wwPDB EM Validation Summary Report

Feb 25, 2024 – 11:51 AM EST

PDB ID	:	6X50
EMDB ID	:	EMD-22045
Title	:	Mfd-bound E.coli RNA polymerase elongation complex - V state
Authors	:	Llewelyn, E.; Chen, J.; Kang, J.Y.; Darst, S.A.
Deposited on	:	2020-05-24
Resolution	:	3.30 Å(reported)

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:

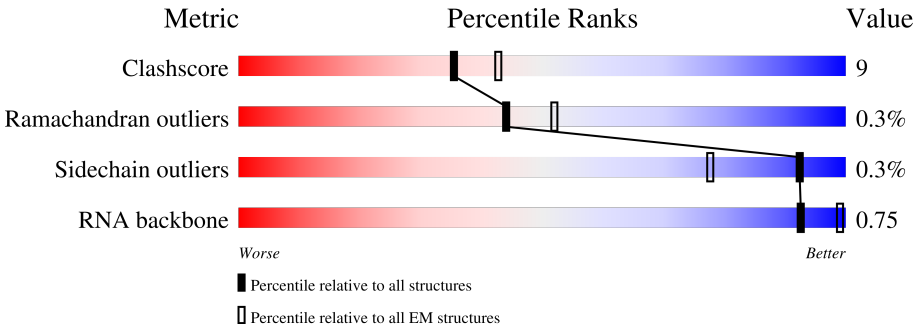
EMDB validation analysis	:	0.0.1.dev70
Mogul	:	1.8.5 (274361), CSD as541be (2020)
MolProbity	:	4.02b-467
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.30 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
RNA backbone	4643	859

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	1148	80% 19%
2	G	329	52% 16% 32%
2	H	329	50% 16% 34%
3	I	1342	78% 20%
4	J	1407	5% 76% 18% 5%
5	K	91	73% 14% 13%
6	R	20	30% 10% 5% 55%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	P	64	
8	Q	64	

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 36077 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 Transcription-repair-coupling factor.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1143	Total	C	N	O	S	0	0
			8946	5661	1598	1651	36		

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	G	224	Total	C	N	O	S	0	0
			1725	1080	303	336	6		
2	H	217	Total	C	N	O	S	0	0
			1667	1043	293	325	6		

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	I	1316	Total	C	N	O	S	0	0
			10368	6507	1808	2010	43		

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	J	1336	Total	C	N	O	S	0	0
			10383	6524	1852	1957	50		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	1384	VAL	MET	conflict	UNP A0A4S1NBU2

- Molecule 5 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	K	79	Total	C	N	O	S	0	0
			627	382	118	126	1		

- Molecule 6 is a RNA chain called RNA (20-mer).

Mol	Chain	Residues	Atoms					AltConf	Trace
6	R	9	Total	C	N	O	P	0	0
			201	89	42	61	9		

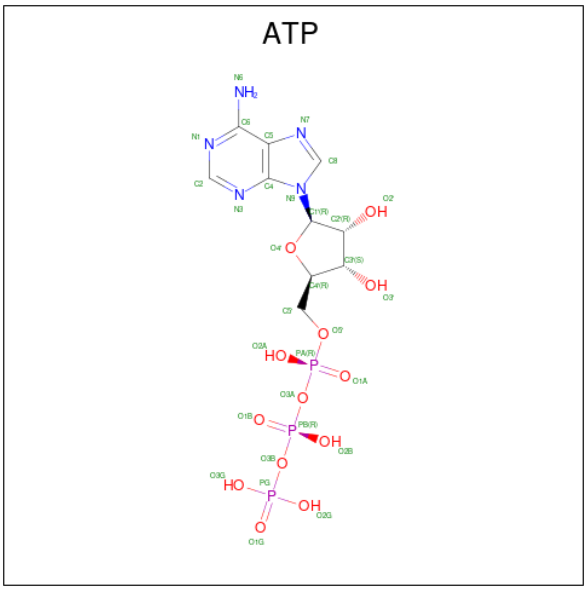
- Molecule 7 is a DNA chain called DNA (64-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
7	P	55	Total	C	N	O	P	0	0
			1112	531	192	335	54		

- Molecule 8 is a DNA chain called DNA (64-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
8	Q	49	Total	C	N	O	P	0	0
			1013	478	197	289	49		

- Molecule 9 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: $C_{10}H_{16}N_5O_{13}P_3$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					AltConf
9	A	1	Total	C	N	O	P	0
			31	10	5	13	3	

- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	AltConf
10	A	1	Total Mg 1 1	0
10	J	1	Total Mg 1 1	0

- Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

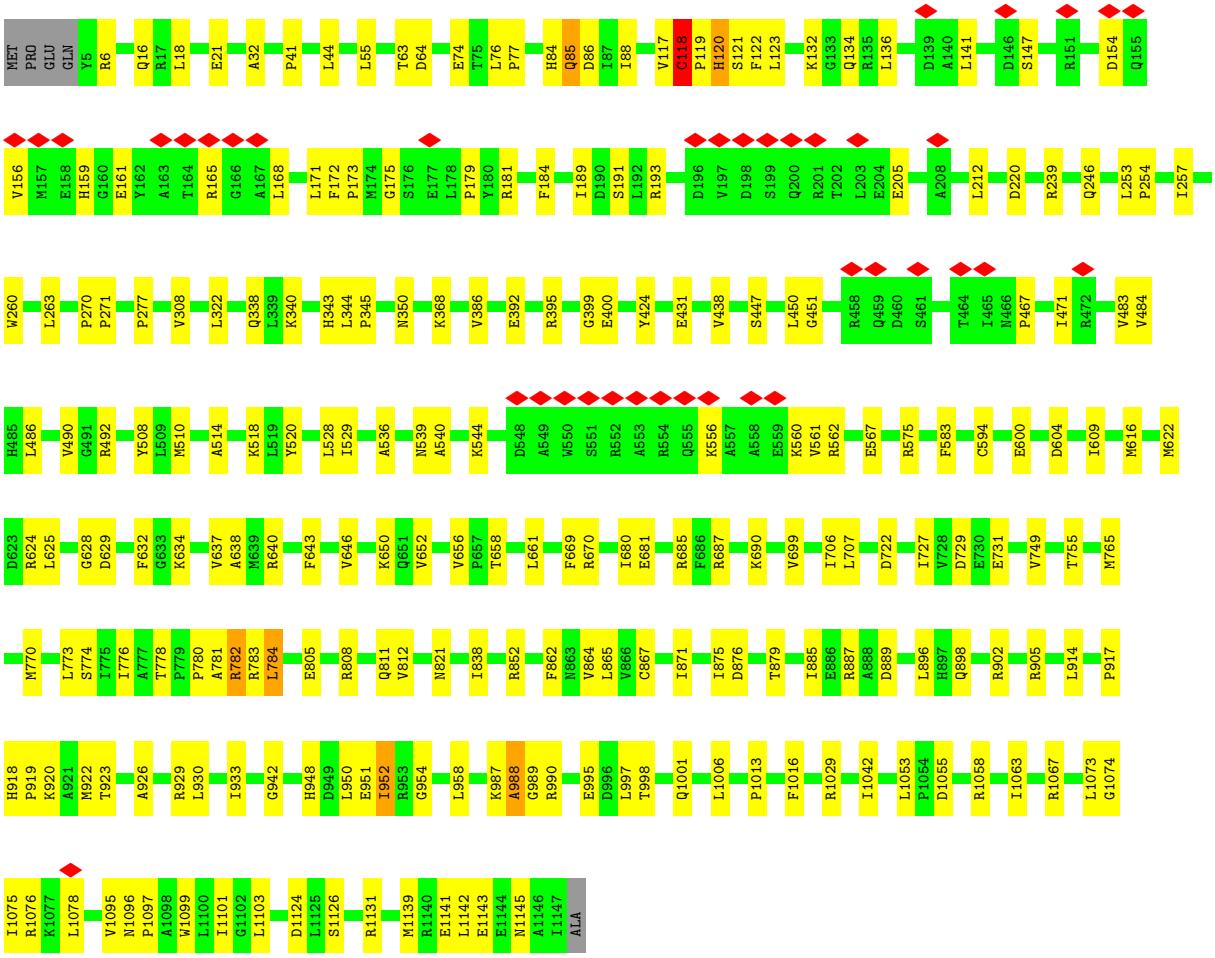
Mol	Chain	Residues	Atoms	AltConf
11	J	2	Total Zn 2 2	0

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.

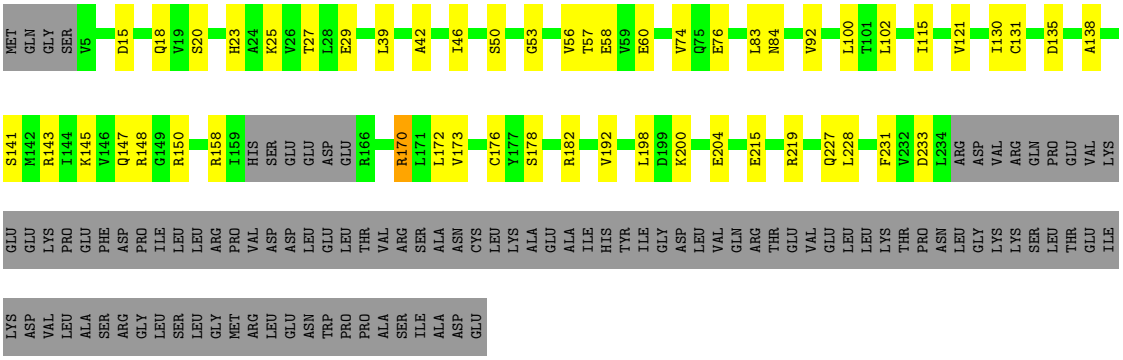
• Molecule 1: Transcription-repair-coupling factor

Chain A:



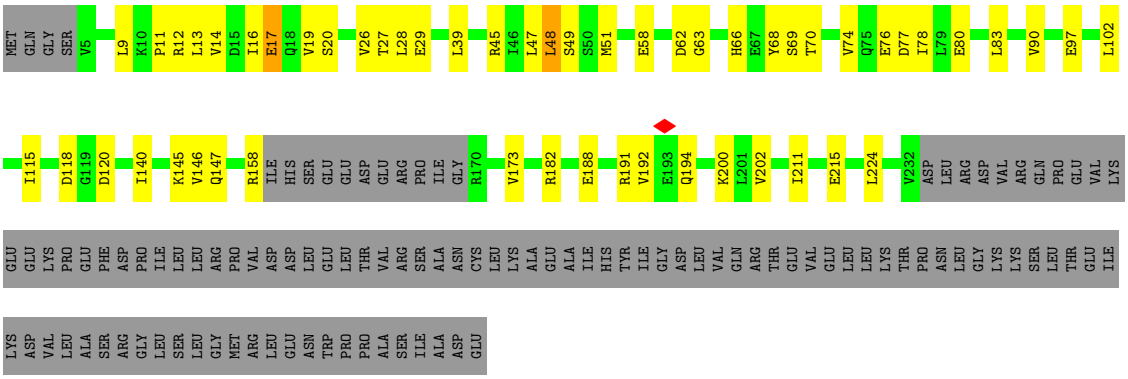
• Molecule 2: DNA-directed RNA polymerase subunit alpha

Chain G:



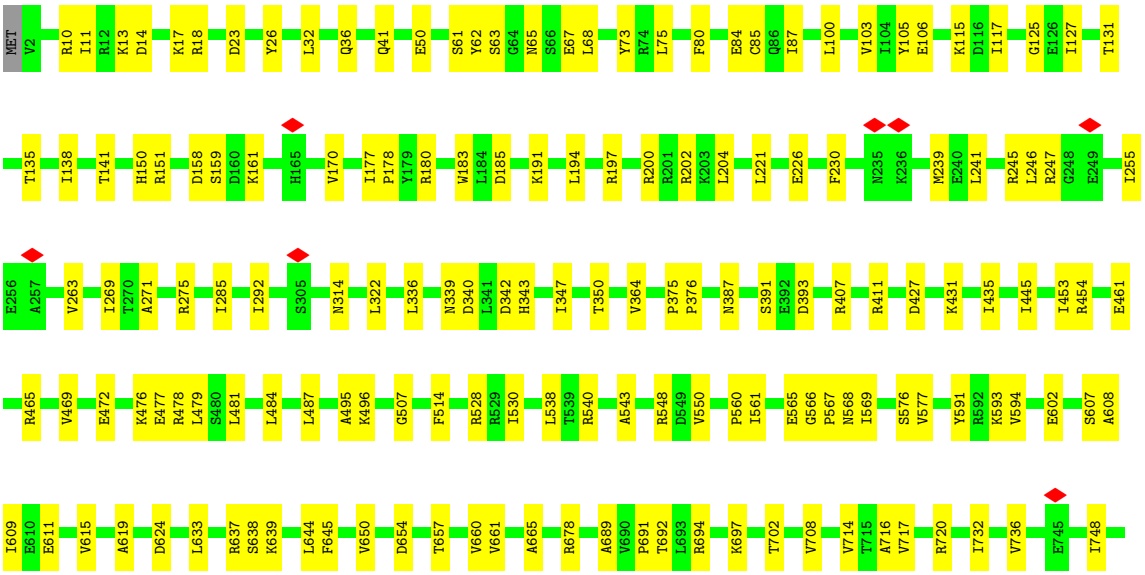
• Molecule 2: DNA-directed RNA polymerase subunit alpha

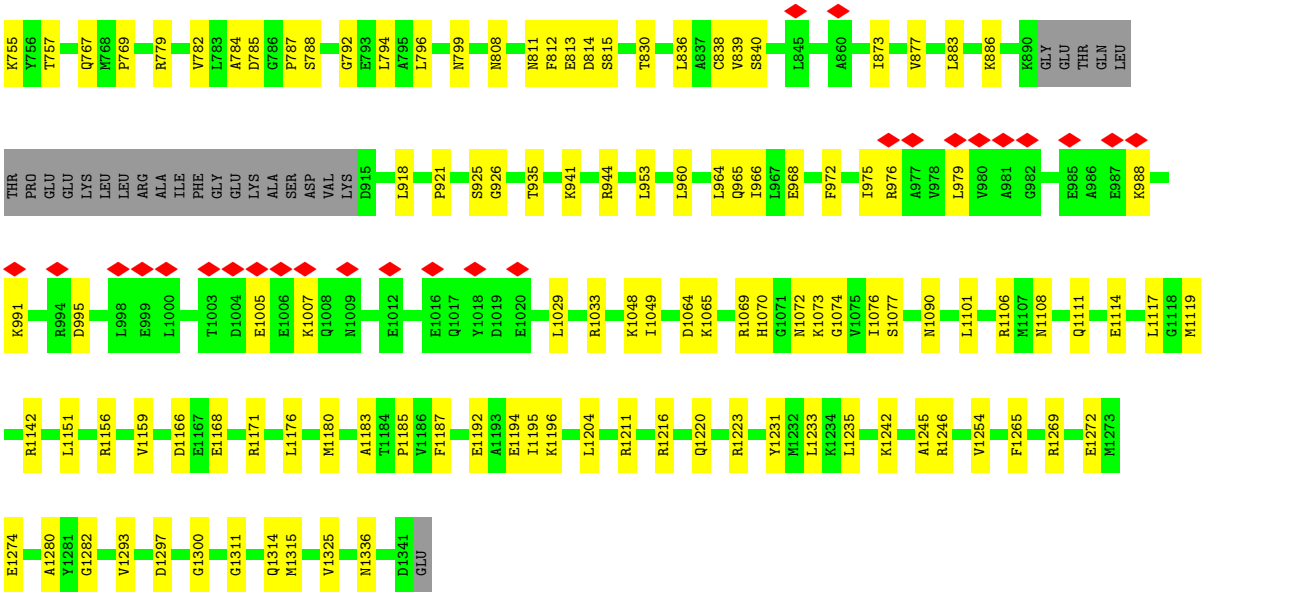
Chain H:



• Molecule 3: DNA-directed RNA polymerase subunit beta

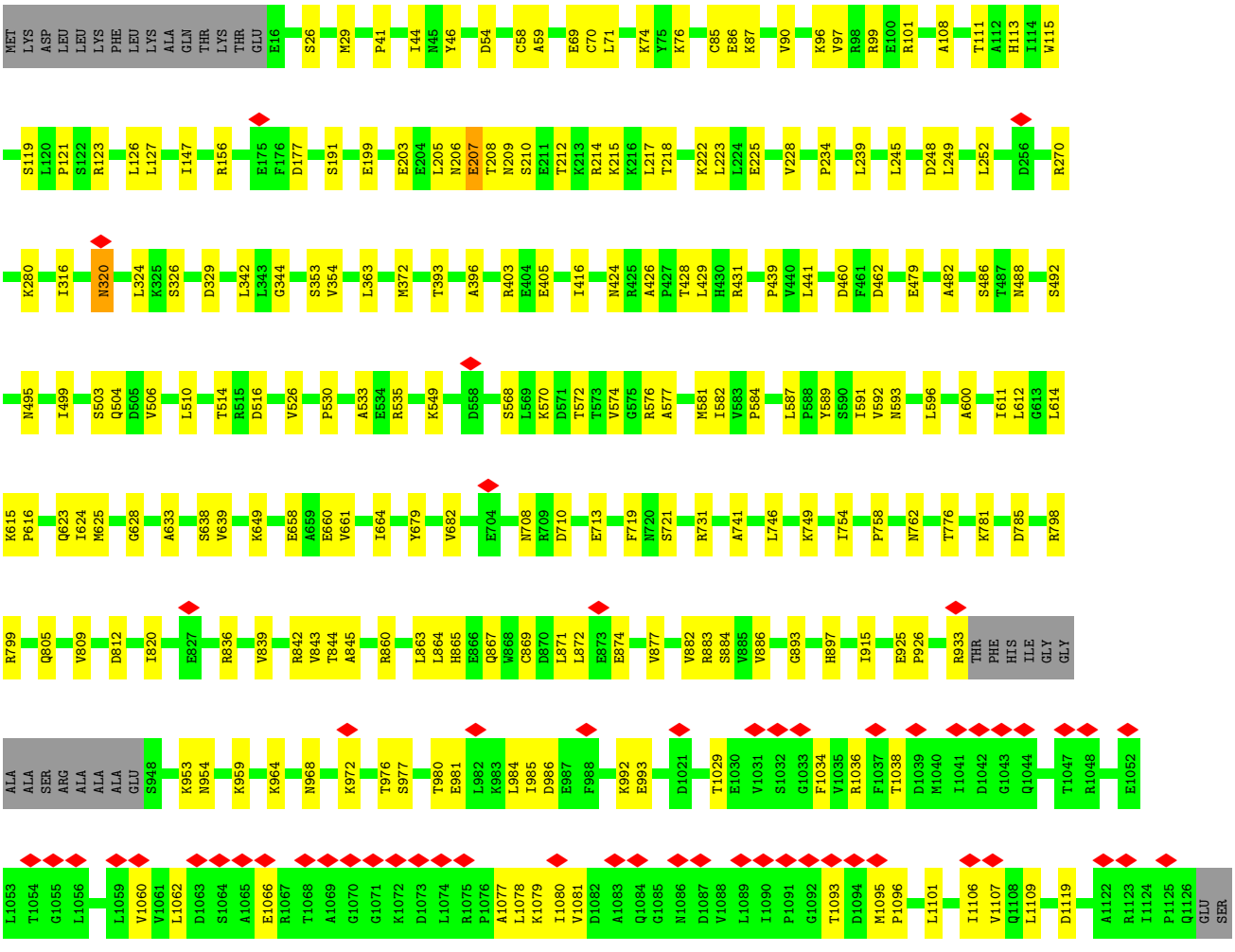
Chain I:

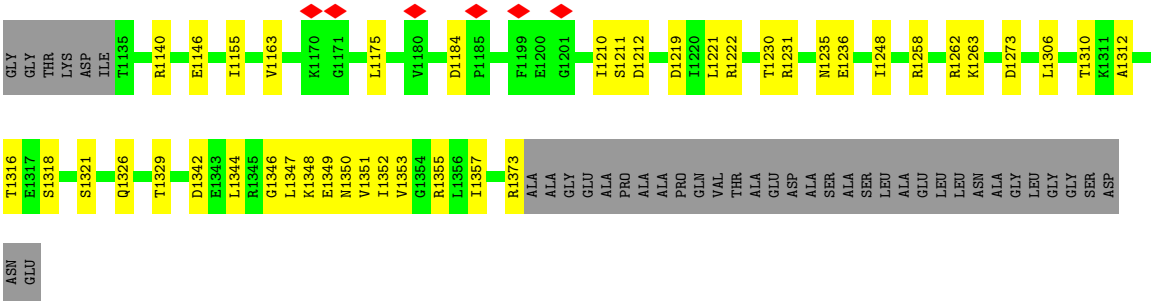




• Molecule 4: DNA-directed RNA polymerase subunit beta'

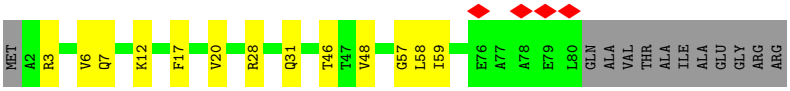
Chain J:





• Molecule 5: DNA-directed RNA polymerase subunit omega

Chain K:

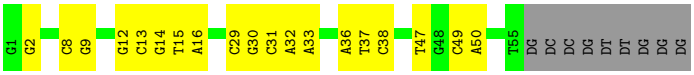


• Molecule 6: RNA (20-mer)

Chain R:

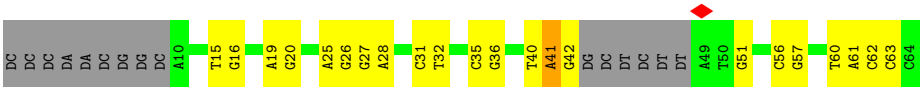
• Molecule 7: DNA (64-MER)

Chain P:



• Molecule 8: DNA (64-MER)

Chain Q:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	103000	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 ($e^-/Å^2$)	60	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	43.752	Depositor
Minimum map value	-24.685	Depositor
Average map value	0.004	Depositor
Map value standard deviation	1.079	Depositor
Recommended contour level	3.5	Depositor
Map size (Å)	390.0, 390.0, 390.0	wwPDB
Map dimensions	300, 300, 300	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.3, 1.3, 1.3	Depositor

5 Model quality

5.1 Standard geometry

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

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	0.34	0/9124	0.53	1/12371 (0.0%)
2	G	0.30	0/1745	0.54	2/2366 (0.1%)
2	H	0.27	0/1686	0.57	2/2286 (0.1%)
3	I	0.31	0/10533	0.52	1/14214 (0.0%)
4	J	0.29	0/10540	0.53	3/14231 (0.0%)
5	K	0.27	0/629	0.45	0/847
6	R	0.45	0/226	0.92	0/352
7	P	0.63	0/1242	0.97	0/1913
8	Q	0.59	0/1138	0.92	3/1753 (0.2%)
All	All	0.34	0/36863	0.57	12/50333 (0.0%)

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	I	1151	LEU	CA-CB-CG	7.93	133.53	115.30
8	Q	41	DA	O4'-C4'-C3'	-6.80	101.78	104.50
8	Q	41	DA	C3'-C2'-C1'	-6.47	94.73	102.50
1	A	784	LEU	CB-CG-CD1	-6.46	100.01	111.00
2	H	48	LEU	CA-CB-CG	6.06	129.24	115.30

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	8946	0	8855	173	0
2	G	1725	0	1763	30	0
2	H	1667	0	1699	41	0
3	I	10368	0	10379	201	0
4	J	10383	0	10597	164	0
5	K	627	0	634	12	0
6	R	201	0	98	2	0
7	P	1112	0	622	19	0
8	Q	1013	0	549	15	0
9	A	31	0	12	1	0
10	A	1	0	0	0	0
10	J	1	0	0	0	0
11	J	2	0	0	0	0
All	All	36077	0	35208	609	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 9.

The worst 5 of 609 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)
3:I:62:TYR:CZ	3:I:476:LYS:HD3	1.48	1.47
1:A:780:PRO:HG2	1:A:783:ARG:CG	1.50	1.39
1:A:604:ASP:CB	1:A:778:THR:CG2	2.07	1.32
1:A:604:ASP:CB	1:A:778:THR:HG21	1.61	1.30
3:I:62:TYR:CZ	3:I:476:LYS:CD	2.14	1.30

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	1141/1148 (99%)	1030 (90%)	100 (9%)	11 (1%)	15 46
2	G	220/329 (67%)	199 (90%)	21 (10%)	0	100 100
2	H	213/329 (65%)	195 (92%)	17 (8%)	1 (0%)	29 61
3	I	1312/1342 (98%)	1218 (93%)	94 (7%)	0	100 100
4	J	1330/1407 (94%)	1234 (93%)	96 (7%)	0	100 100
5	K	77/91 (85%)	71 (92%)	6 (8%)	0	100 100
All	All	4293/4646 (92%)	3947 (92%)	334 (8%)	12 (0%)	44 71

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	118	CYS
1	A	120	HIS
1	A	467	PRO
1	A	988	ALA
1	A	990	ARG

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	928/973 (95%)	924 (100%)	4 (0%)	91 95
2	G	191/286 (67%)	191 (100%)	0	100 100
2	H	184/286 (64%)	183 (100%)	1 (0%)	88 93
3	I	1132/1157 (98%)	1131 (100%)	1 (0%)	93 97
4	J	1118/1168 (96%)	1114 (100%)	4 (0%)	91 95
5	K	67/75 (89%)	67 (100%)	0	100 100
All	All	3620/3945 (92%)	3610 (100%)	10 (0%)	92 96

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	J	207	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	J	320	ASN
4	J	1373	ARG
1	A	952	ILE
2	H	191	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	I	120	GLN
3	I	799	ASN
4	J	865	HIS
4	J	424	ASN
1	A	821	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
6	R	8/20 (40%)	1 (12%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
6	R	14	A

There are no RNA pucker outliers to report.

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 5 ligands modelled in this entry, 4 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	$\#\|Z\| > 2$
9	ATP	A	2000	10	26,33,33	0.90	1 (3%)	31,52,52	1.61	6 (19%)

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
9	ATP	A	2000	10	-	3/18/38/38	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
9	A	2000	ATP	C5-C4	2.27	1.46	1.40

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
9	A	2000	ATP	PA-O3A-PB	-3.99	119.12	132.83
9	A	2000	ATP	PB-O3B-PG	-3.47	120.94	132.83
9	A	2000	ATP	N3-C2-N1	-3.13	123.78	128.68
9	A	2000	ATP	C3'-C2'-C1'	2.87	105.31	100.98
9	A	2000	ATP	C4-C5-N7	-2.73	106.56	109.40

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
9	A	2000	ATP	C5'-O5'-PA-O1A
9	A	2000	ATP	O4'-C4'-C5'-O5'
9	A	2000	ATP	C5'-O5'-PA-O3A

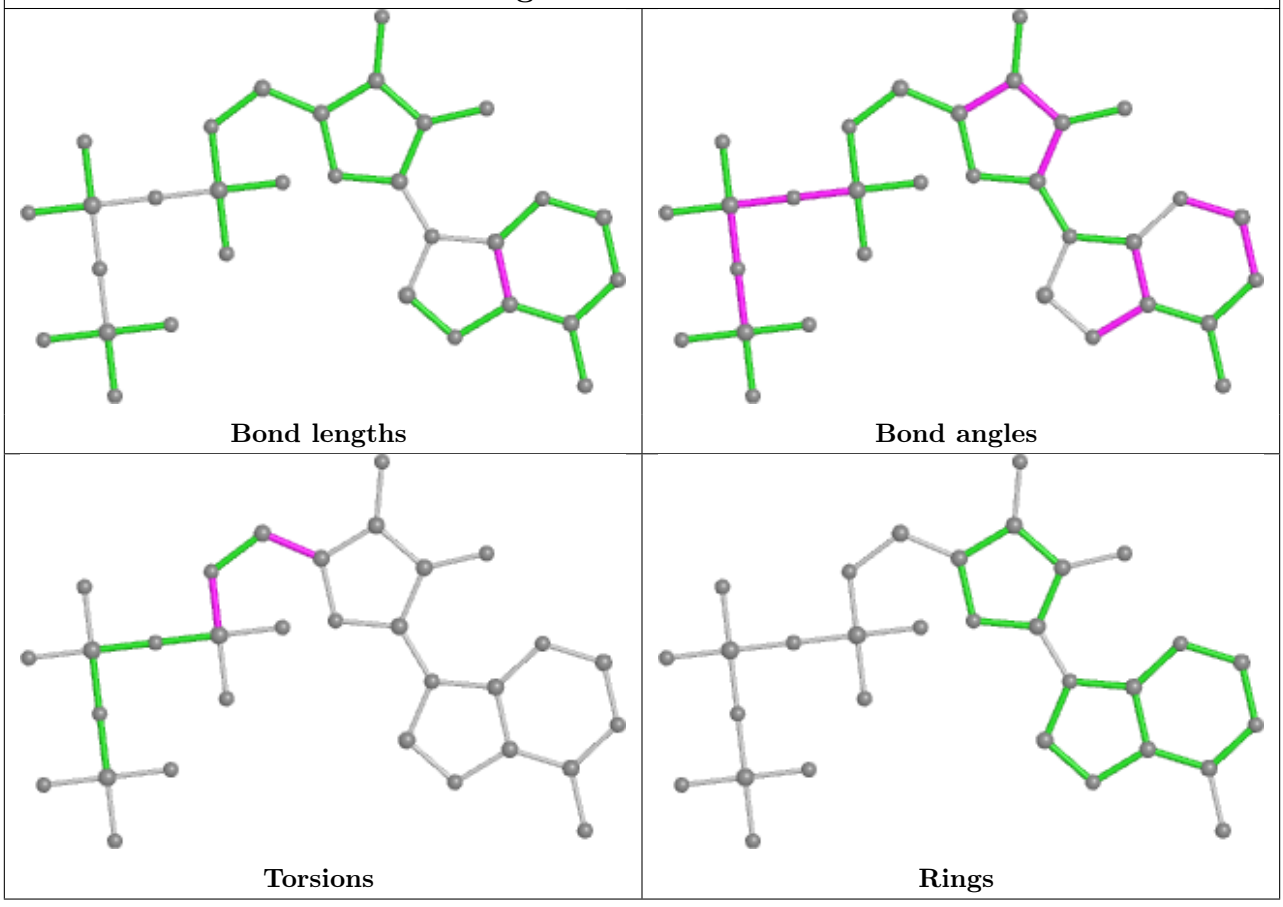
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
9	A	2000	ATP	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand ATP A 2000



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.

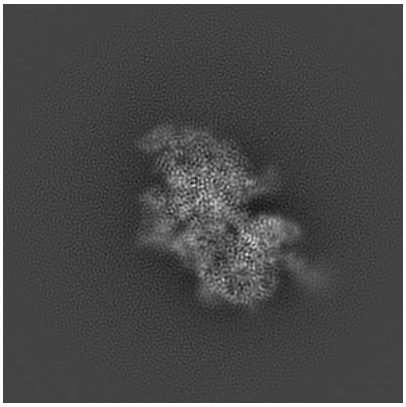
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-22045. These allow visual inspection of the internal detail of the map and identification of artifacts.

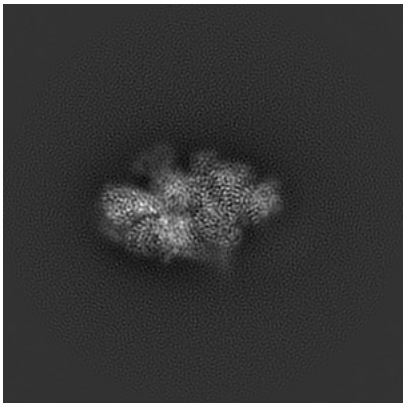
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

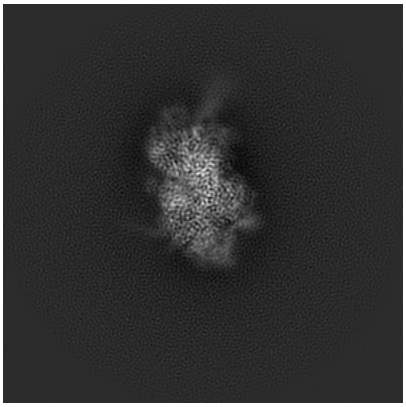
6.1.1 Primary map



X



Y

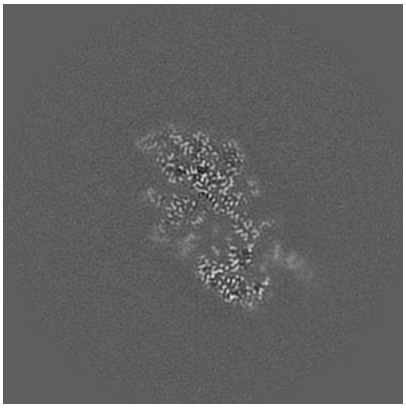


Z

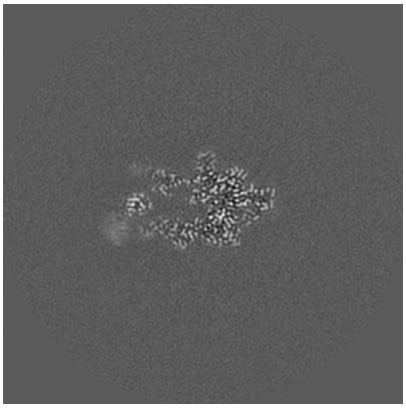
The images above show the map projected in three orthogonal directions.

6.2 Central slices

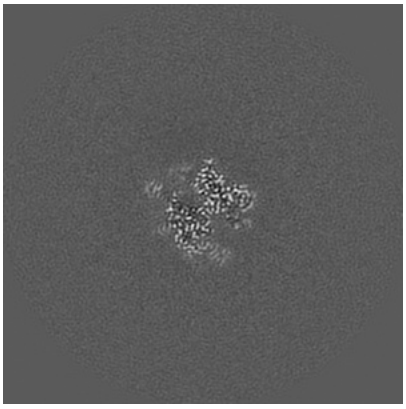
6.2.1 Primary map



X Index: 150



Y Index: 150

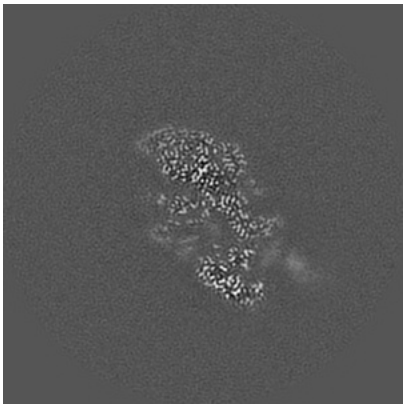


Z Index: 150

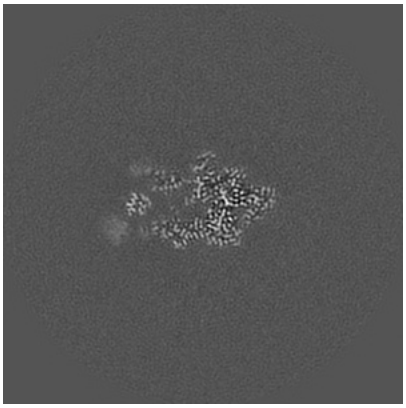
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

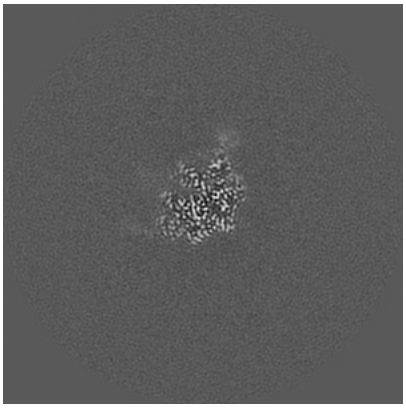
6.3.1 Primary map



X Index: 152



Y Index: 149

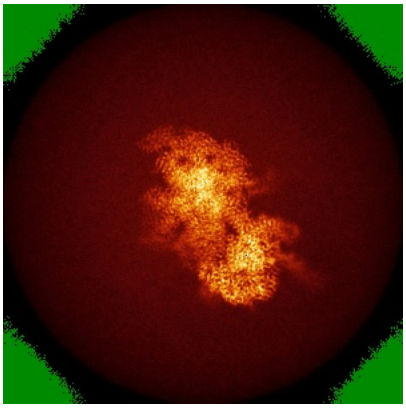


Z Index: 169

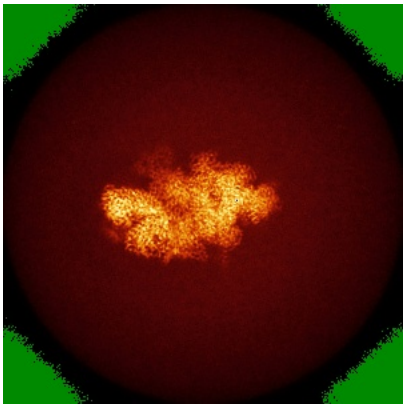
The images above show the largest variance slices of the map in three orthogonal directions.

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

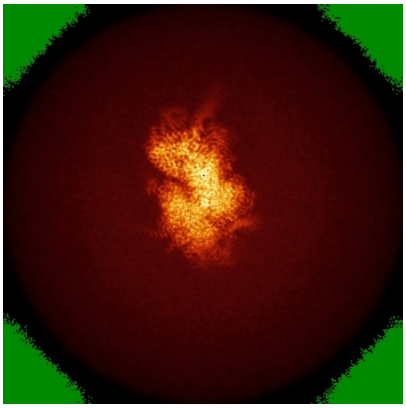
6.4.1 Primary map



X



Y



Z

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 3.5. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

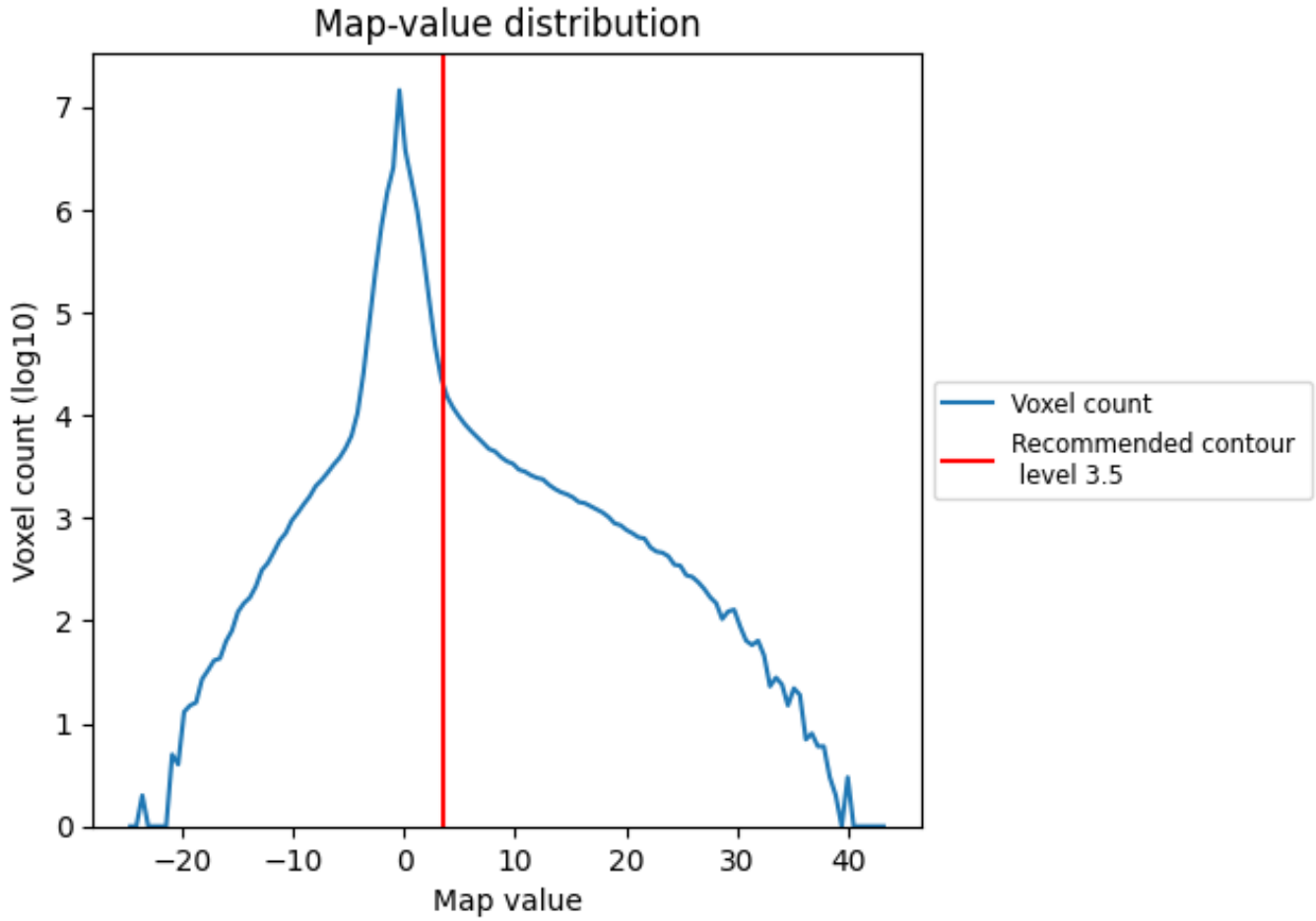
6.6 Mask visualisation (i)

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

7 Map analysis

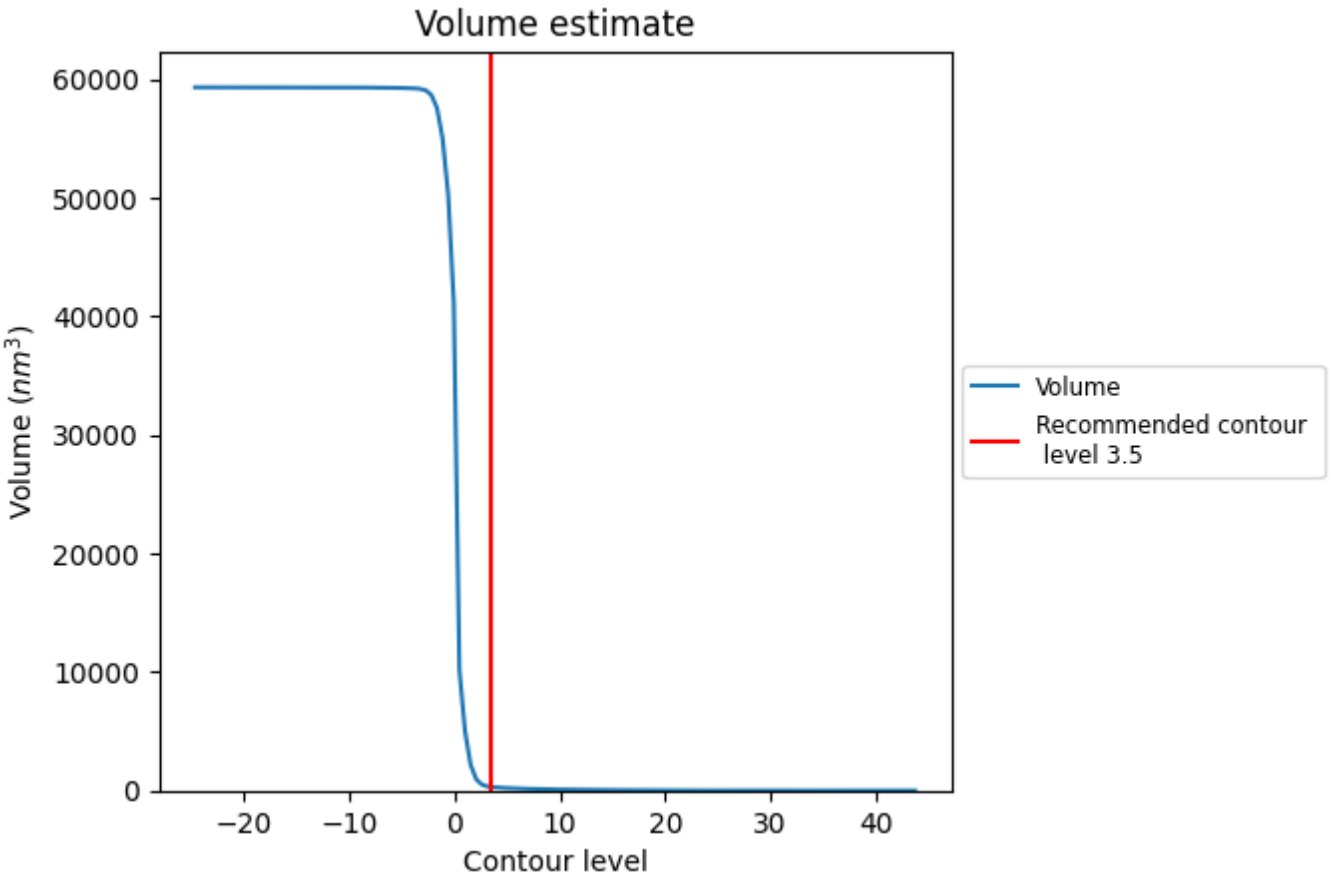
This section contains the results of statistical analysis of the map.

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.

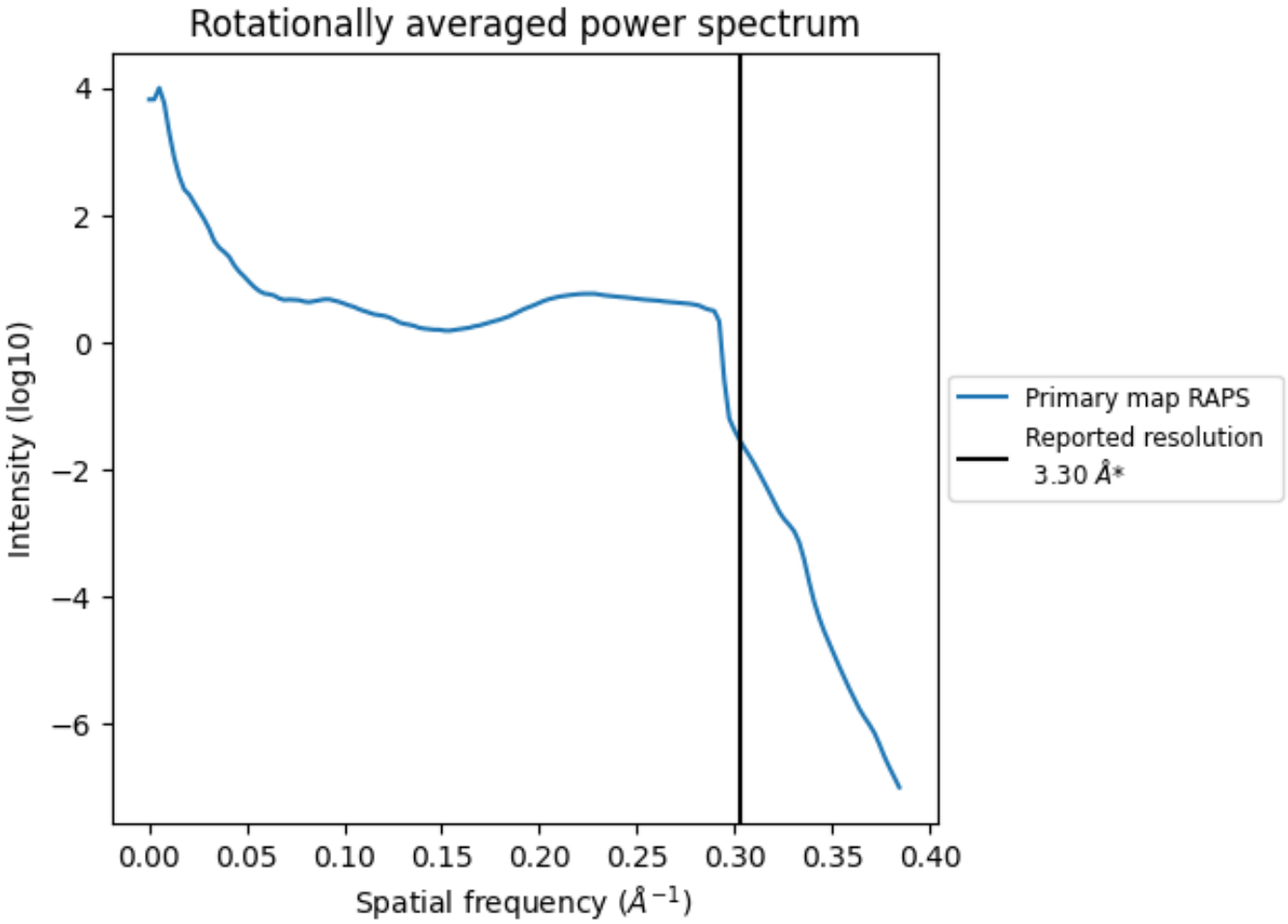
7.2 Volume estimate (i)



The volume at the recommended contour level is 304 nm^3; this corresponds to an approximate mass of 275 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 0.303 Å^{-1}

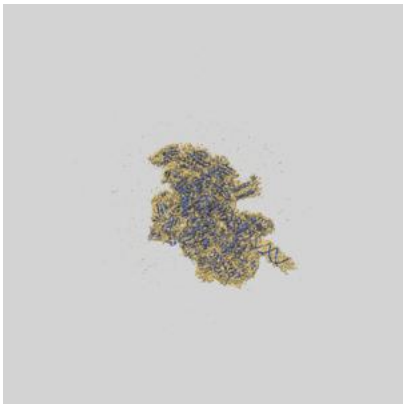
8 Fourier-Shell correlation

This section was not generated. No FSC curve or half-maps provided.

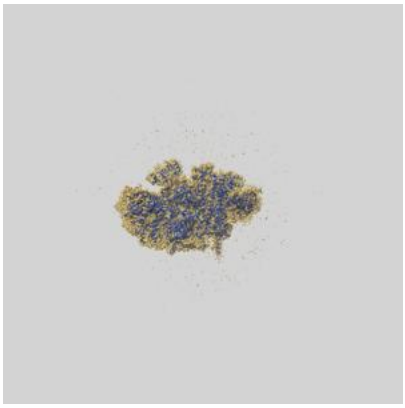
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-22045 and PDB model 6X50. Per-residue inclusion information can be found in section 3 on page 7.

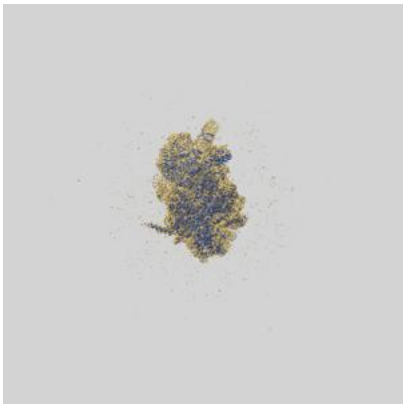
9.1 Map-model overlay



X



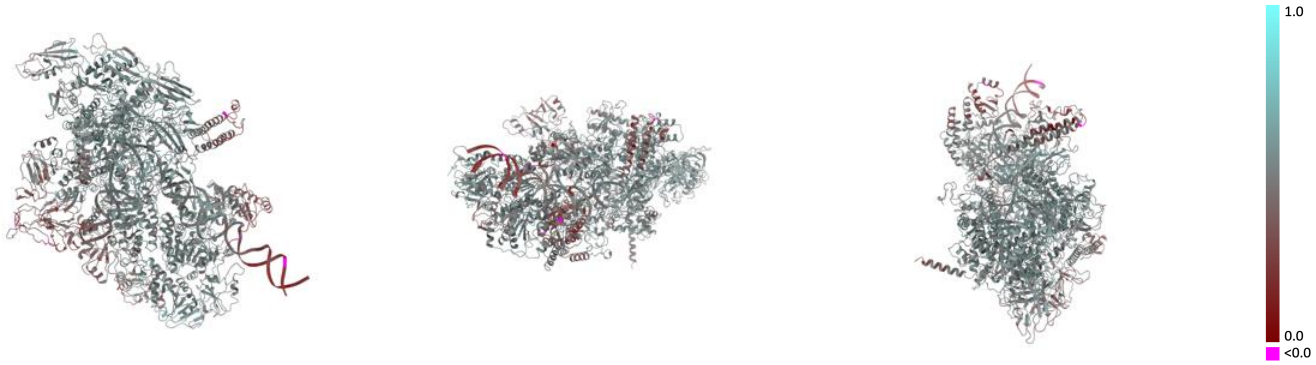
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 3.5 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.

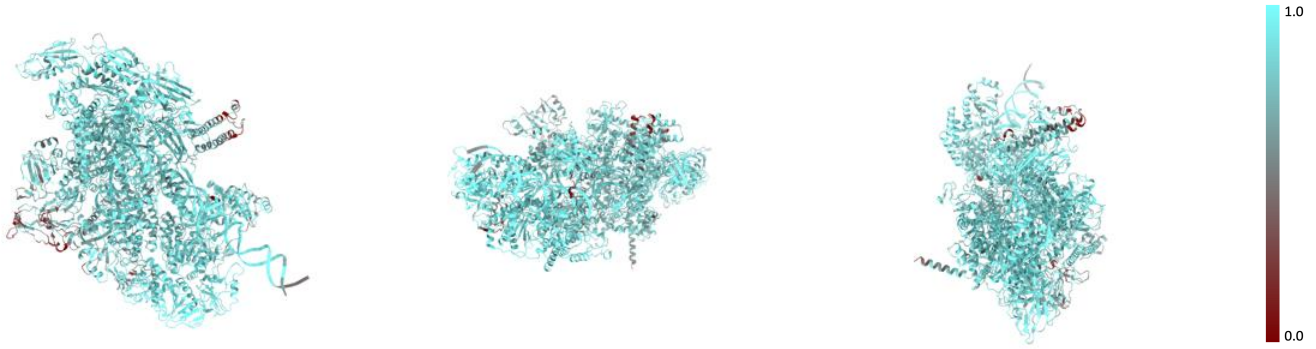
9.2 Q-score mapped to coordinate model



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.

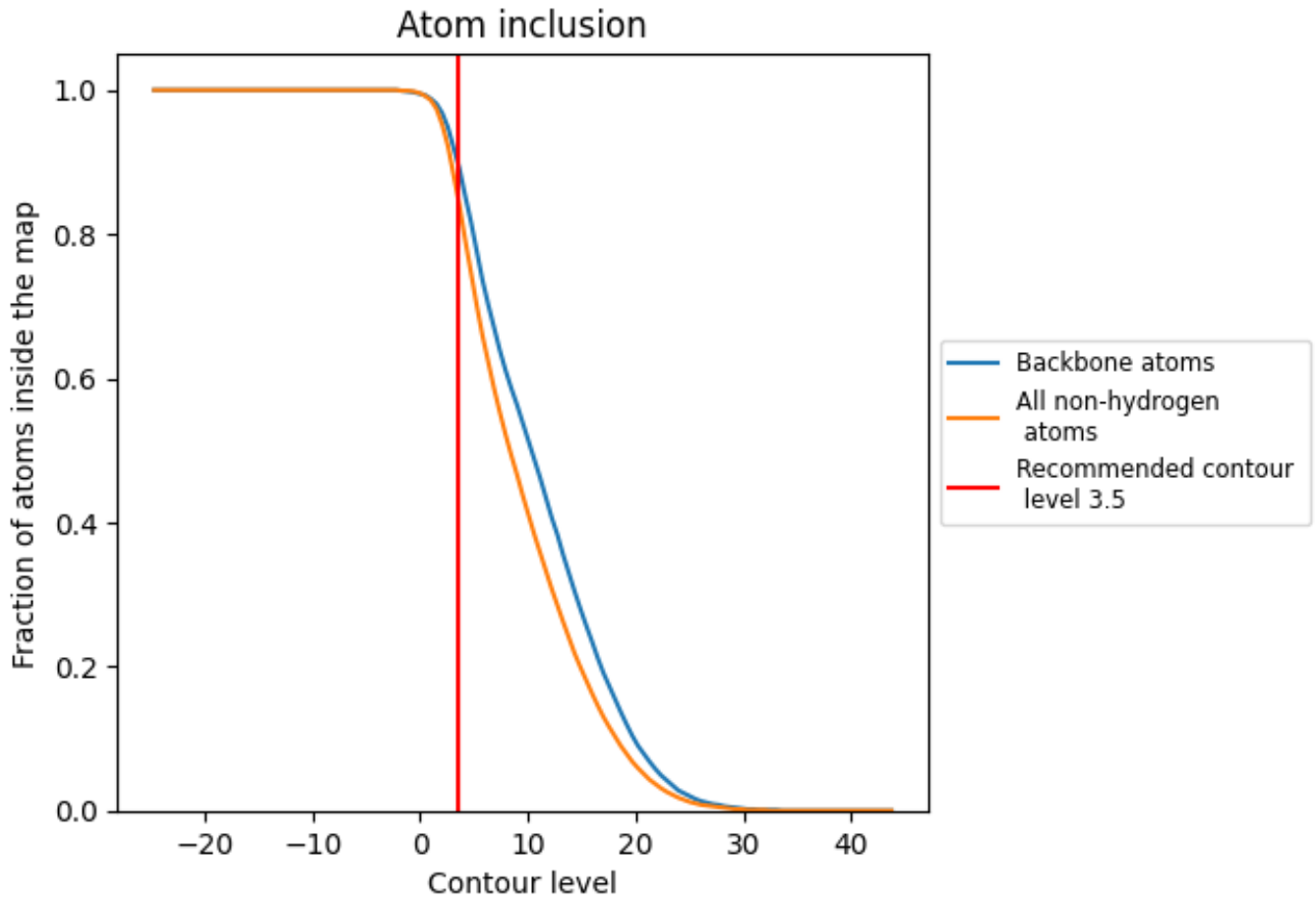
9.3 Atom inclusion mapped to coordinate model



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 (3.5).

9.4 Atom inclusion (i)



At the recommended contour level, 90% of all backbone atoms, 85% 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 (3.5) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8530	0.4940
A	0.8930	0.4870
G	0.8780	0.5340
H	0.8530	0.5060
I	0.8470	0.5100
J	0.8130	0.4940
K	0.7680	0.4920
P	0.9270	0.4300
Q	0.8710	0.3870
R	0.9450	0.5140