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
MolProbity Version 4.5.2
EMDB validation analysis Version 0.0.1.dev127
ChimeraX Version 1.9
Chimera Version 1.19
MapQ Version 1.8.1

PDB ID	9A89 | pdb_00009a89
PDB-Dev ID	PDBDEV_00000374
Structure Title	Structure of the pre-incision complex in nucleotide excision repair with tilted XPG core
Structure Authors	Yu, J.; Yan, C.Y.; Paul, T.; Brewer, L.; Tsutakawa, S.E.; Tsai, C.-L.; Hamdan, S.; Tainer, J.A.; Ivanov, I.
Deposited on	2024-03-11

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

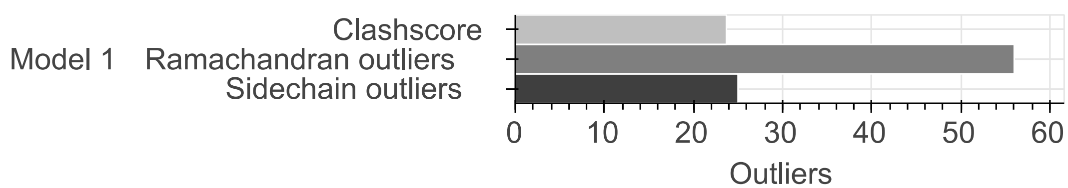
This entry consists of 1 model(s). A total of 25 dataset(s) were used to build this entry.

Name	Type	Count
3DEM volume	Experimental data	1
Crosslinking-MS data	Experimental data	1
Experimental model	Starting model	11
De Novo model	Starting model	12

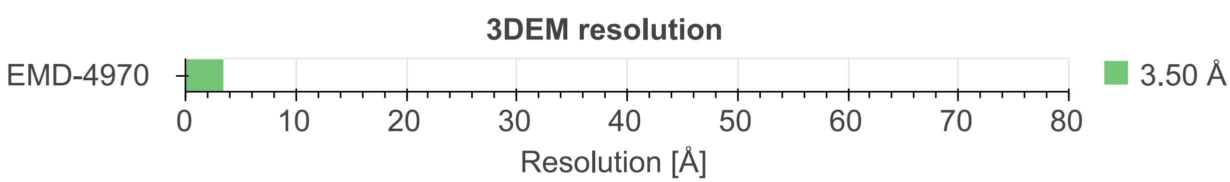
1.2. Overall quality

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.

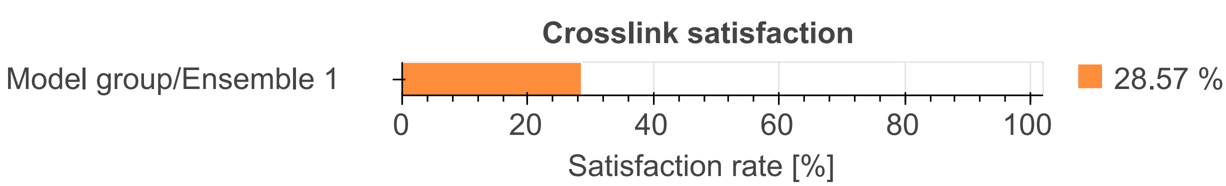
Model Quality: MolProbity Analysis

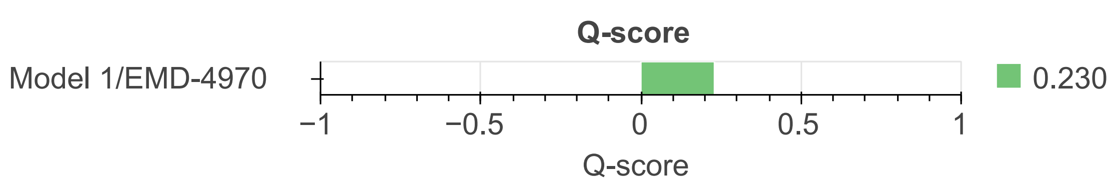


Data Quality



Fit to Data Used for Modeling





2. Model Details

2.1. Ensemble information

This entry consists of 0 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	General transcription and DNA repair factor IIH helicase subunit XPB	A	720	-	34-203, 248-720	89.31 / 100.00	Atomic
		2	General transcription and DNA repair factor IIH helicase subunit XPD	B	760	-	1-760	100.00 / 0.00	Atomic
		3	General transcription factor IIH subunit 4	C	441	-	1-441	100.00 / 100.00	Atomic
		4	General transcription factor IIH subunit 2	D	377	-	1-377	100.00 / 100.00	Atomic
		5	General transcription factor IIH subunit 3	E	292	-	1-292	100.00 / 0.00	Atomic
		6	General transcription factor IIH subunit 5	F	66	-	1-66	100.00 / 100.00	Atomic
		7	DNA repair protein complementing XP-A cells	G	273	-	1-273	100.00 / 0.00	Atomic
		8	General transcription factor IIH subunit 1	H	154	-	1-154	100.00 / 100.00	Atomic
		9	DNA excision repair protein ERCC-5	I	985	-	1-296, 733-985	55.74 / 100.00	Atomic
		10	DNA repair endonuclease XPF Gene: ERCC4, ERCC11, XPF	J	227	-	1-227	100.00 / 100.00	Atomic
		11	DNA excision repair	K	198	-	1-198	100.00 / 100.00	Atomic
		12	Replication protein A 70 kDa DNA-binding subunit, N-terminally processed	L	434	-	1-434	100.00 / 100.00	Atomic
		13	Replication protein A 14 kDa subunit	M	115	-	1-115	100.00 / 100.00	Atomic
		14	Replication protein A 32 kDa subunit	N	225	-	1-225	100.00 / 100.00	Atomic
		15	DNA (66-MER)	O [X]	66	-	1-66	100.00 / 100.00	Atomic
		16	DNA (66-MER)	P [Y]	66	-	1-66	100.00 / 100.00	Atomic
		17	IRON/SULFUR CLUSTER	Q [B]	Non-polymeric	-	-	Not available / Not available	Atomic

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
		18	ZINC ION	R [D]	Non-polymeric	-	-	Not available / Not available	Atomic
				S [D]					
				T [D]					
				U [E]					
				V [E]					
				W [G]					
				Y [L]					
		19	MAGNESIUM ION	X [J]	Non-polymeric	-	-	Not available / Not available	Atomic
				Z [Y]					
		20	water	AA [J]	Non-polymeric	-	-	Not available / Not available	Atomic

2.3. Datasets used for modeling

There are 25 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	Experimental model	PDB	pdb_00006ro4
2	3DEM volume	EMDB	EMD-4970
3	Experimental model	PDB	pdb_00006tuw
4	Crosslinking-MS data	Not available	10.1038/s41467-019-10745-5
5	De Novo model	Not available	Not available
6	De Novo model	Not available	Not available
7	De Novo model	Not available	Not available
8	Experimental model	PDB	pdb_00006sxa
9	Experimental model	PDB	pdb_00006sxb
10	Experimental model	PDB	pdb_00002bgw
11	De Novo model	Not available	Not available
12	Experimental model	PDB	pdb_00004gop
13	Experimental model	PDB	pdb_00006i52
14	Experimental model	PDB	pdb_00001jmc
15	Experimental model	PDB	pdb_00001l1o
16	Experimental model	PDB	pdb_00002jnw

ID	Dataset type	Database name	Data access code
17	Experimental model	PDB	pdb_00004mqv
18	De Novo model	Not available	Not available
19	De Novo model	Not available	Not available
20	De Novo model	Not available	Not available
21	De Novo model	Not available	Not available
22	De Novo model	Not available	Not available
23	De Novo model	Not available	Not available
24	De Novo model	Not available	Not available
25	De Novo model	MODEL ARCHIVE	ma-2chon

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
				To construct a model of the pre-incision complex (PlnC), we systematically examined the cryo-EM structures and densities of human apo-TFIIH, TFIIH/XPA/DNA, and XPF/ERCC1, the NMR structure of XPA-ERCC1, and the X-ray structures of the XPG catalytic core and RPA-ssDNA (RPA70, RPA32, and RPA14). The TFIIH/XPA/DNA structure (PDB ID: 6RO4 and EMDB accession code: EMD-4970) was the starting point for model building. The PInC hybrid model has an NER bubble size of 23 nucleotides, matching the 27-nucleotide optimal length of the excision products and the XPF and XPG incision patterns. FEN1 shares 30% sequence identity with the XPG catalytic core (PDB ID: 6TUR, 6TUW, and 6VBH). Thus, we modeled DNA-bound XPG based on the human FEN1/DNA X-ray structure (PDB ID: 5UM9). XPG positioning into the hybrid model was based on existing XL-MS data. In addition, positioning of the XPG core required placement of the 3' DNA junction 8 nucleotides away from the expected position of the DNA lesion near XPD's His135 residue. The two XPG gateway helices (GH1, residues 82-126) and (GH2, residues 734-761) were predicted with AlphaFold2 and positioned in the gap between XPD's Arch and Fe-S domains in accordance with the crosslink data. The XPD-anchor domain (residues 157-296) was predicted by AlphaFold2 and fitted into the TFIIH/XPA/DNA cryo-EM density. The loop connecting GH1 and the XPD-anchor was built with Modeller. To model XPF/ERCC1, we used the cryo-			

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	EM structures of XPF/ERCC1 (PDB ID: 6SXA and 6SXB). We first docked the XPF nuclease domain to the 5' junction. The catalytic metal was oriented 3A away from the scissile phosphodiester bond. Mg2+ ion coordination was based on the Aeropyrum pernix SNF2 structure (PDB ID: 2BGW). A water molecule was placed between Mg2+ ion and the DNA backbone phosphate group. The ERCC1 (HhH)2 domain was oriented to interact with the ssDNA through two DNA hairpins based on the 6SXB structure. The long linkers from the ERCC1 central domain to the ERCC1 (HhH)2 (residues 214-230) and from the XPF nuclease domain to the XPF (HhH)2 (residues 817-847) were built with Modeller. The SF2 helicase-like N-terminal domain of XPF was omitted from the hybrid PInC model due to lack of sufficient structural or biochemical restraints. To model RPA, we used following X-ray structures: Ustilago maydis RPA/ssDNA (PDB ID: 4GOP), yeast RPA/ssDNA (PDB ID: 6I52) and human RPA (PDB ID: 1JMC and 1L1O). The RPA70AB/ssDNA complex was modeled by superimposing the yeast RPA/ssDNA structure (PDB ID: 1JMC) onto the human apo-RPA 70AB (PDB ID: 6I52). Within PInC, only RPA70A, 70B, and 70C can engage DNA due to the size of the NER bubble. RPA70AB was placed close to the 3' junction where it interacts with XPG. We reoriented RPA70C to bind ssDNA near the 5' junction. The RPA70C/ssDNA was modeled by aligning the Ustilago maydis RPA/ssDNA structure (PDB ID: 4GOP) with the human trimer core structure (PDB ID: 1L1O). The orientation of RPA32D and RPA14 follows from the placement of the RPA70C module as they are all connected, forming the trimer core (70C/32D/14). To model XPA, we used the following structures: the cryo-EM TFIIH/XPA/DNA structure (PDB ID: 6RO4), the NMR structure of XPA/ERCC1 (PDB ID: 2JNW), and the human X-ray structure of RPA32C/Smarcal1 N-terminus (PDB ID: 4MQV). The XPA N-terminal extension (residues 1-103), which includes the RPA32C binding helix (residues 22-40), and the C-terminal extension (beta-domain) (residues 235-273) lacked known structural homologues and were modeled using AlphaFold2. The beta-domain was fitted into the TFIIH/XPA/DNA density. To position XPA's N-terminal helix (residues 22-40) we used the X-ray structure of RPA32C/Smarcal1 N-terminus. To assemble the complete PInC model, we also modelled loop regions of TFIIH's core subunits (XPB, XPD, p44, p34, and p52) into the TFIIH/XPA/DNA density.	Not available	False	False

There are 6 software packages reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	AlphaFold2	Not available	model building	https://alphafold.ebi.ac.uk/
2	Modeller	10.40	model building	https://salilab.org/modeller/
3	Clustal Omega	Not available	sequence alignments	https://www.ebi.ac.uk/jdispatcher/msa/clustalo
4	Coot	0.9.8.92	real-space refinement	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
5	Phenix	1.20.1	real-space refinement	https://phenix-online.org/
6	UCSF Chimera	1.18	model visualization	https://www.cgl.ucsf.edu/chimera/

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).

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

3.3. 3DEM ?

This section describes quality of the 3DEM datasets

EMD-4970

3.3.1. Experimental information ?

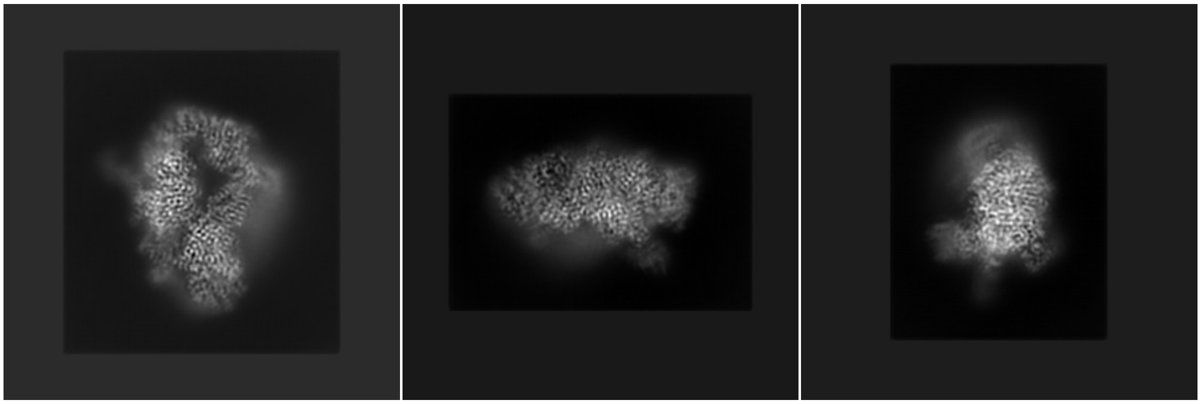
EM reconstruction method:	SINGLE PARTICLE
Resolution:	3.50 Å
Recommended level:	0.015
Estimated volume:	131.36 nm^3
Specimen preparation:	Preparation ID 1 Vitrification
Map-only validation report:	wwPDB validation report

3.3.2. Map visualisation ?

This section contains visualisations of the EMDB entry EMD-4970. 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

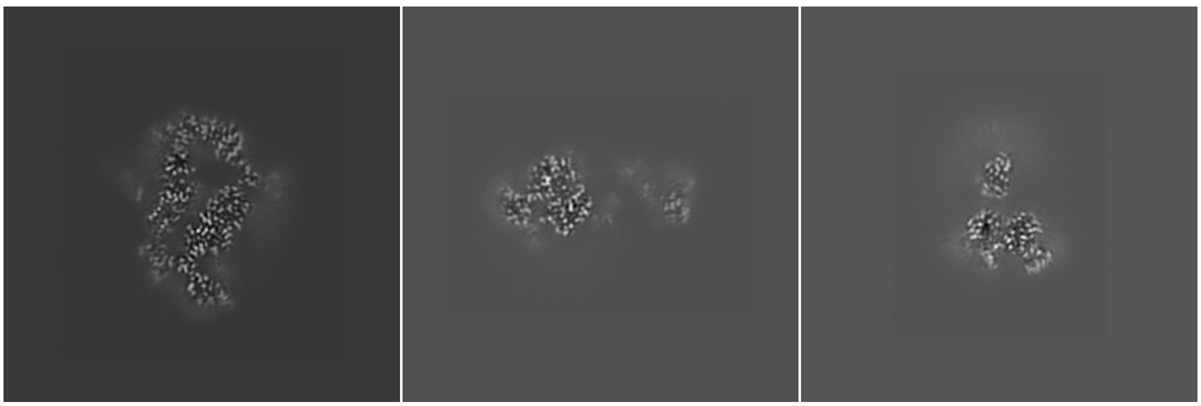


X **Y** **Z**

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

3.3.2.2. Central slices ?

Primary map

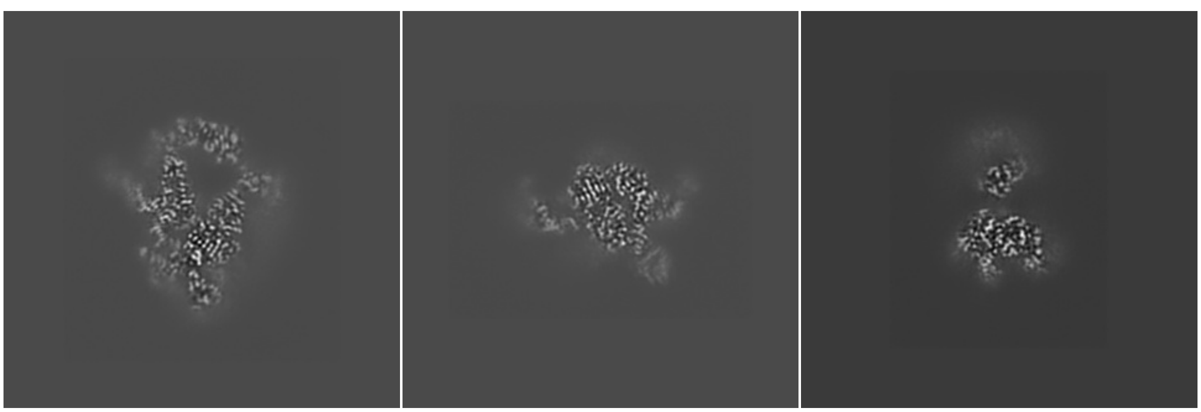


X Index: 160 **Y Index: 160** **Z Index: 160**

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

3.3.2.3. Largest variance slices ?

Primary map

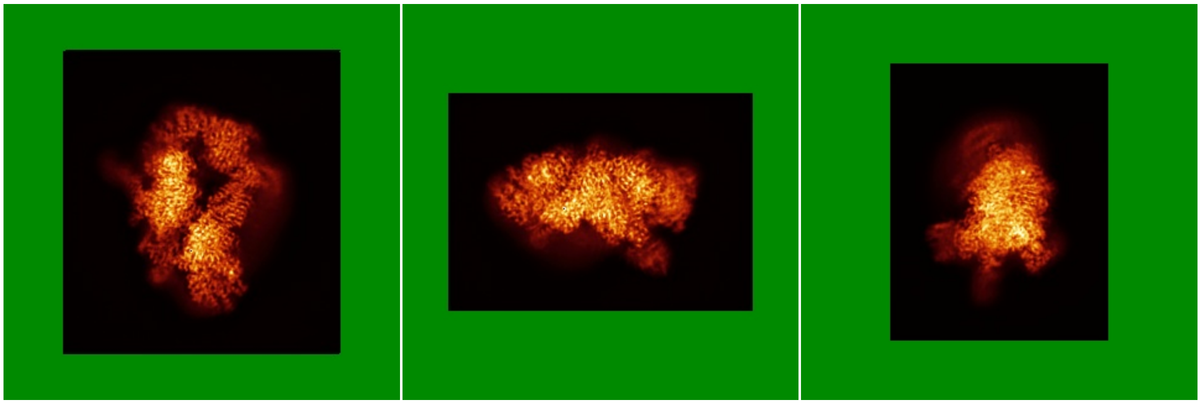


X Index: 153 **Y Index: 134** **Z Index: 166**

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



X **Y** **Z**

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

X **Y** **Z**

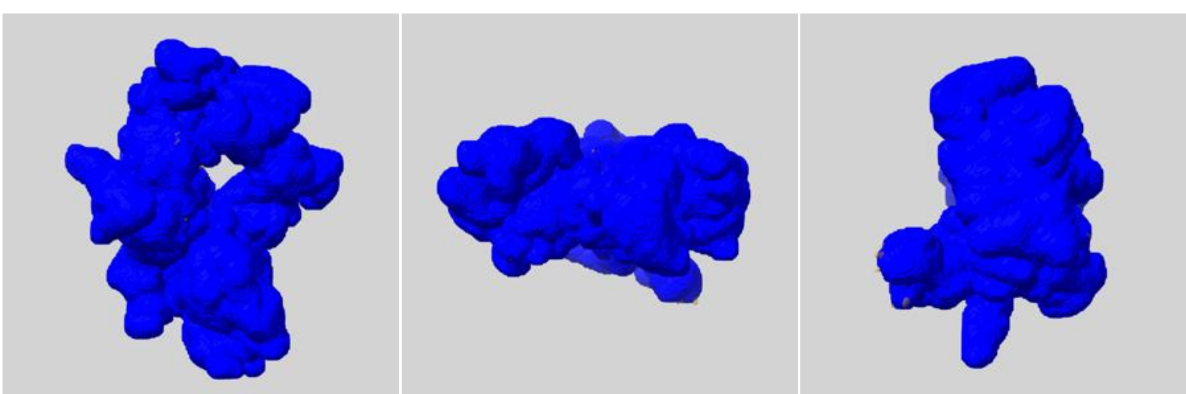
The images above show the 3D surface view of the map at the recommended contour level 0.015 . These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

3.3.2.6. Mask visualisation ?

This section shows the 3D surface view of the primary map at 50% transparency overlaid with the specified mask at 0% transparency. A mask typically either:

- Encompasses the whole structure;
- Separates out a domain, a functional unit, a monomer or an area of interest from a larger structure.

emd_4970_msk_1.map ?

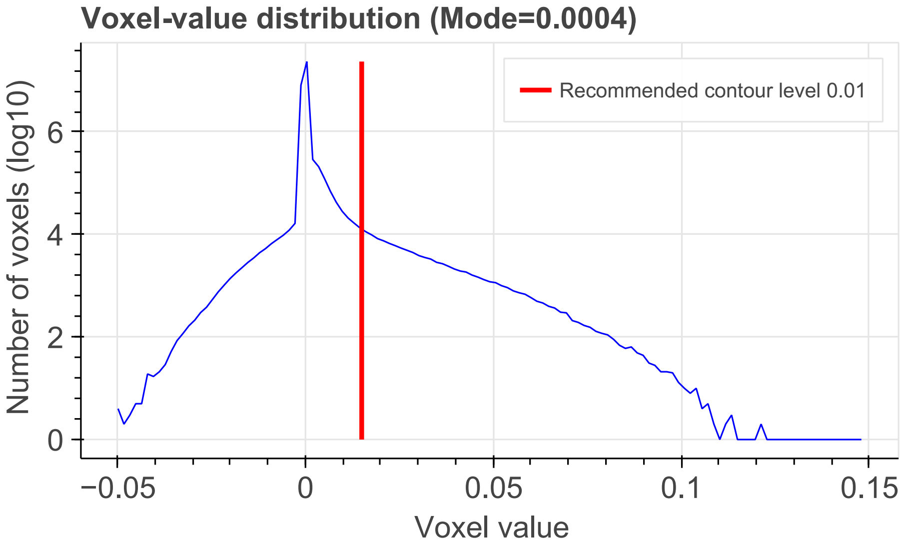


X **Y** **Z**

3.3.3. Map analysis

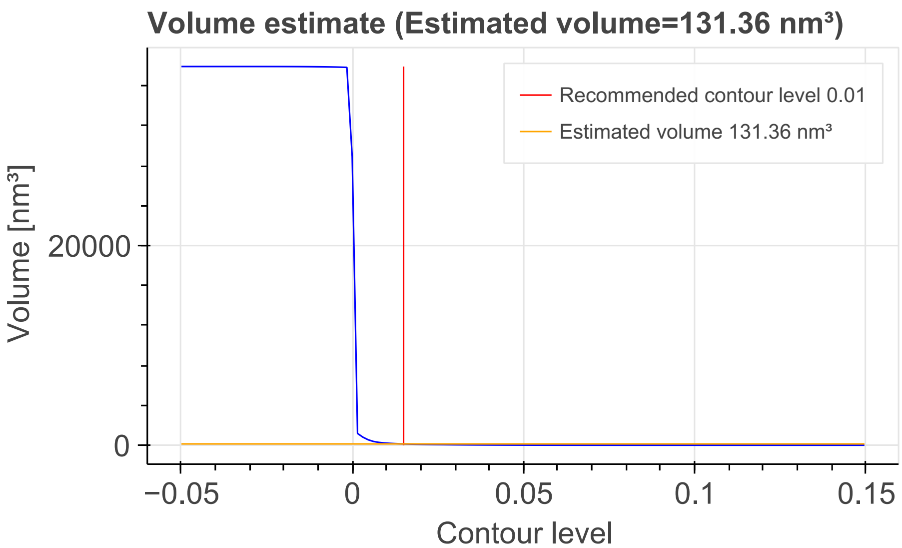
This section contains the results of statistical analysis of the map.

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.

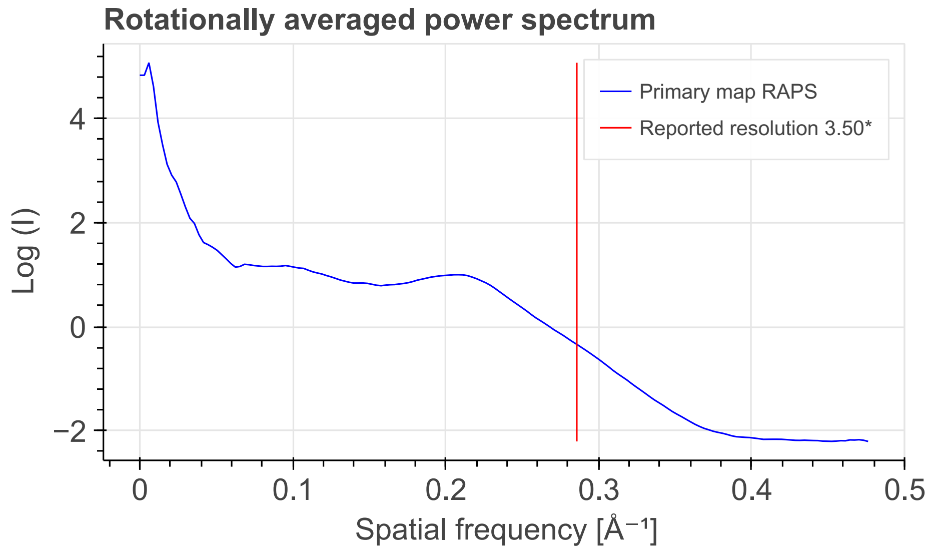
3.3.3.2. Volume estimate



The volume at the recommended contour level is 131.36 nm^3.

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 0.286 Å⁻¹

3.3.4. Fourier-Shell correlation

3.3.4.2. Resolution estimates

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	**0.5**	**Half-bit**
Reported by author	3.50	-	-

Author-provided FSC curve is not available.

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 5 bond length outliers in this entry (0.01% of 41849 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
I	198	PRO	CG-CD	12.87	1.07	1.50	1	1
I	198	PRO	N-CD	5.87	1.56	1.47	1	1
J	84	PRO	CG-CD	4.97	1.33	1.50	1	1
L	104	VAL	C-N	4.73	1.26	1.33	1	1
I	198	PRO	CB-CG	4.71	1.73	1.49	1	1

Standard geometry: angle outliers

There are 45 bond angle outliers in this entry (0.08% of 57165 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
P	50	DC	OP1-P-OP2	13.17	80.49	120.00	1	1
P	49	DG	O3'-P-OP1	11.75	72.75	108.00	1	1
I	198	PRO	N-CD-CG	11.75	85.58	103.20	1	1
I	198	PRO	CA-N-CD	11.60	95.76	112.00	1	1
A	511	TRP	C-N-CA	11.39	142.21	121.70	1	1
P	50	DC	OP2-P-O5'	10.65	76.05	108.00	1	1
P	49	DG	O3'-P-OP2	9.04	135.13	108.00	1	1
J	84	PRO	N-CD-CG	8.78	90.03	103.20	1	1
I	198	PRO	CA-CB-CG	8.06	89.19	104.50	1	1
P	50	DC	OP1-P-O5'	7.65	131.96	109.00	1	1
J	84	PRO	CA-N-CD	7.48	101.53	112.00	1	1
J	14	SER	C-N-CA	6.61	133.59	121.70	1	1
I	294	GLY	C-N-CA	6.56	133.52	121.70	1	1
J	158	THR	C-N-CA	6.24	132.94	121.70	1	1
N	68	ASP	C-N-CA	6.09	132.66	121.70	1	1
J	84	PRO	CA-CB-CG	5.92	93.25	104.50	1	1
H	20	TYR	C-N-CA	5.84	132.21	121.70	1	1
B	273	ILE	CA-CB-CG1	5.81	120.27	110.40	1	1
L	258	THR	C-N-CA	5.51	131.61	121.70	1	1
I	965	TYR	CA-CB-CG	5.50	123.79	113.90	1	1
P	49	DG	O3'-P-O5'	5.44	95.84	104.00	1	1
L	257	ASN	C-N-CA	5.31	131.26	121.70	1	1
J	84	PRO	CB-CG-CD	5.25	89.31	106.10	1	1
L	104	VAL	C-N-CA	5.02	130.74	121.70	1	1
B	272	ARG	C-N-CA	4.95	112.78	121.70	1	1
J	84	PRO	N-CA-CB	4.84	97.67	103.00	1	1
I	41	GLY	N-CA-C	4.80	127.23	113.30	1	1
B	273	ILE	CG1-CB-CG2	4.72	96.54	110.70	1	1
I	270	GLY	C-N-CA	4.72	130.19	121.70	1	1
J	85	SER	C-N-CA	4.66	130.09	121.70	1	1
I	198	PRO	N-CA-CB	4.63	97.90	103.00	1	1
B	223	LYS	CB-CG-CD	4.50	121.66	111.30	1	1
I	284	GLU	C-N-CA	4.50	129.79	121.70	1	1
O	21	DC	C3'-C2'-C1'	4.34	95.10	101.60	1	1
J	82	PHE	C-CA-CB	4.23	118.14	110.10	1	1

Chain	Res	Type	Atoms	|Z|	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
B	350	VAL	C-N-CA	4.23	129.31	121.70	1	1
H	150	LEU	CA-CB-CG	4.22	131.07	116.30	1	1
G	96	PRO	C-N-CA	4.21	129.27	121.70	1	1
B	273	ILE	N-CA-CB	4.19	118.63	111.50	1	1
G	90	VAL	C-N-CA	4.19	129.24	121.70	1	1
G	101	ASP	C-N-CA	4.17	129.21	121.70	1	1
C	82	GLN	CA-CB-CG	4.13	122.37	114.10	1	1
B	406	LEU	CA-CB-CG	4.11	130.68	116.30	1	1
H	22	PRO	CA-N-CD	4.10	106.26	112.00	1	1
G	127	ASP	C-N-CA	4.09	114.33	121.70	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	23.69	1901

There are 1901 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)
I:44:ASP:HB2	I:50:ILE:HG12	1.04	1	1
K:146:VAL:HB	K:150:ASP:HB2	1.00	1	1
A:35:GLN:HG2	A:36:GLU:HG2	0.98	1	1
K:137:THR:HB	K:151:SER:HB3	0.94	1	1
B:345:ARG:HH21	L:223:ILE:HD11	0.94	1	1
B:270:VAL:HG13	I:50:ILE:HD13	0.92	1	1
B:345:ARG:HD2	L:220:ASN:HA	0.92	1	1
B:273:ILE:N	I:44:ASP:O	0.91	1	1
G:218:LYS:HG3	K:148:LYS:HB3	0.91	1	1
B:386:LEU:HD13	I:48:ASN:HD21	0.91	1	1
D:248:THR:H	D:279:TYR:HB2	0.91	1	1
I:101:ASP:HB2	I:757:GLN:HE22	0.91	1	1
I:292:ILE:HG23	K:48:ASN:HD21	0.90	1	1
B:121:VAL:HG23	B:133:LYS:HB3	0.90	1	1
B:343:ARG:HH21	L:251:GLY:HA2	0.89	1	1
B:269:THR:HB	I:45:ARG:HB2	0.89	1	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
D:281:CYS:HB2	D:302:LEU:HD11	0.89	1	1
I:44:ASP:OD2	I:48:ASN:ND2	0.89	1	1
B:273:ILE:HD12	I:44:ASP:N	0.88	1	1
A:584:THR:HA	K:185:ARG:HG2	0.87	1	1
B:369:ARG:HB3	L:257:ASN:H	0.87	1	1
B:253:ARG:O	B:257:ASP:HB2	0.86	1	1
L:253:VAL:HG22	L:272:GLY:HA2	0.86	1	1
A:568:LEU:HD11	A:606:PHE:HB3	0.85	1	1
B:278:GLU:HB2	I:43:ARG:HE	0.85	1	1
B:304:HIS:NE2	P:45:DT:O2	0.85	1	1
J:26:ASP:HA	J:217:GLN:HG2	0.85	1	1
I:39:LEU:O	I:43:ARG:NH2	0.85	1	1
B:382:LEU:O	B:386:LEU:N	0.85	1	1
N:37:ILE:HB	N:80:GLU:H	0.84	1	1
N:11:LEU:HD21	N:31:VAL:HG21	0.83	1	1
G:225:GLU:HB3	K:152:GLN:HG3	0.83	1	1
D:246:GLN:HE21	E:251:TYR:HB3	0.83	1	1
B:273:ILE:HG13	I:45:ARG:C	0.82	1	1
I:279:ARG:HG3	I:280:ARG:HG2	0.82	1	1
B:270:VAL:HA	I:44:ASP:HA	0.82	1	1
B:503:ALA:O	J:185:ARG:NH1	0.82	1	1
B:266:LEU:O	B:269:THR:OG1	0.82	1	1
I:292:ILE:HD12	L:236:GLU:HA	0.82	1	1
L:397:GLU:OE1	O:31:DA:N6	0.81	1	1
B:349:VAL:HG13	B:351:GLN:HB3	0.81	1	1
G:170:PRO:HB3	L:400:ASN:HB3	0.80	1	1
I:46:HIS:N	I:749:SER:OG	0.80	1	1
B:388:ILE:HD13	I:744:LEU:HD21	0.80	1	1
L:279:TYR:HB2	L:391:ARG:HH21	0.80	1	1
I:292:ILE:HG23	K:48:ASN:ND2	0.79	1	1
L:20:ARG:HG3	L:72:VAL:HG22	0.79	1	1
I:294:GLY:HA2	K:48:ASN:HD22	0.79	1	1
J:14:SER:HA	P:30:DA:H4'	0.79	1	1
K:136:VAL:HG12	K:151:SER:HA	0.79	1	1
C:84:LEU:HD12	C:86:GLY:H	0.79	1	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
D:277:GLY:HA2	D:285:ARG:HH22	0.79	1	1
B:273:ILE:O	B:275:GLU:N	0.79	1	1
B:280:ARG:NH1	I:753:GLN:O	0.79	1	1
G:103:VAL:HB	G:113:MET:HA	0.78	1	1
J:10:ARG:HG2	J:151:ALA:HB3	0.78	1	1
B:324:ARG:O	B:378:ARG:NH2	0.78	1	1
D:340:GLY:HA3	E:146:ARG:HH12	0.78	1	1
B:266:LEU:O	B:270:VAL:HG23	0.78	1	1
I:200:VAL:O	I:204:ILE:HG13	0.77	1	1
B:485:LEU:HD12	B:486:ALA:H	0.77	1	1
B:515:ALA:HB3	J:185:ARG:HG3	0.77	1	1
B:280:ARG:HG2	I:756:GLN:CB	0.77	1	1
B:176:ASN:OD1	B:177:LEU:N	0.77	1	1
G:237:ARG:HG2	G:240:ILE:H	0.77	1	1
J:9:MET:HE1	K:132:PHE:HA	0.76	1	1
A:449:LYS:NZ	O:47:DG:OP2	0.76	1	1
A:583:PRO:HD3	P:21:DG:H3'	0.76	1	1
B:501:GLN:OE1	K:194:PHE:HB3	0.76	1	1
J:14:SER:HB2	J:16:LEU:H	0.76	1	1
J:165:TYR:HB2	J:167:PRO:HD2	0.76	1	1
B:355:PRO:O	L:273:GLN:NE2	0.76	1	1
B:253:ARG:HE	L:222:ASP:HB3	0.76	1	1
G:244:HIS:ND1	G:244:HIS:O	0.75	1	1
G:113:MET:HG2	G:114:ASP:H	0.75	1	1
G:130:ARG:HB3	L:290:ARG:HH22	0.75	1	1
B:324:ARG:HA	B:378:ARG:HH12	0.75	1	1
B:269:THR:O	B:272:ARG:N	0.75	1	1
H:55:ILE:HD12	H:57:GLN:H	0.75	1	1
I:1:MET:HG2	I:94:ARG:HH22	0.75	1	1
J:25:ILE:O	J:217:GLN:NE2	0.75	1	1
G:30:ARG:HG2	G:31:LYS:HG3	0.75	1	1
K:136:VAL:HG13	K:154:LEU:HD12	0.75	1	1
P:46:DG:N1	Q:1:SF4:S1	0.74	1	1
L:212:LEU:HB3	O:21:DC:C4	0.74	1	1
C:232:TYR:HB2	C:263:ARG:HD2	0.73	1	1

Atom 1	Atom 2	Clash(Å)	Model ID (Worst)	Models (Total)
I:116:ARG:HH22	I:740:GLU:HG3	0.73	1	1
B:497:ARG:NH2	B:707:ASN:O	0.73	1	1
E:6:ASP:HB2	E:155:GLN:HB2	0.73	1	1
B:365:VAL:HG12	B:366:CYS:H	0.73	1	1
B:370:LYS:HE3	I:964:ARG:HH22	0.73	1	1
B:282:ARG:NH1	I:35:LEU:O	0.73	1	1
G:16:GLN:HA	M:1:ASP:H2	0.73	1	1
I:116:ARG:HD2	I:747:GLN:HE22	0.72	1	1
I:962:CYS:HB3	I:968:TRP:HB2	0.72	1	1
G:140:THR:HG21	O:35:DT:H5'	0.72	1	1
B:269:THR:HG22	I:45:ARG:HH11	0.72	1	1
J:77:VAL:HB	J:120:ARG:HB2	0.72	1	1
B:360:GLY:HA2	L:276:LYS:HD2	0.72	1	1
A:580:ILE:HD11	A:589:ARG:HH11	0.72	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	4722	4111	555	56

There are 56 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	71	HIS	1
A	110	PRO	1
A	150	ASP	1
A	199	LEU	1
B	274	LYS	1
B	293	ARG	1
B	296	SER	1
B	300	GLU	1
B	350	VAL	1
B	365	VAL	1
C	32	PRO	1
C	175	GLU	1
C	281	GLY	1

Chain	Res	Type	Models (Total)
D	36	VAL	1
D	37	PHE	1
D	65	ASP	1
D	68	PRO	1
D	230	PRO	1
D	264	ALA	1
E	92	ASN	1
G	8	LEU	1
G	20	LEU	1
G	21	PRO	1
G	91	VAL	1
G	97	VAL	1
G	118	MET	1
G	128	ASN	1
G	129	CYS	1
G	237	ARG	1
H	3	SER	1
H	54	ALA	1
H	55	ILE	1
I	44	ASP	1
I	49	SER	1
I	130	ALA	1
I	279	ARG	1
I	285	ASP	1
I	293	LYS	1
J	6	VAL	1
J	7	VAL	1
J	8	ASP	1
J	84	PRO	1
J	86	LYS	1
J	87	PRO	1
J	119	LEU	1
J	159	LEU	1
J	161	GLU	1
K	21	PRO	1

Chain	Res	Type	Models (Total)
K	147	ASN	1
L	221	PRO	1
L	258	THR	1
L	259	ASN	1
L	270	ASN	1
L	322	ASP	1
L	343	GLU	1
N	43	ALA	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	4194	3055	1114	25

There are 25 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	35	GLN	1
A	64	GLN	1
A	65	MET	1
A	187	HIS	1
A	431	LYS	1
B	112	ARG	1
B	207	TYR	1
B	280	ARG	1
B	629	GLN	1
D	91	ILE	1
D	294	GLU	1
F	48	LEU	1
G	37	MET	1
G	136	HIS	1
G	219	PHE	1
H	150	LEU	1
I	94	ARG	1
I	119	ILE	1

Chain	Res	Type	Models (Total)
I	244	ASN	1
I	250	VAL	1
J	23	ARG	1
J	69	MET	1
K	15	LYS	1
M	40	ILE	1
M	83	VAL	1

5. Fit to Data Used for Modeling Assessment ?

5.2. Crosslinking-MS ?

5.2.1. Restraint types ?

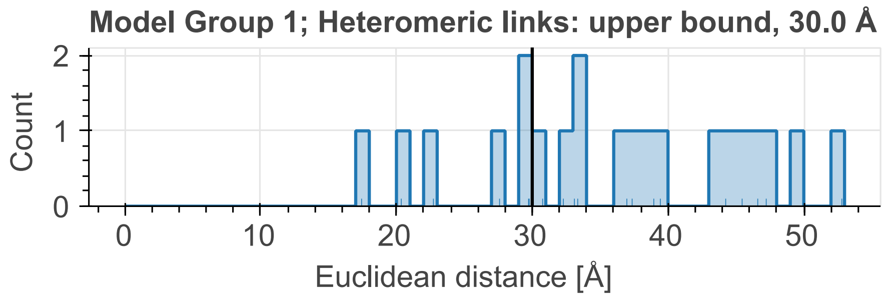
This table summarizes information about crosslinker(s) used for data generation, and how crosslinking information was translated into actual modeling restraints. Restraints assigned "by-residue" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "coarse-grained". Restraint group represents a set of crosslinking restraints applied collectively in the modeling.

There are 21 crosslinking restraints combined in 21 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
BS3	ALA	CA	GLY	CA	upper bound	30.00	1
BS3	ALA	CA	LEU	CA	upper bound	30.00	2
BS3	ASP	CA	GLU	CA	upper bound	30.00	2
BS3	GLU	CA	LEU	CA	upper bound	30.00	4
BS3	ASP	CA	PHE	CA	upper bound	30.00	1
BS3	GLY	CA	PHE	CA	upper bound	30.00	1
BS3	ARG	CA	PHE	CA	upper bound	30.00	1
BS3	LEU	CA	PHE	CA	upper bound	30.00	1
BS3	ARG	CA	GLU	CA	upper bound	30.00	1
BS3	LEU	CA	SER	CA	upper bound	30.00	1
BS3	ASP	CA	GLN	CA	upper bound	30.00	1
BS3	ASP	CA	LEU	CA	upper bound	30.00	1
BS3	LEU	CA	LEU	CA	upper bound	30.00	2
BS3	ARG	CA	ASP	CA	upper bound	30.00	1
BS3	GLN	CA	LEU	CA	upper bound	30.00	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.



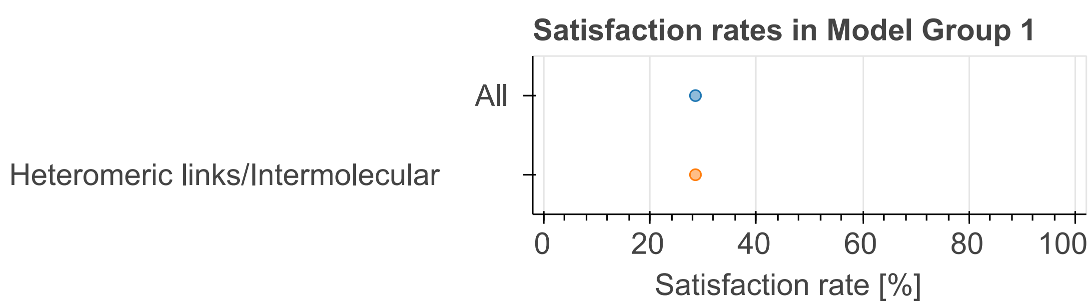
5.2.2. Satisfaction of restraints

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=21)
1	1	1	1/1	All	28.57	71.43	21
				Heteromeric links/ Intermolecular	28.57	71.43	21

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

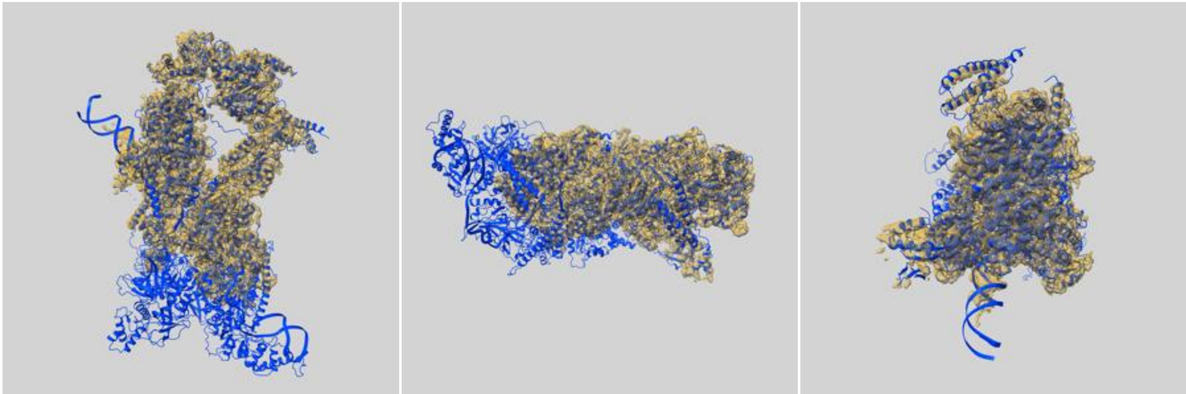
This section describes fit of models to the 3DEM data. Only results for the representative model, selected as a first model with the largest number of asymmetric units.

EMD-4970

5.3.1. Map-model fit

Only results for the representative Model 1 are shown.

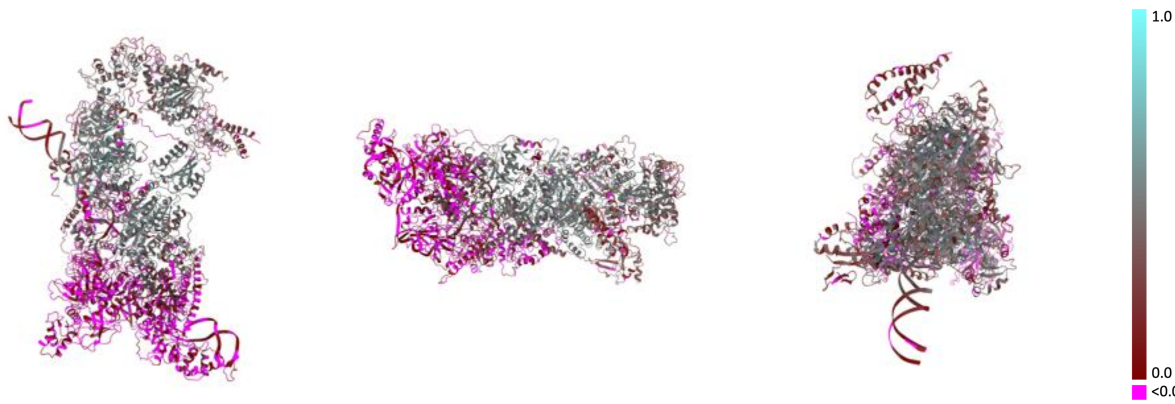
5.3.1.1 Map-model overlay ?



X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.015 at 50% transparency in yellow overlaid with a ribbon representation of the model colored in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

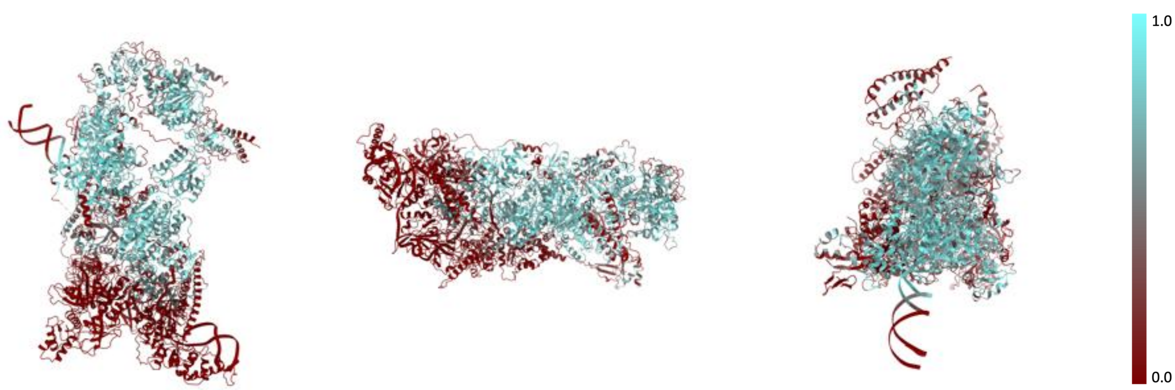
5.3.1.2. Q-score mapped to coordinate model ?



X Y Z

The images above show the model with each residue colored according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

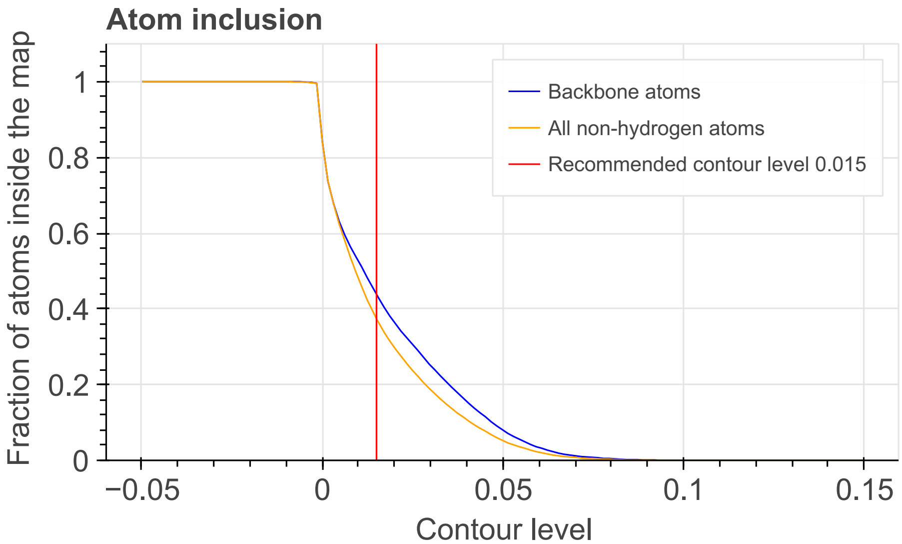
5.3.1.3. Atom inclusion mapped to coordinate model ?



X Y Z

The images above show the model with each residue colored according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level 0.015 .

5.3.1.4. Atom inclusion ?



At the recommended contour level, 44% of all backbone atoms, 37% of all non-hydrogen atoms, are inside the map.

5.3.1.5. Map-model fit summary

The table lists the average atom inclusion at the recommended contour level (0.015) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.374	0.230
A	0.726	0.449
B	0.655	0.388
C	0.616	0.382
D	0.573	0.371
E	0.601	0.393
F	0.657	0.416
G	0.210	0.159
H	0.296	0.215
I	0.052	0.034
J	0.009	-0.025
K	0.005	-0.001
L	0.006	0.004
M	0.000	0.013
N	0.000	-0.017
O	0.268	0.144
P	0.291	0.133
Q	1.000	0.347
R	1.000	0.342
S	1.000	0.570
T	1.000	0.507

Chain	Atom inclusion	Q-score
U	1.000	0.553
V	1.000	0.527
W	1.000	0.117
X	0.000	-0.075
Y	0.000	-0.082
Z	0.000	0.124

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.