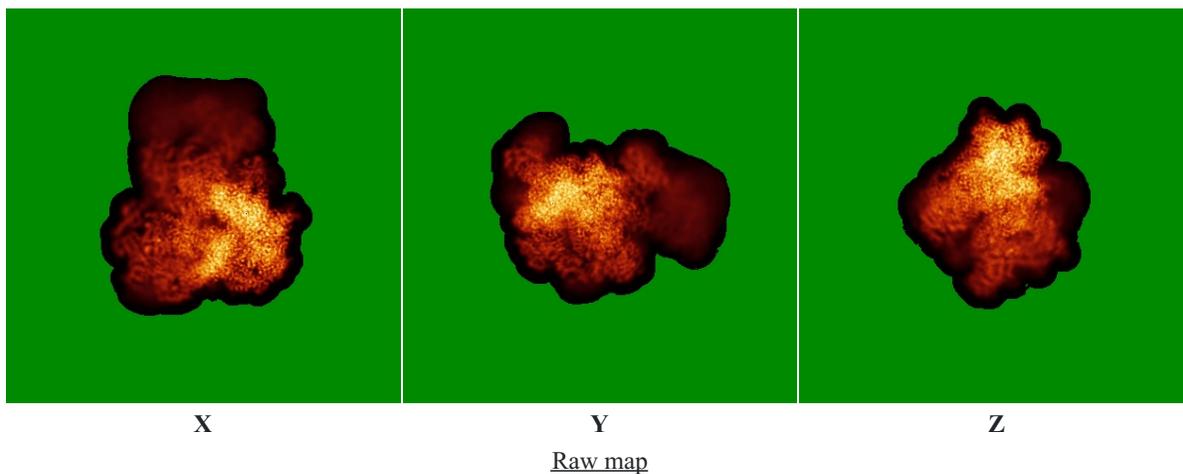
Y Index: 224

Z Index: 164

The images above show the largest variance slices of the map in three orthogonal directions.

3.3.2.4 Orthogonal standard-deviation projections (false-color) ?

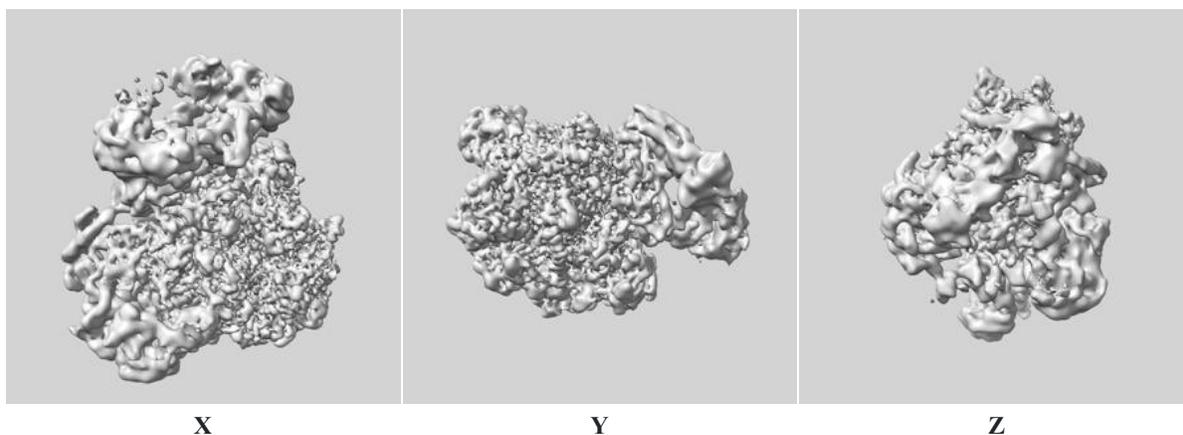
Primary map



The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

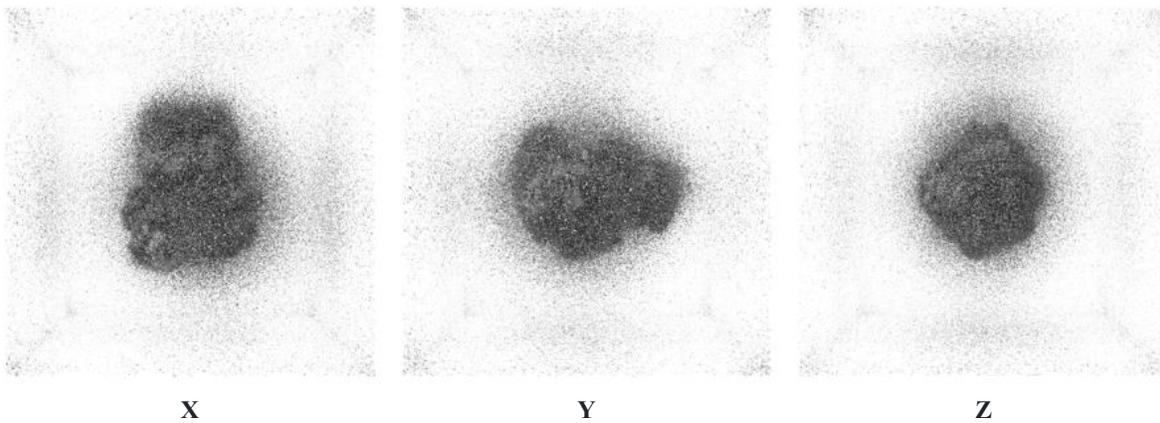
3.3.2.5. Orthogonal surface views

Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.150 . These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

Raw map

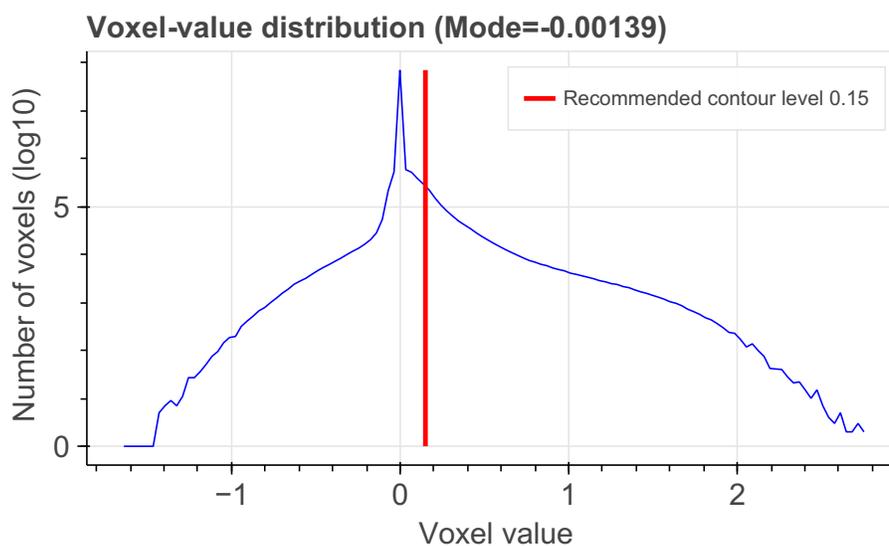


These images show the 3D surface of the raw map. The raw map's contour level 0.140 was selected so that its surface encloses the same volume as the primary map does at its recommended contour level.

3.3.3. Map analysis ?

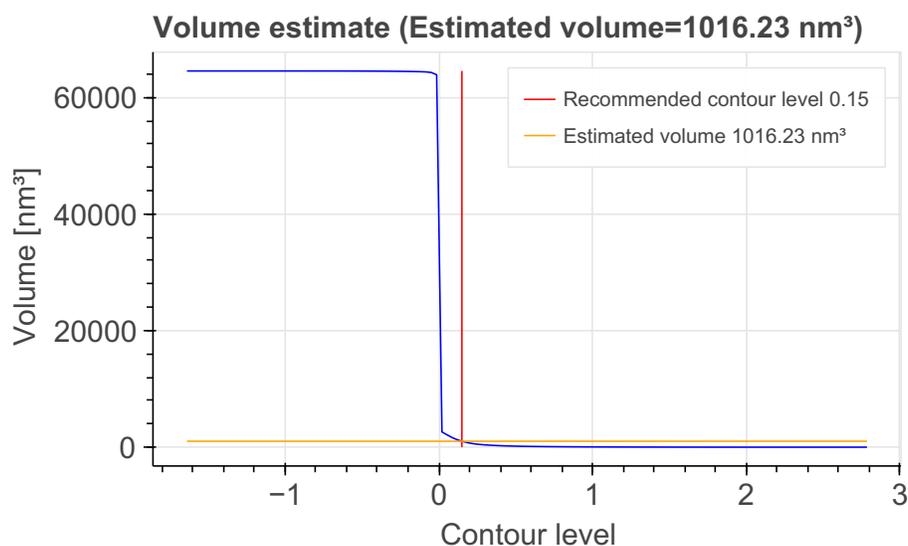
This section contains the results of statistical analysis of the map.

3.3.3.1. Map-value distribution ?



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

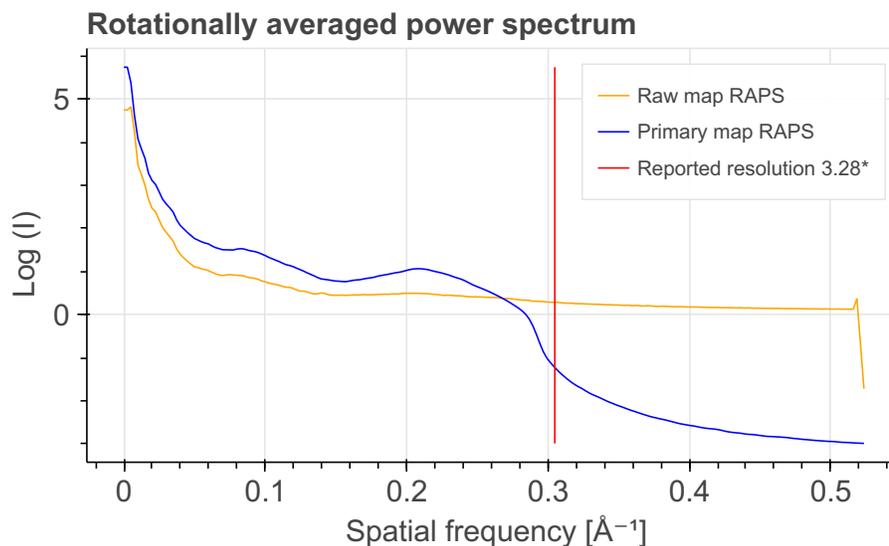
3.3.3.2. Volume estimate ?



The volume at the recommended contour level is 1016.23 nm³.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

3.3.3.3. Rotationally averaged power spectrum ?

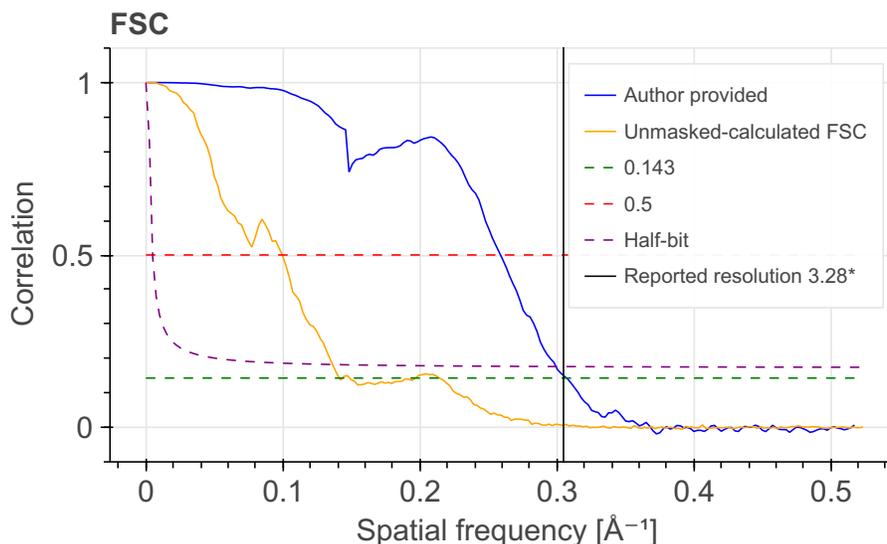


*Reported resolution corresponds to spatial frequency of 0.305 \AA^{-1}

3.3.4. Fourier-Shell correlation ?

3.3.4.1. FSC ?

Fourier-Shell Correlation (FSC) is the most commonly used method to estimate the resolution of single-particle and subtomogram-averaged maps. The shape of the curve depends on the imposed symmetry, mask and whether or not the two 3D reconstructions used were processed from a common reference. The reported resolution is shown as a black line. A curve is displayed for the half-bit criterion in addition to lines showing the 0.143 gold standard cut-off and 0.5 cut-off.



*Reported resolution corresponds to spatial frequency of 0.305 \AA^{-1}

3.3.4.2. Resolution estimates ?

Resolution estimate (\AA)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.28	-	-

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Author-provided FSC curve	3.26	3.86	3.34
Unmasked-calculated*	7.08	10.03	7.34

*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from deposited half-maps intersecting FSC 0.143 CUT-OFF 7.08 differs from the reported value 3.28 by more than 10%.

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.

4.1a. Excluded Volume Analysis ?

Excluded volume satisfaction for the models in the entry are listed below. The Analysed column shows the number of particle-particle or particle-atom pairs for which excluded volume was analysed.

Model ID	Analysed	Number of violations	Excluded Volume Satisfaction (%)
1	1381953	3973	99.71

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 251 crosslinking restraints combined in 251 restraint groups.

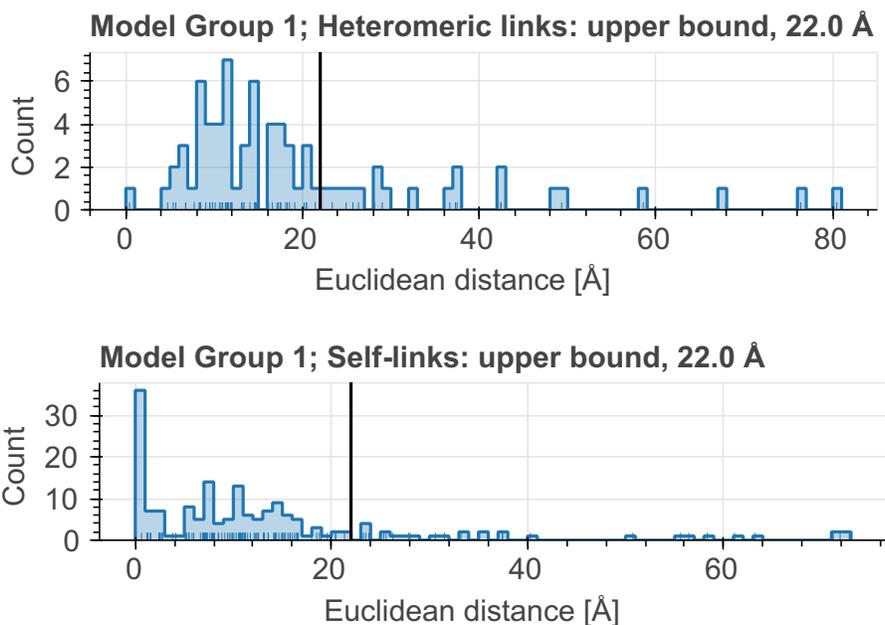
Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
SDA	LYS	CA	PHE	CA	upper bound	22.00	3
SDA	GLU	CA	LYS	CA	upper bound	22.00	44
SDA	LYS	CA	SER	CA	upper bound	22.00	2
SDA	HIS	CA	LYS	CA	upper bound	22.00	6
SDA	LYS	CA	VAL	CA	upper bound	22.00	5
SDA	HIS	coarse-grained	LYS	coarse-grained	upper bound	22.00	9
SDA	LYS	coarse-grained	THR	coarse-grained	upper bound	22.00	5
SDA	GLU	coarse-grained	LYS	coarse-grained	upper bound	22.00	22
SDA	LYS	coarse-grained	TYR	coarse-grained	upper bound	22.00	5

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
SDA	MET	coarse-grained	SER	coarse-grained	upper bound	22.00	1
SDA	LYS	CA	TYR	CA	upper bound	22.00	14
SDA	ALA	CA	LYS	CA	upper bound	22.00	4
SDA	GLN	CA	LYS	CA	upper bound	22.00	2
SDA	PRO	coarse-grained	THR	coarse-grained	upper bound	22.00	2
SDA	GLU	coarse-grained	SER	coarse-grained	upper bound	22.00	9
SDA	GLU	coarse-grained	THR	coarse-grained	upper bound	22.00	5
SDA	ALA	coarse-grained	LYS	coarse-grained	upper bound	22.00	9
SDA	GLY	CA	SER	CA	upper bound	22.00	1
SDA	GLY	CA	PRO	CA	upper bound	22.00	2
SDA	LYS	coarse-grained	SER	coarse-grained	upper bound	22.00	10
SDA	GLU	CA	GLY	CA	upper bound	22.00	6
SDA	ASP	coarse-grained	LYS	coarse-grained	upper bound	22.00	4
SDA	LEU	coarse-grained	LYS	coarse-grained	upper bound	22.00	10
SDA	GLU	CA	MET	CA	upper bound	22.00	2
SDA	ALA	coarse-grained	TYR	coarse-grained	upper bound	22.00	1
SDA	MET	CA	TYR	CA	upper bound	22.00	1
SDA	LYS	CA	PRO	CA	upper bound	22.00	4
SDA	GLY	coarse-grained	LYS	coarse-grained	upper bound	22.00	4
SDA	LYS	CA	TRP	CA	upper bound	22.00	1
SDA	PRO	CA	TYR	CA	upper bound	22.00	3
SDA	SER	CA	TYR	CA	upper bound	22.00	1
SDA	LYS	coarse-grained	PRO	coarse-grained	upper bound	22.00	4
SDA	LEU	CA	LYS	CA	upper bound	22.00	5
SDA	GLY	CA	TYR	CA	upper bound	22.00	2
SDA	SER	coarse-grained	TYR	coarse-grained	upper bound	22.00	2
SDA	GLY	CA	LYS	CA	upper bound	22.00	1
SDA	LYS	CA	LYS	CA	upper bound	22.00	3
SDA	THR	coarse-grained	VAL	coarse-grained	upper bound	22.00	2
SDA	MET	CA	PRO	CA	upper bound	22.00	1
SDA	LYS	coarse-grained	PHE	coarse-grained	upper bound	22.00	2
SDA	HIS	CA	TYR	CA	upper bound	22.00	1
SDA	GLU	CA	SER	CA	upper bound	22.00	2
SDA	ILE	CA	LYS	CA	upper bound	22.00	5
SDA	GLU	coarse-grained	GLY	coarse-grained	upper bound	22.00	1

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
SDA	LEU	CA	TYR	CA	upper bound	22.00	1
SDA	HIS	coarse-grained	THR	coarse-grained	upper bound	22.00	1
SDA	GLU	coarse-grained	MET	coarse-grained	upper bound	22.00	1
SDA	GLN	coarse-grained	LYS	coarse-grained	upper bound	22.00	4
SDA	LYS	coarse-grained	LYS	coarse-grained	upper bound	22.00	2
SDA	LYS	coarse-grained	VAL	coarse-grained	upper bound	22.00	3
SDA	GLY	coarse-grained	TYR	coarse-grained	upper bound	22.00	1
SDA	ASP	CA	TYR	CA	upper bound	22.00	1
SDA	ASP	CA	LYS	CA	upper bound	22.00	2
SDA	ARG	coarse-grained	LYS	coarse-grained	upper bound	22.00	1
SDA	ASN	coarse-grained	LYS	coarse-grained	upper bound	22.00	1
SDA	LYS	coarse-grained	MET	coarse-grained	upper bound	22.00	1
SDA	MET	coarse-grained	TYR	coarse-grained	upper bound	22.00	1
SDA	GLU	coarse-grained	TYR	coarse-grained	upper bound	22.00	1
SDA	GLY	CA	PHE	CA	upper bound	22.00	1
SDA	ASP	CA	SER	CA	upper bound	22.00	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.



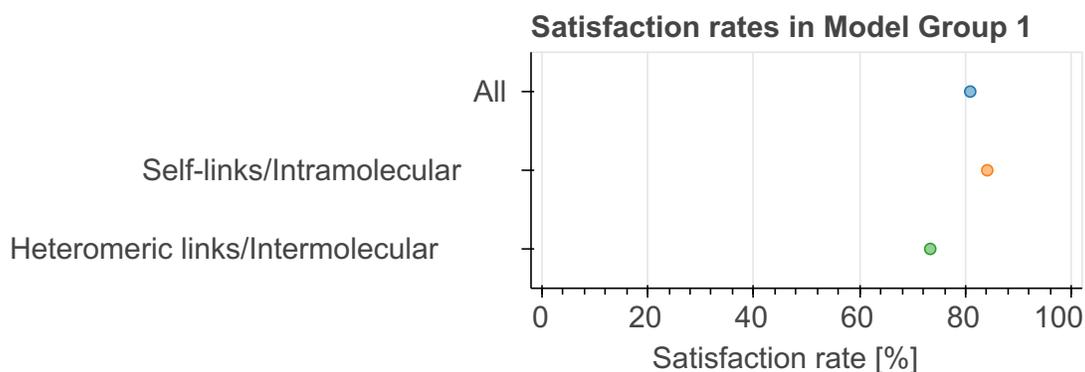
5.2.2. Satisfaction of restraints ?

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=251)
1	1	1	1/5766	All	80.88	19.12	251
				Self-links/ Intramolecular	84.09	15.91	176
				Heteromeric links/ Intermolecular	73.33	26.67	75

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.



5.3. 3DEM

This section describes fit of models to the 3DEM data. Only results for the representative model, selected as a first model with the largest number of asymmetric units.

3DEM validation for coarse-grained structures is under development.

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.