wwPDB EM Validation Summary Report

May 6, 2023 – 05:41 PM EDT

PDB ID	:	8EBY
EMDB ID	:	EMD-28002
Title	:	XPC release from Core7-XPA-DNA (AP)
Authors	:	Kim, J.; Yang, W.
Deposited on	:	2022-08-31
Resolution	:	3.60 Å(reported)
Based on initial model	:	6RO4

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

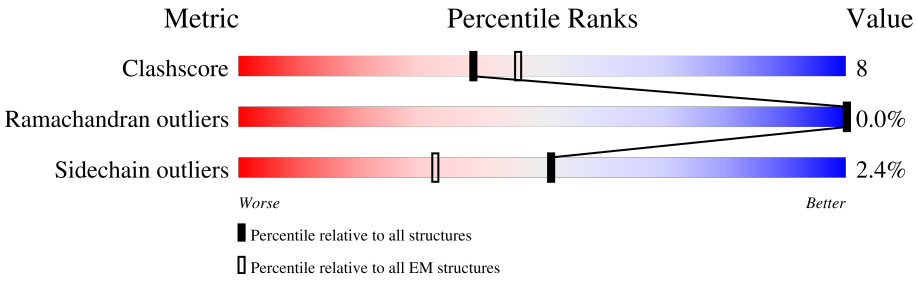
EMDB validation analysis	:	0.0.1.dev50
Mogul	:	1.8.5 (274361), CSD as541be (2020)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.60 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



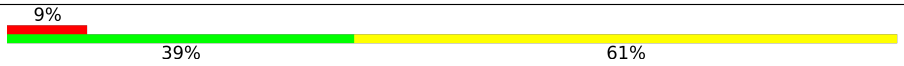
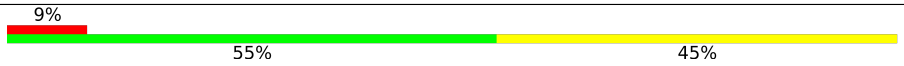
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the EM map (all-atom inclusion < 40%). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	782	56% 21% 23%
2	B	768	64% 24% 12%
3	C	548	6% 29% 5% 65%
4	D	462	78% 15% 6%
5	E	417	76% 14% 9%
6	F	308	69% 14% 16%
7	G	71	79% 14% 7%
8	H	950	95%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	K	273	
10	L	23	
11	M	22	

2 Entry composition (i)

There are 13 unique types of molecules in this entry. The entry contains 23709 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called TFIIH basal transcription factor complex helicase XPB subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	604	Total	C	N	O	S	0	0
			4882	3119	845	889	29		

- Molecule 2 is a protein called General transcription and DNA repair factor IIH helicase subunit XPD.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	677	Total	C	N	O	S	0	0
			5457	3502	948	978	29		

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	761	ASP	-	expression tag	UNP P18074
B	762	TYR	-	expression tag	UNP P18074
B	763	LYS	-	expression tag	UNP P18074
B	764	ASP	-	expression tag	UNP P18074
B	765	ASP	-	expression tag	UNP P18074
B	766	ASP	-	expression tag	UNP P18074
B	767	ASP	-	expression tag	UNP P18074
B	768	LYS	-	expression tag	UNP P18074

- Molecule 3 is a protein called General transcription factor IIH subunit 1.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	191	Total	C	N	O	S	0	0
			1547	981	267	290	9		

- Molecule 4 is a protein called General transcription factor IIH subunit 4, p52.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	434	Total	C	N	O	S	0	0
			3483	2245	605	620	13		

- Molecule 5 is a protein called General transcription factor IIH subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	380	Total	C	N	O	S	0	0
			2989	1886	518	558	27		

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	-21	MET	-	initiating methionine	UNP Q13888
E	-20	GLY	-	expression tag	UNP Q13888
E	-19	SER	-	expression tag	UNP Q13888
E	-18	SER	-	expression tag	UNP Q13888
E	-17	HIS	-	expression tag	UNP Q13888
E	-16	HIS	-	expression tag	UNP Q13888
E	-15	HIS	-	expression tag	UNP Q13888
E	-14	HIS	-	expression tag	UNP Q13888
E	-13	HIS	-	expression tag	UNP Q13888
E	-12	HIS	-	expression tag	UNP Q13888
E	-11	SER	-	expression tag	UNP Q13888
E	-10	SER	-	expression tag	UNP Q13888
E	-9	GLY	-	expression tag	UNP Q13888
E	-8	LEU	-	expression tag	UNP Q13888
E	-7	GLU	-	expression tag	UNP Q13888
E	-6	VAL	-	expression tag	UNP Q13888
E	-5	LEU	-	expression tag	UNP Q13888
E	-4	PHE	-	expression tag	UNP Q13888
E	-3	GLN	-	expression tag	UNP Q13888
E	-2	GLY	-	expression tag	UNP Q13888
E	-1	PRO	-	expression tag	UNP Q13888
E	0	HIS	-	expression tag	UNP Q13888

- Molecule 6 is a protein called General transcription factor IIH subunit 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	258	Total	C	N	O	S	0	0
			2034	1301	337	377	19		

- Molecule 7 is a protein called General transcription factor IIH subunit 5.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	66	Total	C	N	O	S	0	0
			522	337	82	100	3		

- Molecule 8 is a protein called DNA repair protein complementing XP-C cells.

Mol	Chain	Residues	Atoms				AltConf	Trace
8	H	52	Total	C	N	O	0	0
			409	255	77	77		

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	-9	MET	-	initiating methionine	UNP A0A024R2M8
H	-8	ASP	-	expression tag	UNP A0A024R2M8
H	-7	TYR	-	expression tag	UNP A0A024R2M8
H	-6	LYS	-	expression tag	UNP A0A024R2M8
H	-5	ASP	-	expression tag	UNP A0A024R2M8
H	-4	ASP	-	expression tag	UNP A0A024R2M8
H	-3	ASP	-	expression tag	UNP A0A024R2M8
H	-2	ASP	-	expression tag	UNP A0A024R2M8
H	-1	LYS	-	expression tag	UNP A0A024R2M8
H	0	HIS	-	expression tag	UNP A0A024R2M8
H	499	VAL	ALA	conflict	UNP A0A024R2M8

- Molecule 9 is a protein called DNA repair protein complementing XP-A cells.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	K	172	Total	C	N	O	S	0	0
			1450	908	255	273	14		

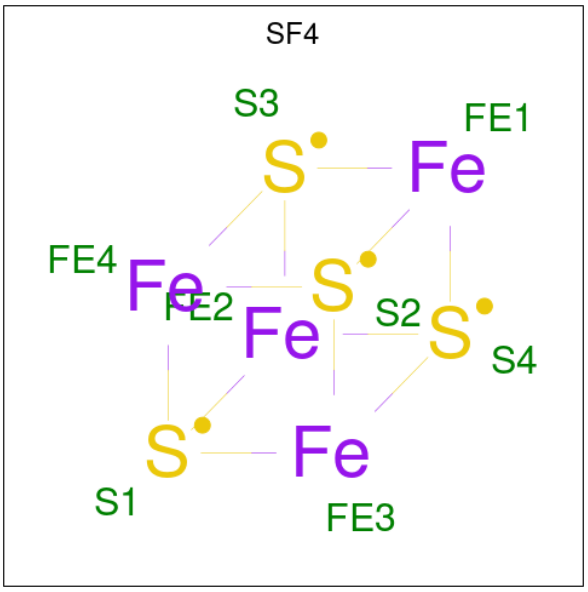
- Molecule 10 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	L	23	Total	C	N	O	P	0	0
			474	225	87	139	23		

- Molecule 11 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	M	22	Total	C	N	O	P	0	0
			448	213	81	132	22		

- Molecule 12 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).



Mol	Chain	Residues	Atoms	AltConf
12	B	1	Total Fe S 8 4 4	0

- Molecule 13 is ZINC ION (three-letter code: ZN) (formula: Zn).

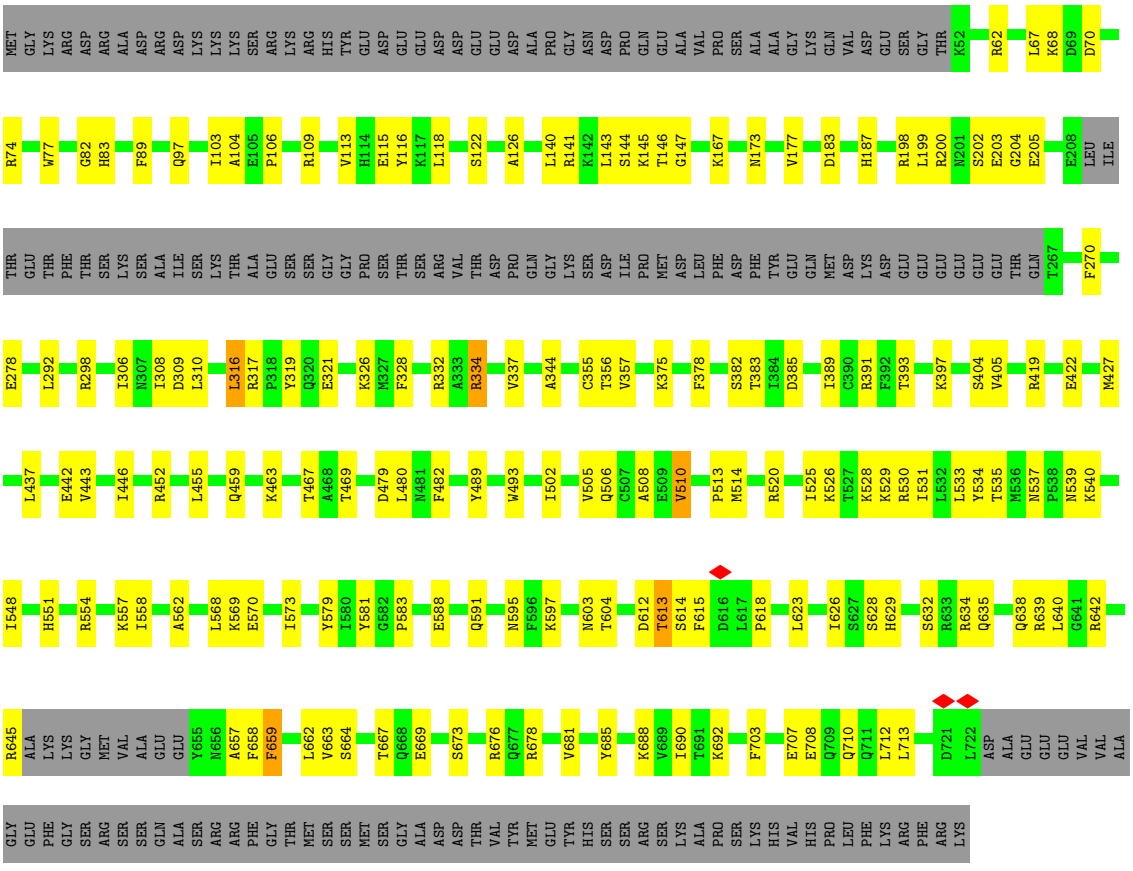
Mol	Chain	Residues	Atoms	AltConf
13	E	3	Total Zn 3 3	0
13	F	2	Total Zn 2 2	0
13	K	1	Total Zn 1 1	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion in map density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

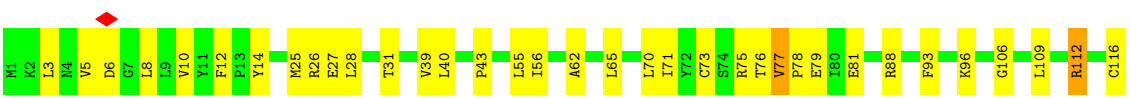
• Molecule 1: TFIIH basal transcription factor complex helicase XPB subunit

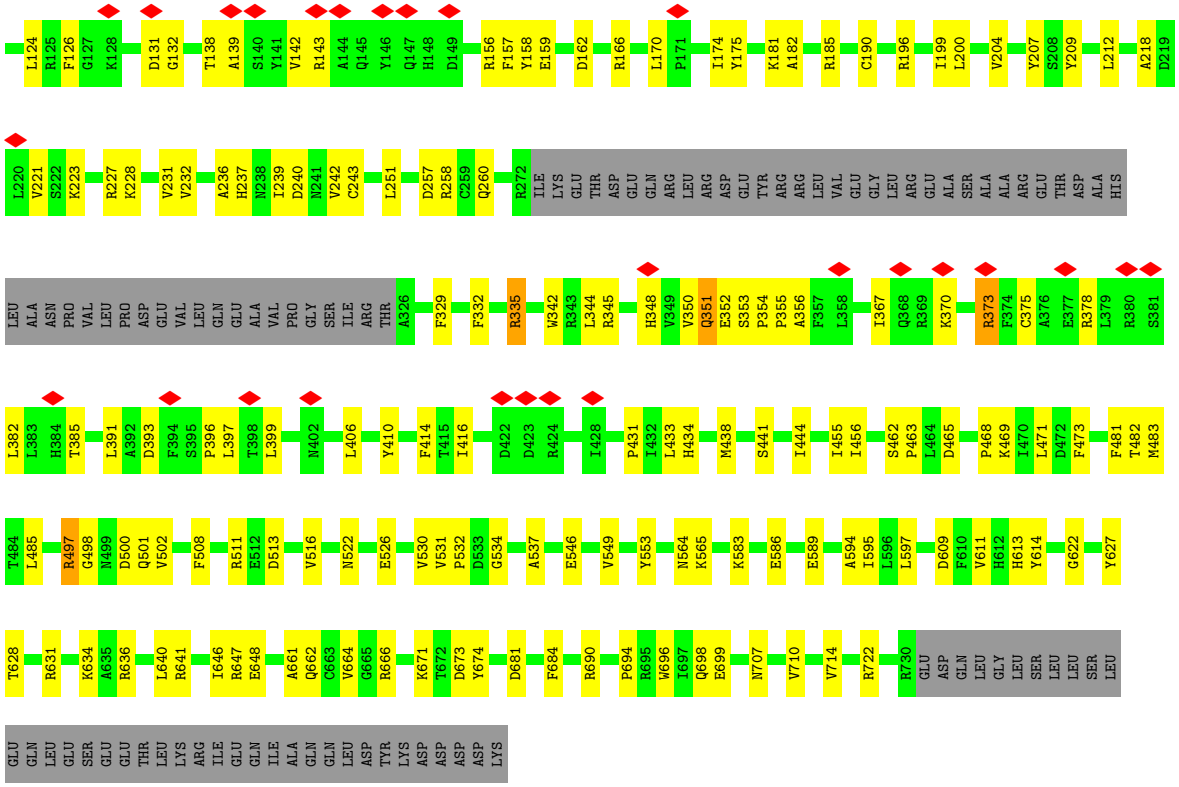
Chain A:



• Molecule 2: General transcription and DNA repair factor IIH helicase subunit XPD

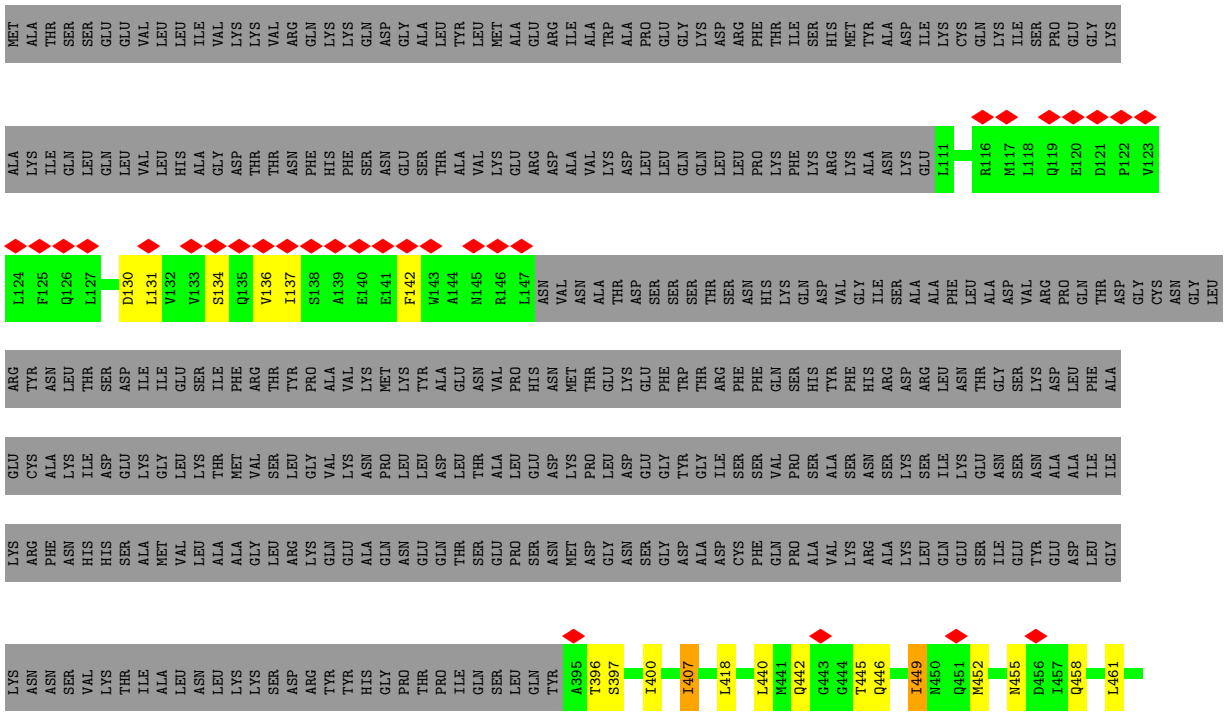
Chain B:

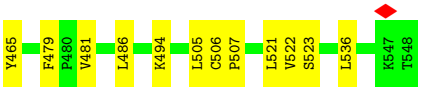




• Molecule 3: General transcription factor IIH subunit 1

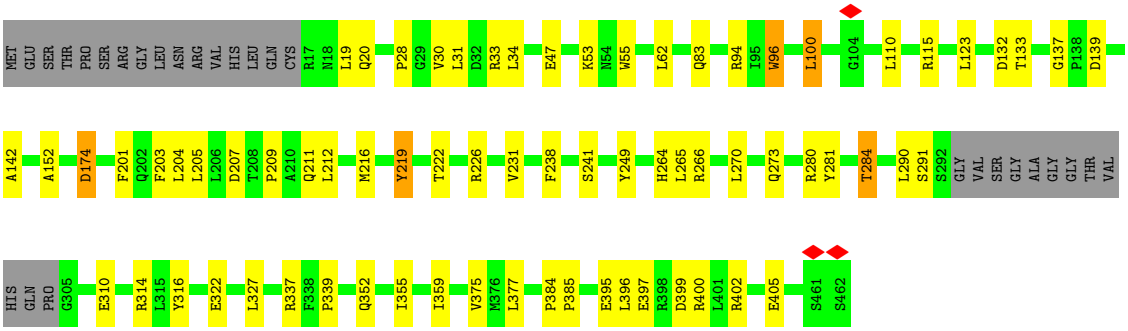
Chain C:





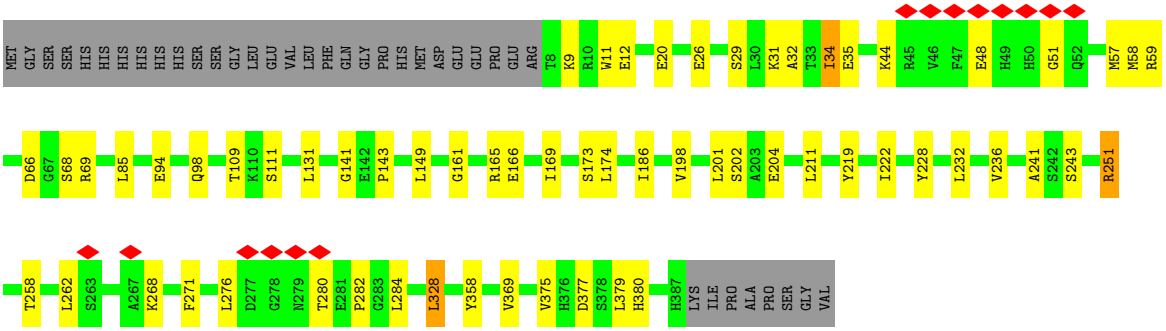
• Molecule 4: General transcription factor IIH subunit 4, p52

Chain D:



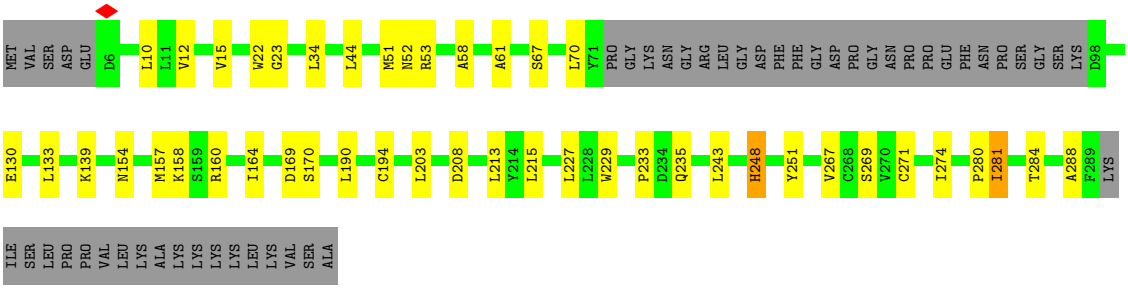
• Molecule 5: General transcription factor IIH subunit 2

Chain E:



• Molecule 6: General transcription factor IIH subunit 3

Chain F:

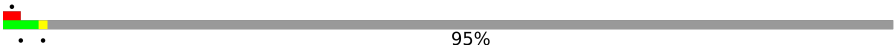


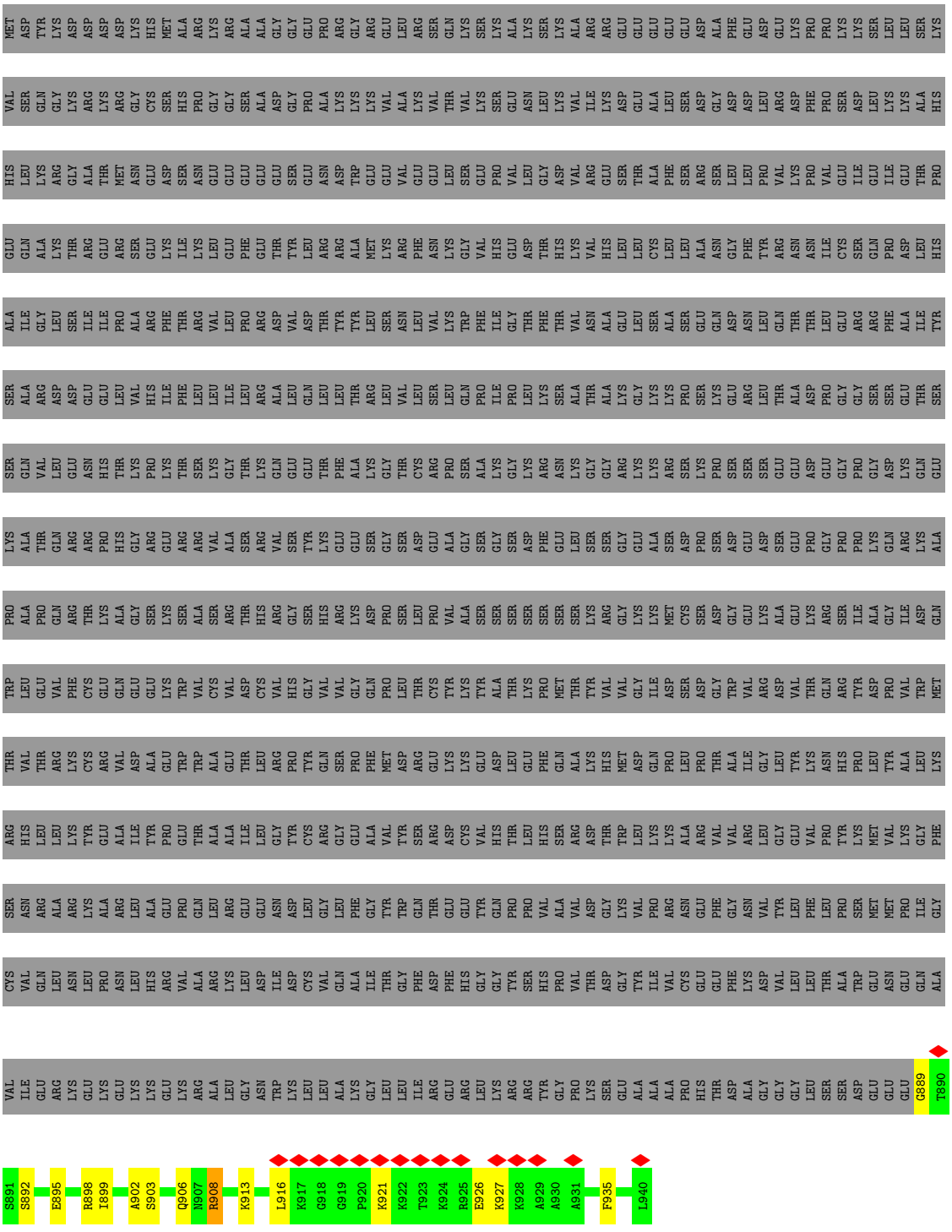
• Molecule 7: General transcription factor IIH subunit 5

Chain G:

• Molecule 8: DNA repair protein complementing XP-C cells

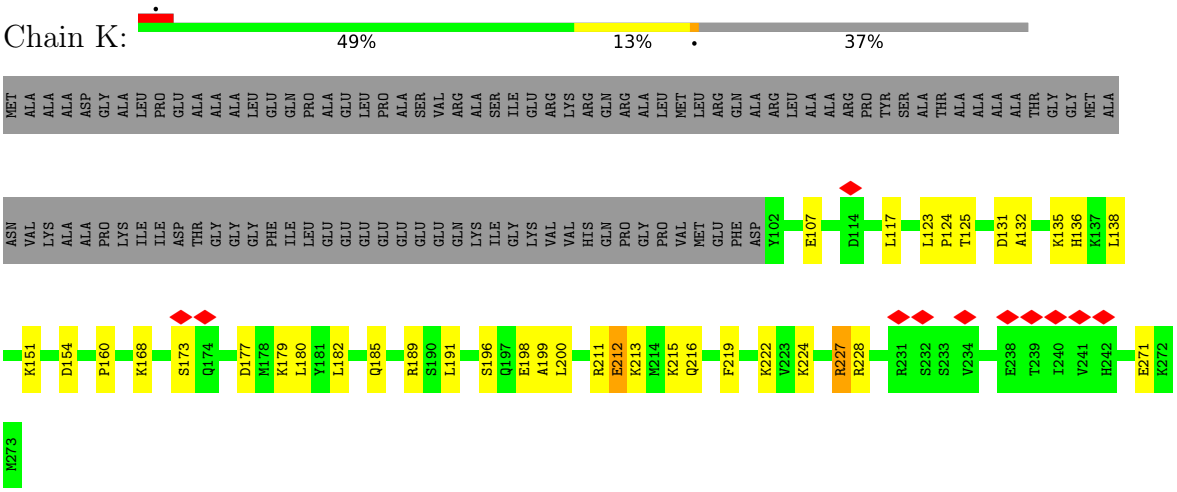
Chain H:





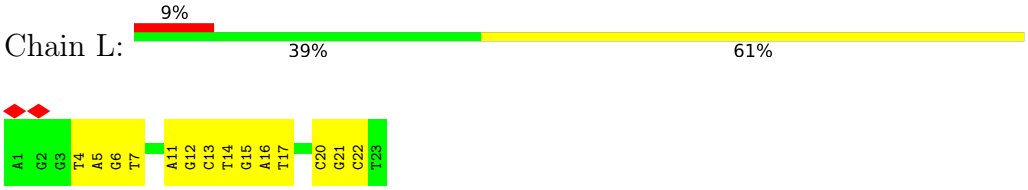
• Molecule 9: DNA repair protein complementing XP-A cells

Chain K:



• Molecule 10: DNA

Chain L:



• Molecule 11: DNA

Chain M:

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	137438	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	54.1	Depositor
Minimum defocus (nm)	2000	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	105000	Depositor
Image detector	GATAN K3 (6k x 4k)	Depositor
Maximum map value	1.831	Depositor
Minimum map value	-0.034	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.022	Depositor
Recommended contour level	0.01	Depositor
Map size (Å)	373.5, 373.5, 373.5	wwPDB
Map dimensions	300, 300, 300	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.245, 1.245, 1.245	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, SF4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.24	0/4985	0.41	0/6734
2	B	0.24	0/5578	0.42	0/7551
3	C	0.23	0/1575	0.36	0/2125
4	D	0.24	0/3564	0.37	0/4827
5	E	0.24	0/3057	0.40	0/4136
6	F	0.24	0/2070	0.37	0/2803
7	G	0.24	0/528	0.40	0/713
8	H	0.24	0/416	0.39	0/552
9	K	0.23	0/1478	0.40	0/1974
10	L	0.50	0/531	0.90	0/818
11	M	0.48	0/501	0.87	0/770
All	All	0.26	0/24283	0.43	0/33003

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4882	0	4923	103	0
2	B	5457	0	5499	111	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	1547	0	1561	17	0
4	D	3483	0	3526	46	0
5	E	2989	0	2948	38	0
6	F	2034	0	2050	27	0
7	G	522	0	531	9	0
8	H	409	0	419	14	0
9	K	1450	0	1434	24	0
10	L	474	0	260	12	0
11	M	448	0	248	8	0
12	B	8	0	0	0	0
13	E	3	0	0	0	0
13	F	2	0	0	0	0
13	K	1	0	0	0	0
All	All	23709	0	23399	372	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 8.

The worst 5 of 372 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
8:H:889:GLY:N	8:H:892:SER:HG	1.58	0.98
1:A:514:MET:HA	1:A:537:ASN:HD21	1.49	0.77
2:B:353:SER:HB2	2:B:356:ALA:HB3	1.69	0.74
2:B:647:ARG:HH21	9:K:179:LYS:HE3	1.53	0.72
2:B:232:VAL:HG12	2:B:455:ILE:HB	1.72	0.71

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	598/782 (76%)	556 (93%)	42 (7%)	0	100	100
2	B	673/768 (88%)	645 (96%)	28 (4%)	0	100	100
3	C	187/548 (34%)	180 (96%)	7 (4%)	0	100	100
4	D	430/462 (93%)	425 (99%)	5 (1%)	0	100	100
5	E	378/417 (91%)	366 (97%)	12 (3%)	0	100	100
6	F	254/308 (82%)	246 (97%)	8 (3%)	0	100	100
7	G	64/71 (90%)	58 (91%)	6 (9%)	0	100	100
8	H	50/950 (5%)	42 (84%)	8 (16%)	0	100	100
9	K	170/273 (62%)	164 (96%)	5 (3%)	1 (1%)	25	64
All	All	2804/4579 (61%)	2682 (96%)	121 (4%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
9	K	199	ALA

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	536/688 (78%)	522 (97%)	14 (3%)	46	74
2	B	591/672 (88%)	575 (97%)	16 (3%)	44	73
3	C	175/484 (36%)	168 (96%)	7 (4%)	31	65
4	D	377/399 (94%)	370 (98%)	7 (2%)	57	80
5	E	338/371 (91%)	330 (98%)	8 (2%)	49	75
6	F	230/272 (85%)	227 (99%)	3 (1%)	69	86
7	G	59/64 (92%)	59 (100%)	0	100	100
8	H	41/815 (5%)	39 (95%)	2 (5%)	25	59
9	K	162/233 (70%)	159 (98%)	3 (2%)	57	80
All	All	2509/3998 (63%)	2449 (98%)	60 (2%)	51	75

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	613	HIS
8	H	908	ARG
3	C	465	TYR
6	F	281	ILE
9	K	228	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	595	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 7 ligands modelled in this entry, 6 are monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
12	SF4	B	1000	2	0,12,12	-	-	-		

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
12	SF4	B	1000	2	-	-	0/6/5/5

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Map visualisation

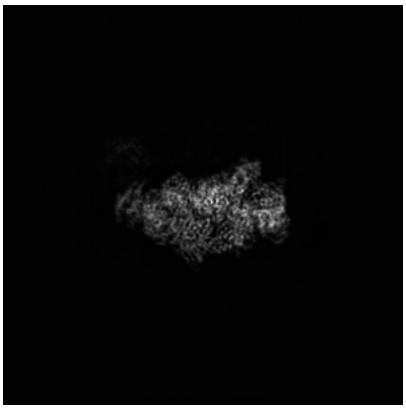
This section contains visualisations of the EMDB entry EMD-28002. These allow visual inspection of the internal detail of the map and identification of artifacts.

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.

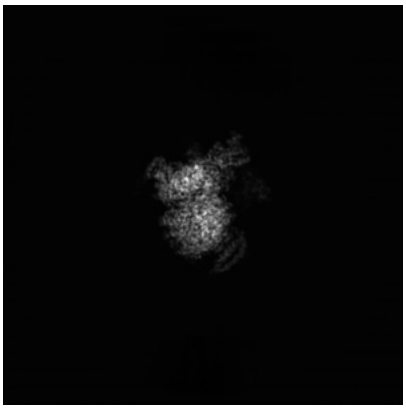
6.1 Orthogonal projections

6.1.1 Primary map

X

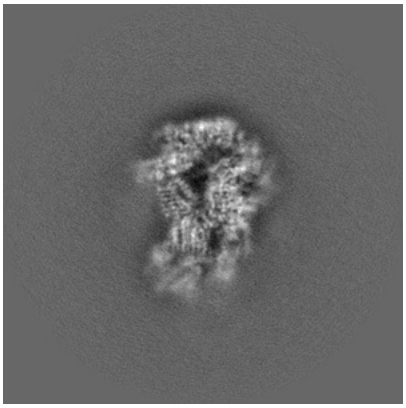


Y

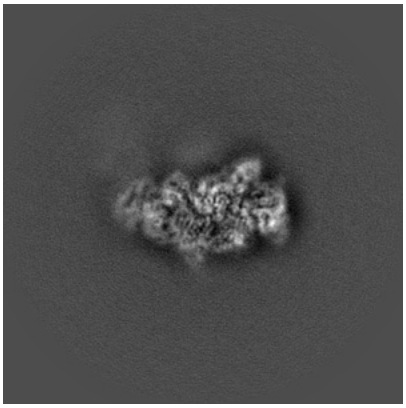


Z

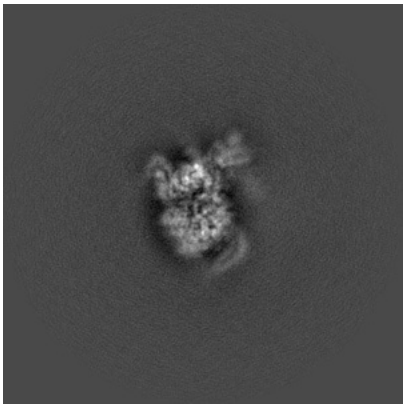
6.1.2 Raw map



X



Y



Z

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

6.2 Central slices (i)

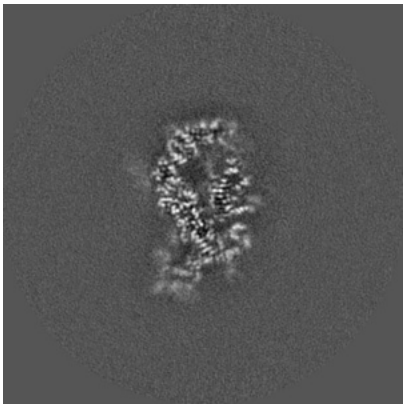
6.2.1 Primary map

X Index: 150

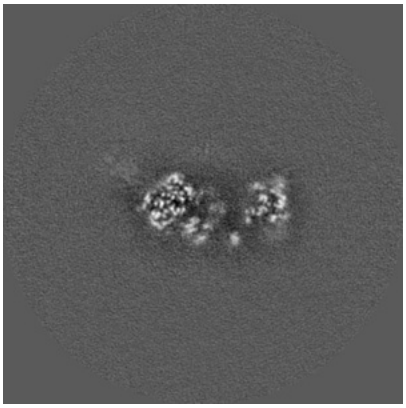
Y Index: 150

Z Index: 150

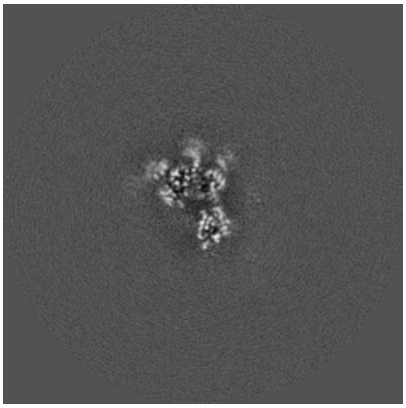
6.2.2 Raw map



X Index: 150



Y Index: 150



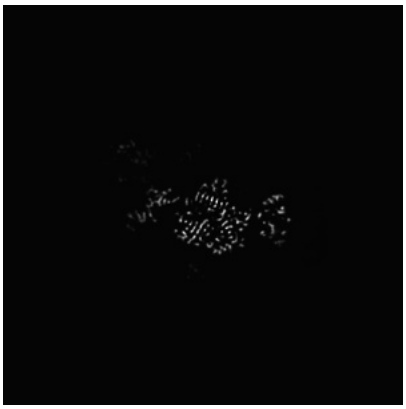
Z Index: 150

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

6.3 Largest variance slices (i)

6.3.1 Primary map

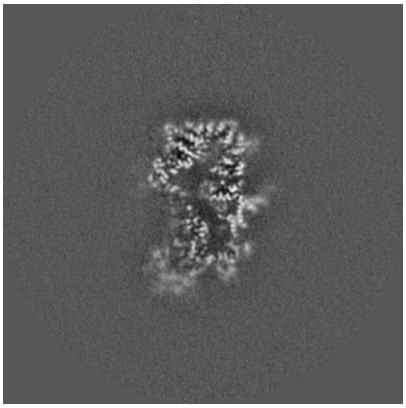
X Index: 145



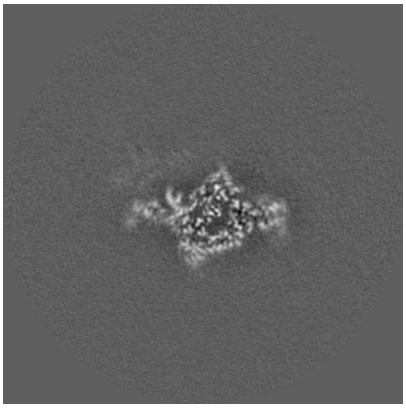
Y Index: 164

Z Index: 160

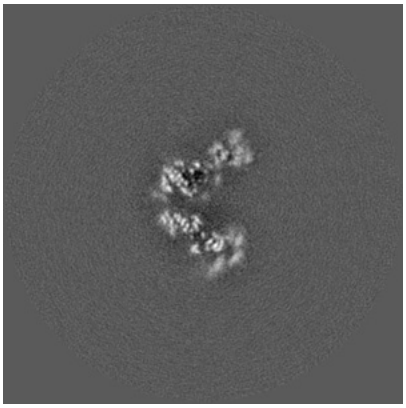
6.3.2 Raw map



X Index: 145



Y Index: 171

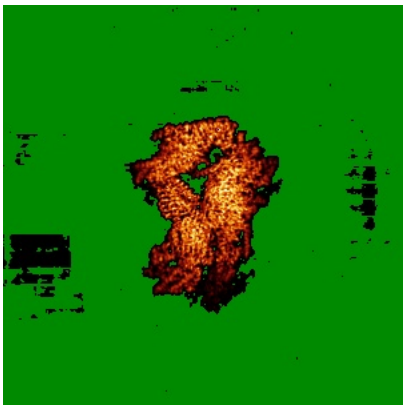


Z Index: 178

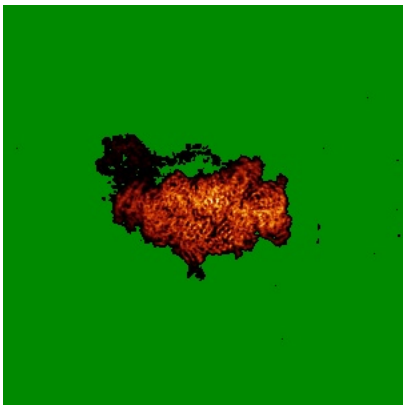
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) (i)

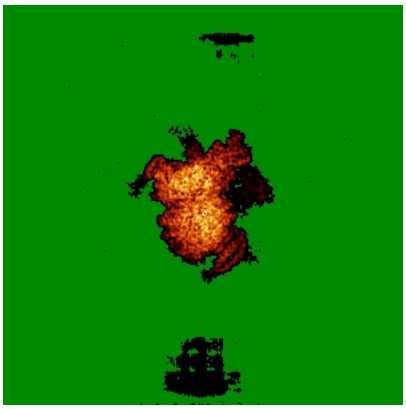
6.4.1 Primary map



X

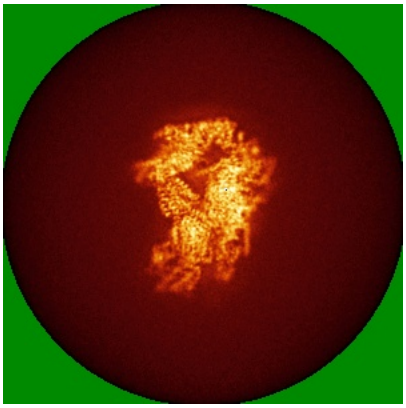


Y

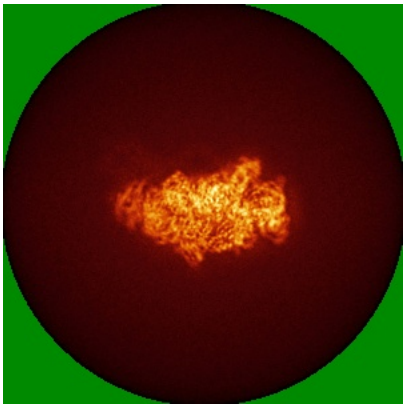


Z

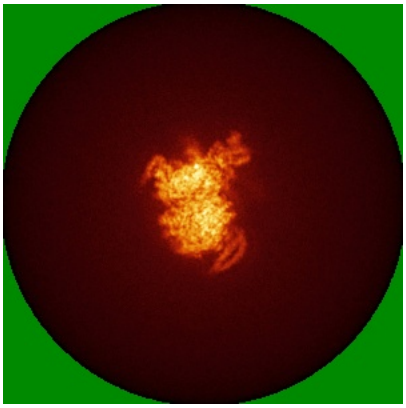
6.4.2 Raw map



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.

6.5 Orthogonal surface views (i)

6.5.1 Primary map

X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.01. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.5.2 Raw map

X Y Z

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

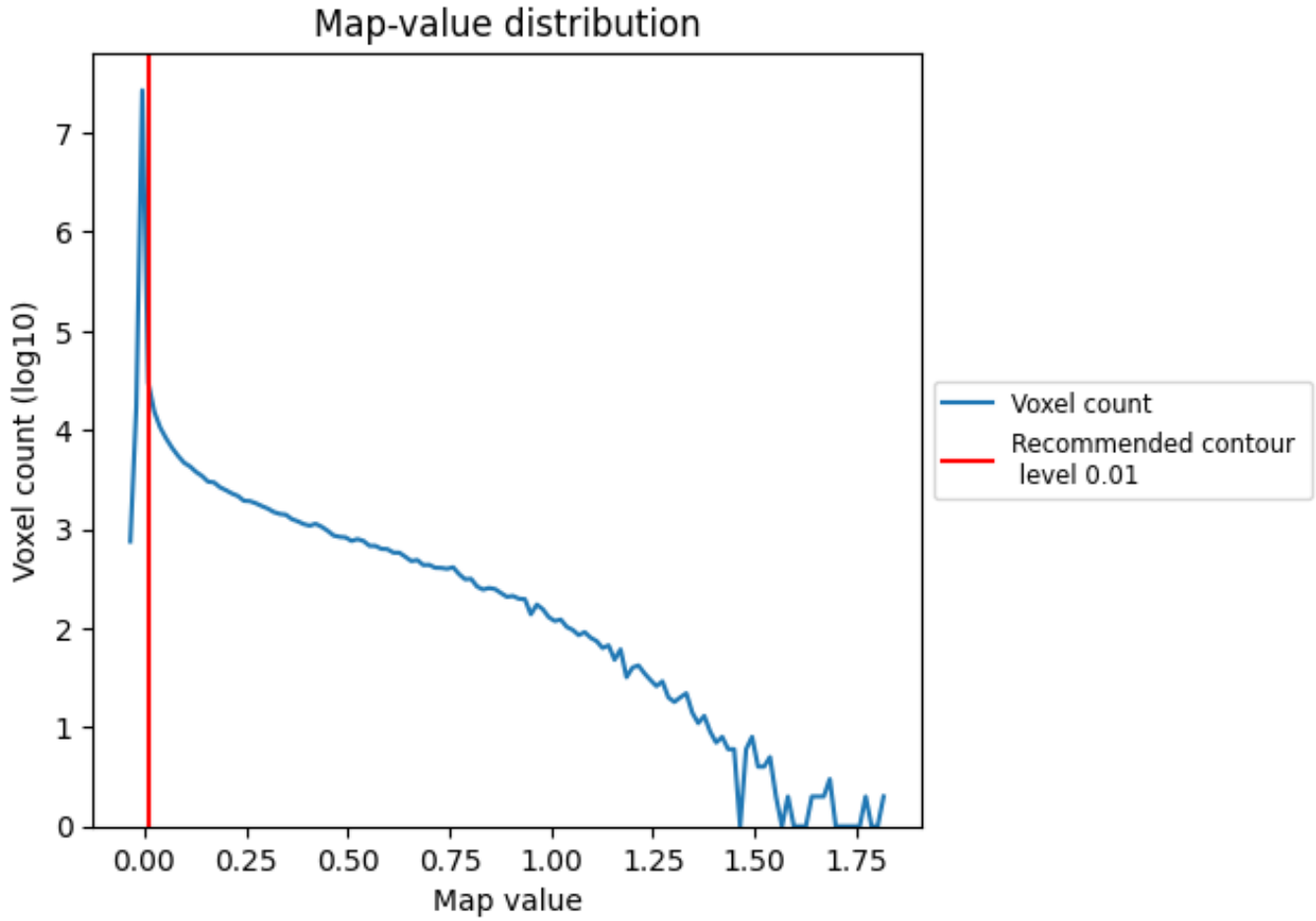
6.6 Mask visualisation (i)

This section was not generated. No masks/segmentation were deposited.

7 Map analysis

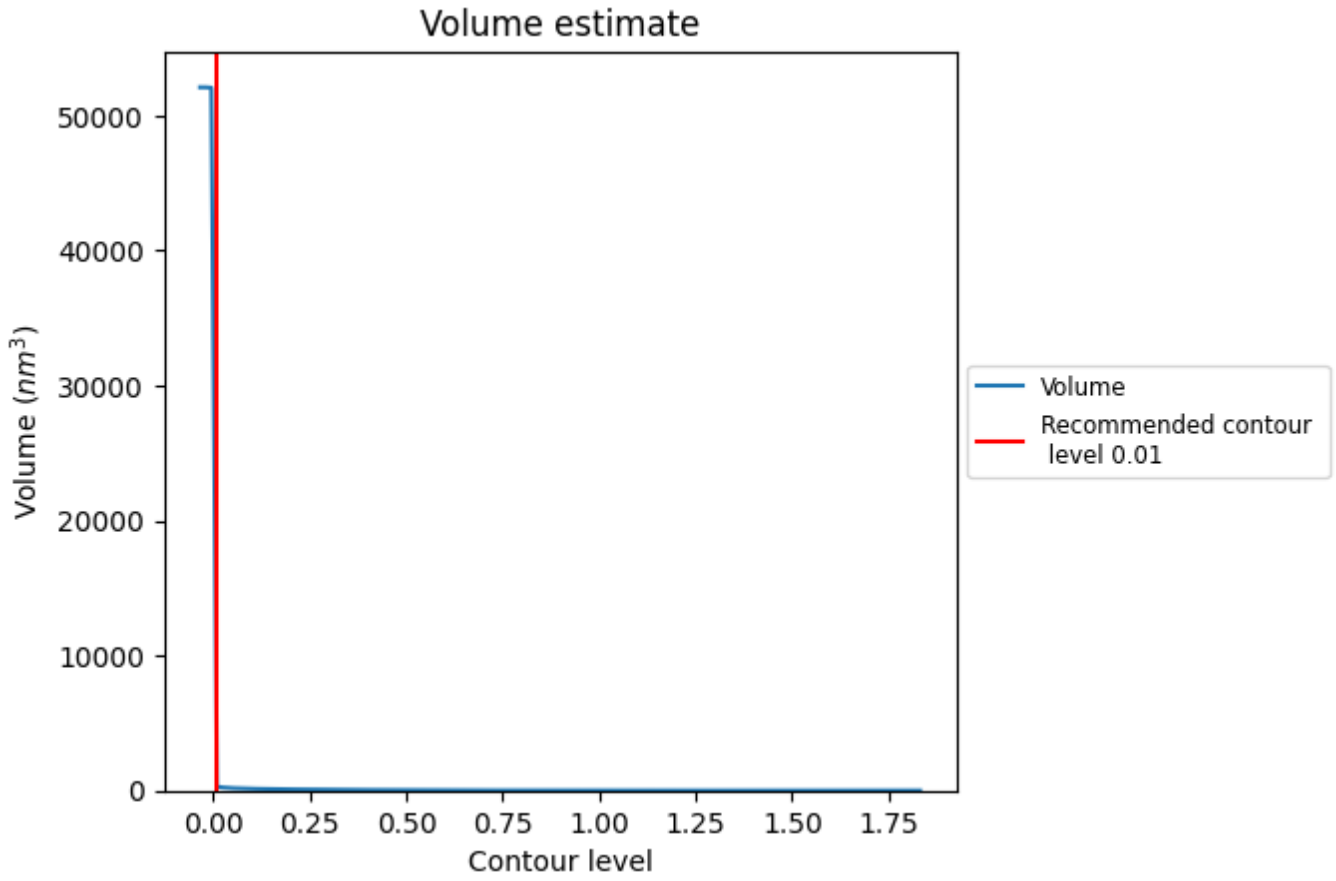
This section contains the results of statistical analysis of the map.

7.1 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.

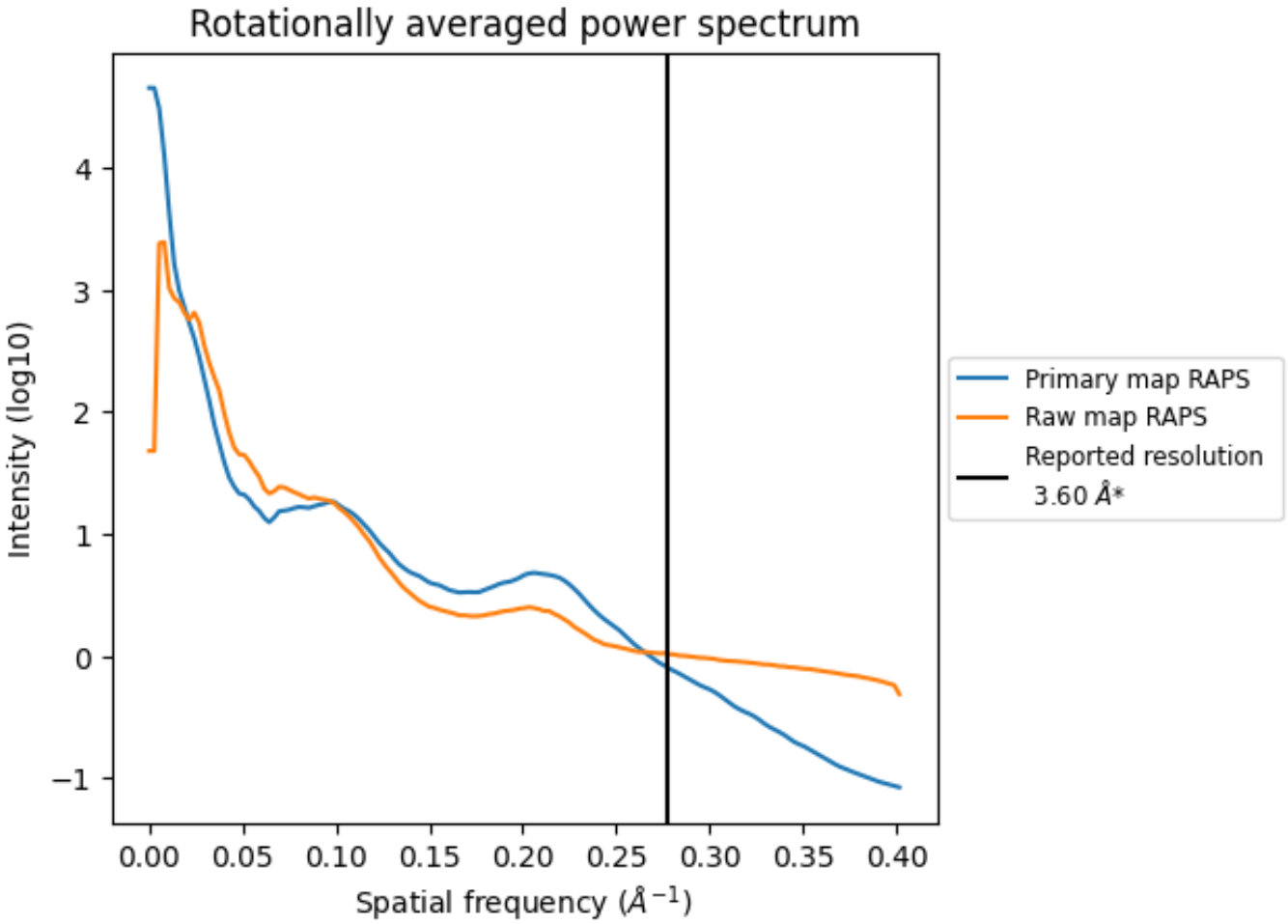
7.2 Volume estimate (i)



The volume at the recommended contour level is 282 nm^3; this corresponds to an approximate mass of 255 kDa.

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.

7.3 Rotationally averaged power spectrum (i)

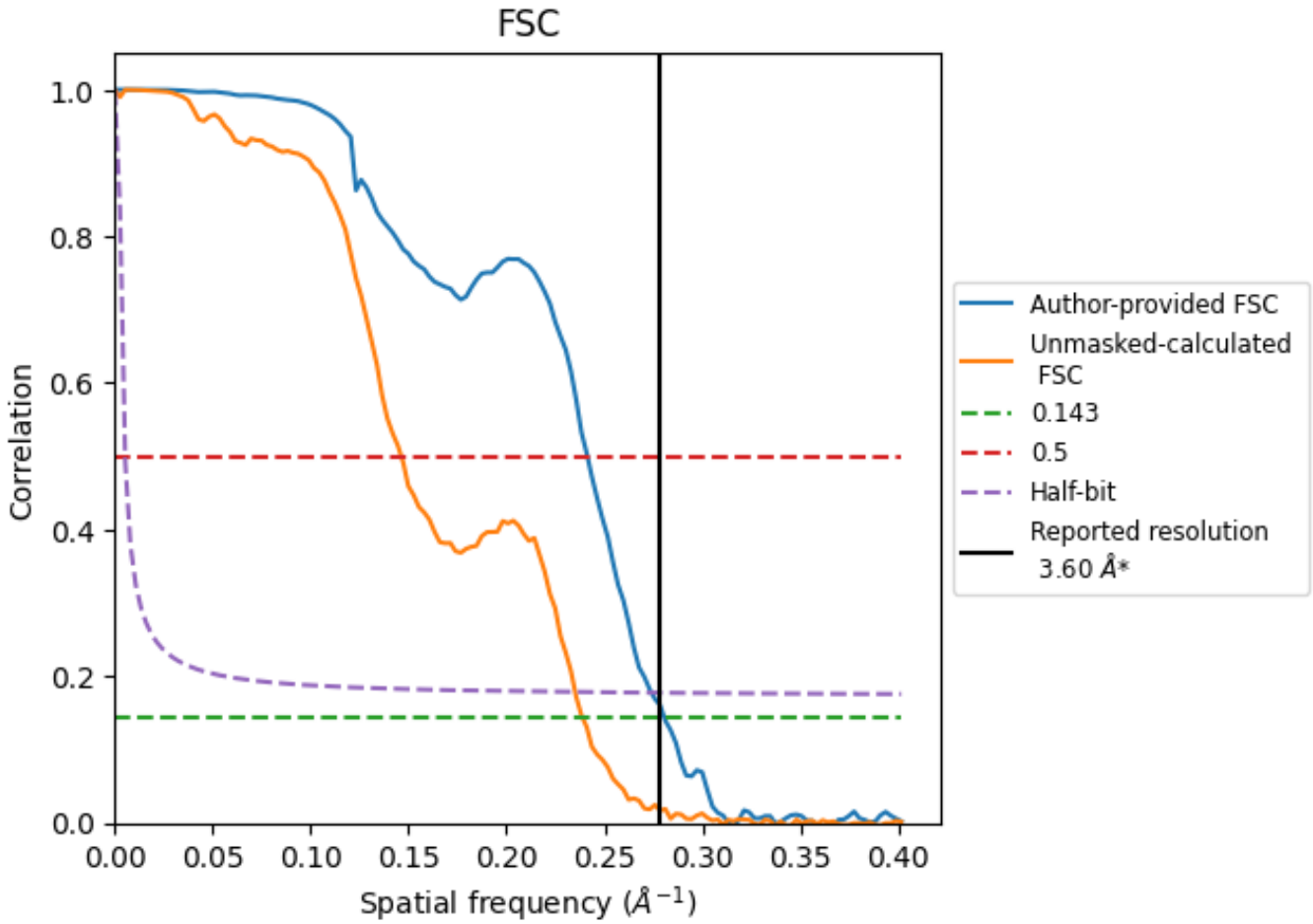


*Reported resolution corresponds to spatial frequency of 0.278 Å^{-1}

8 Fourier-Shell correlation

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.

8.1 FSC



*Reported resolution corresponds to spatial frequency of 0.278 Å^{-1}

8.2 Resolution estimates (i)

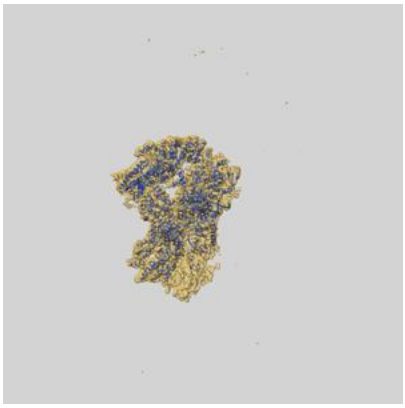
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.60	-	-
Author-provided FSC curve	3.56	4.14	3.65
Unmasked-calculated*	4.19	6.84	4.25

*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 4.19 differs from the reported value 3.6 by more than 10 %

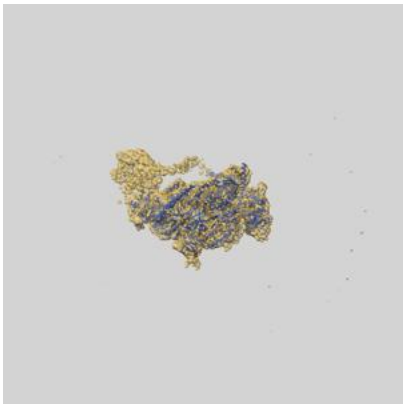
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-28002 and PDB model 8EBY. Per-residue inclusion information can be found in section 3 on page 8.

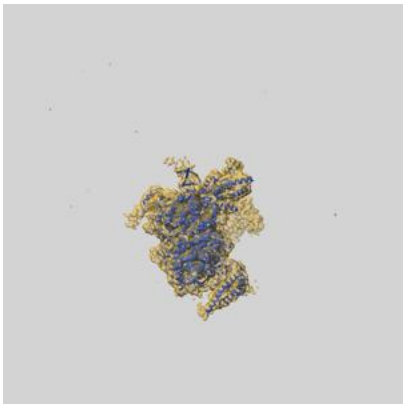
9.1 Map-model overlay



X



Y



Z

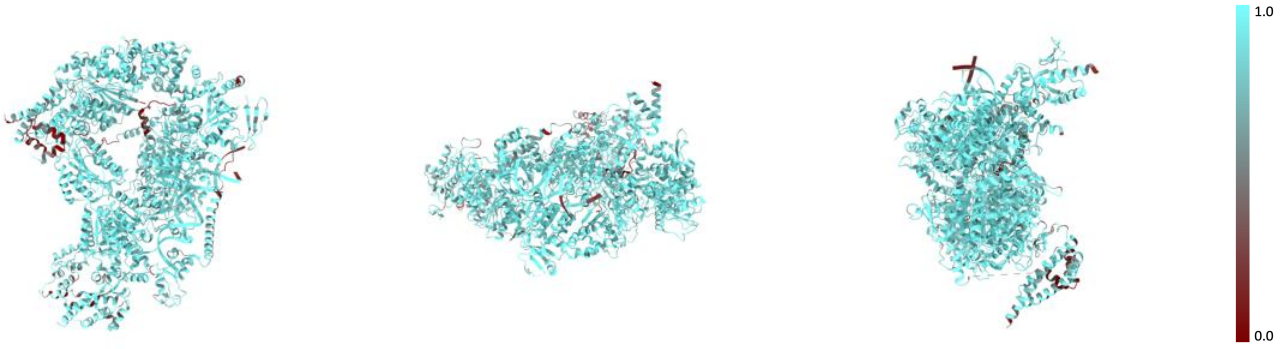
The images above show the 3D surface view of the map at the recommended contour level 0.01 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2 Q-score mapped to coordinate model (i)

X Y Z

The images above show the model with each residue coloured according 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.

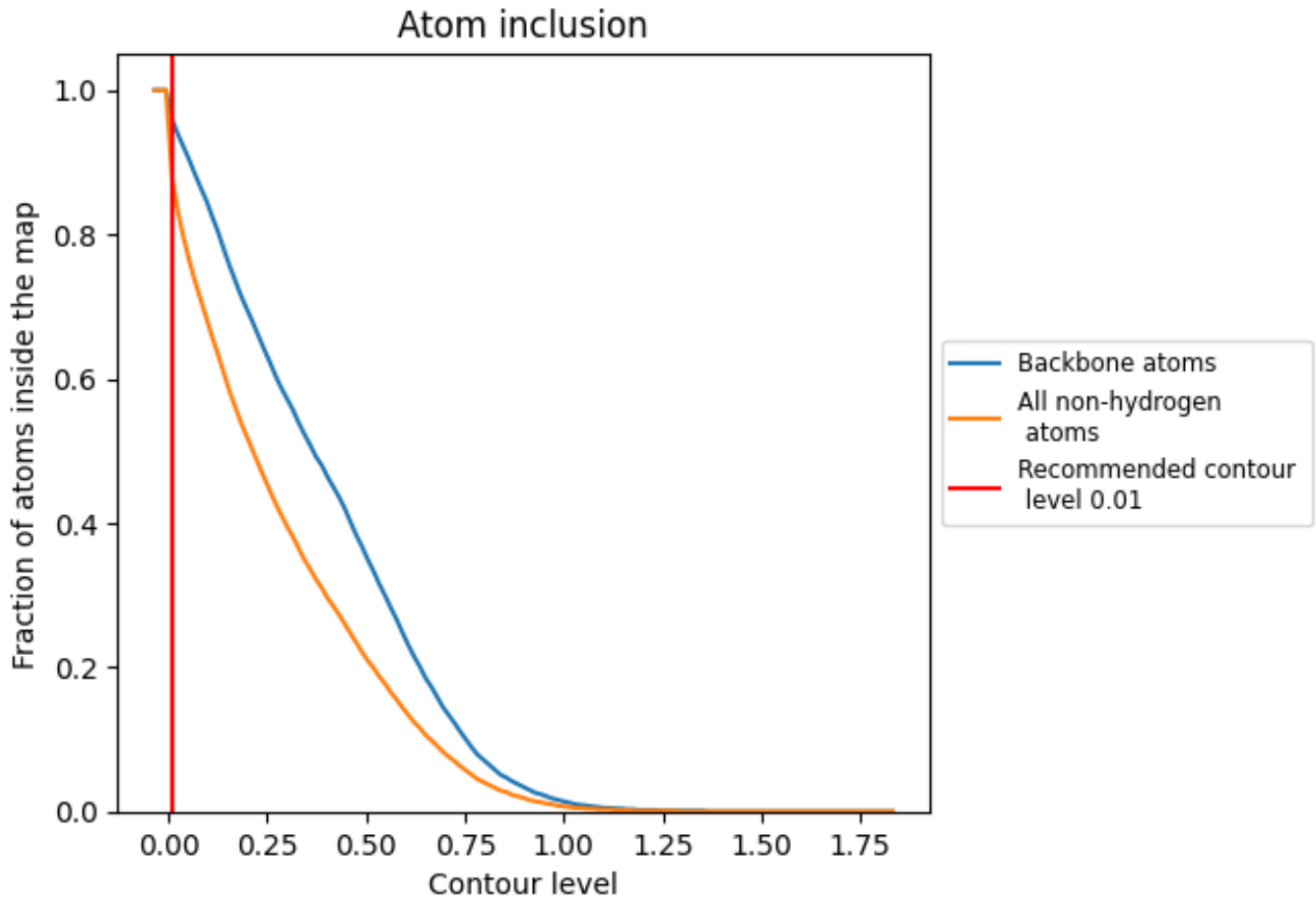
9.3 Atom inclusion mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.01).

9.4 Atom inclusion (i)



At the recommended contour level, 96% of all backbone atoms, 88% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary

The table lists the average atom inclusion at the recommended contour level (0.01) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8790	0.3000
A	0.9330	0.3750
B	0.8740	0.2470
C	0.7130	0.1310
D	0.9070	0.3130
E	0.8950	0.3710
F	0.9250	0.4020
G	0.9210	0.3070
H	0.5800	0.1190
K	0.7770	0.1530
L	0.8900	0.2820
M	0.9110	0.3270