wwPDB EM Validation Summary Report

May 26, 2025 – 09:14 PM EDT

PDB ID	:	7K19 / pdb_00007k19
EMDB ID	:	EMD-22622
Title	:	CryoEM structure of DNA-PK catalytic subunit complexed with DNA (Complex I)
Authors	:	Chen, X.; Gellert, M.; Yang, W.
Deposited on	:	2020-09-07
Resolution	:	4.30 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:

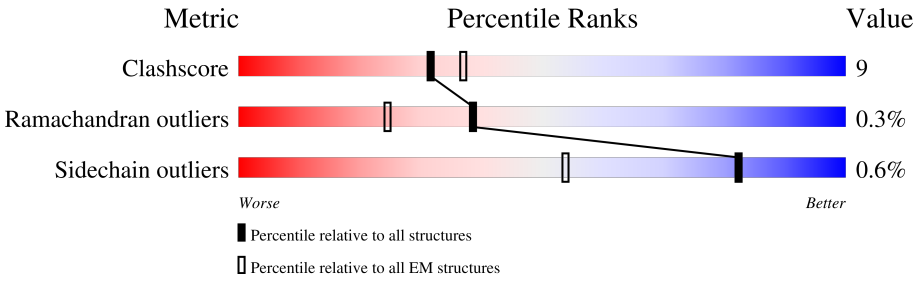
EMDB validation analysis	:	0.0.1.dev118
MolProbity	:	4-5-2 with Phenix2.0rc1
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.43.1

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 4.30 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	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415

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	4128	67% 19% 14%
2	D	24	12% 83% 17%
2	F	24	8% 33% 67%
3	G	16	6% 81% 19%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 29158 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 DNA-dependent protein kinase catalytic subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	3569	28180	18091	4776	5128	185	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*CP*AP*TP*GP*CP*TP*CP*TP*AP*CP*TP*GP*CP*TP*TP*CP*GP*AP*TP*AP*TP*CP*G)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	D	24	484	233	82	146	23	0	0
			Total	C	N	O	P		
2	F	8	164	78	30	48	8	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*AP*GP*CP*AP*GP*TP*AP*GP*AP*GP*CP*AP*TP*GP*C)-3').

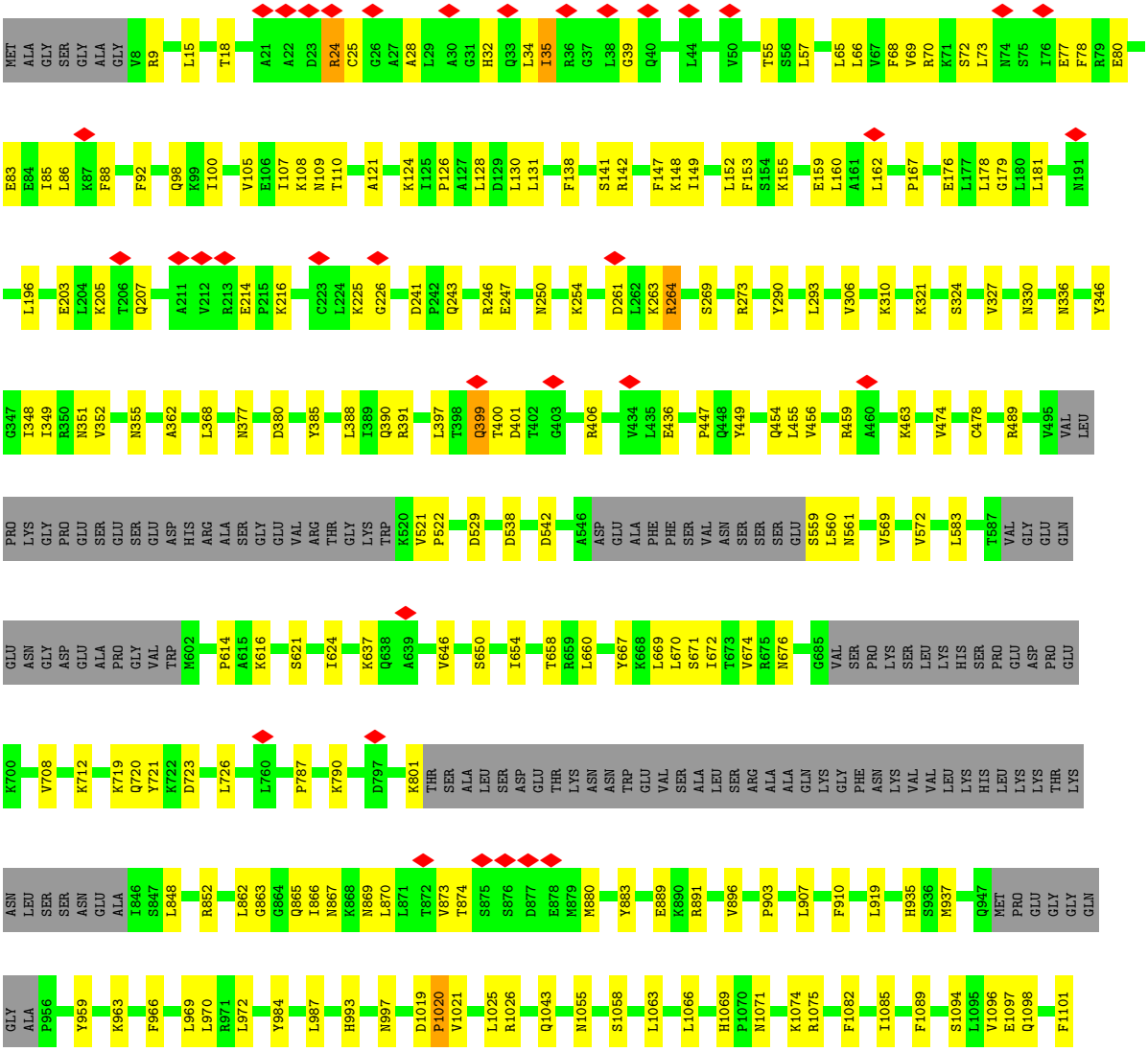
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
3	G	16	330	157	68	90	15	0	0

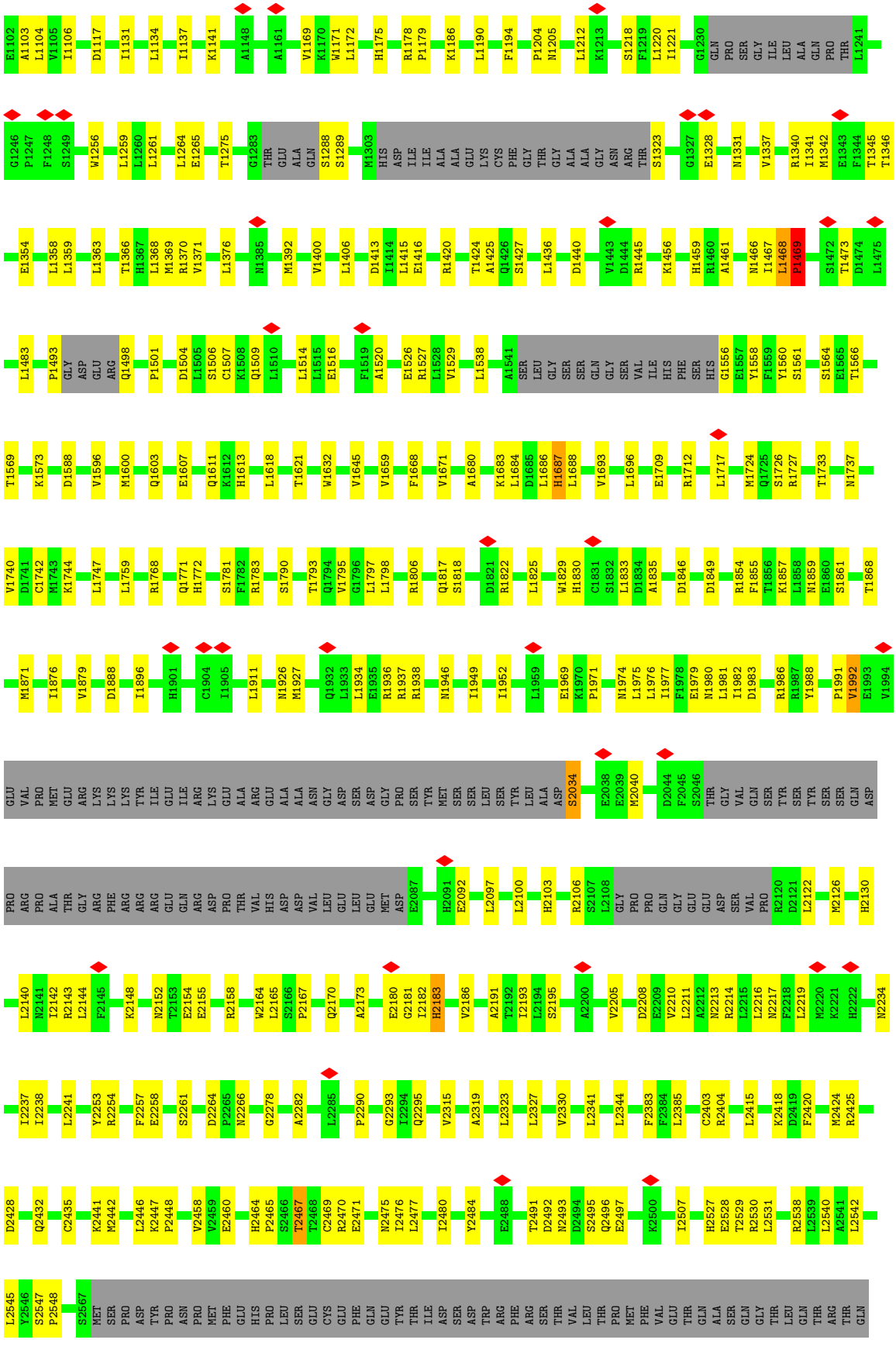
3 Residue-property plots (i)

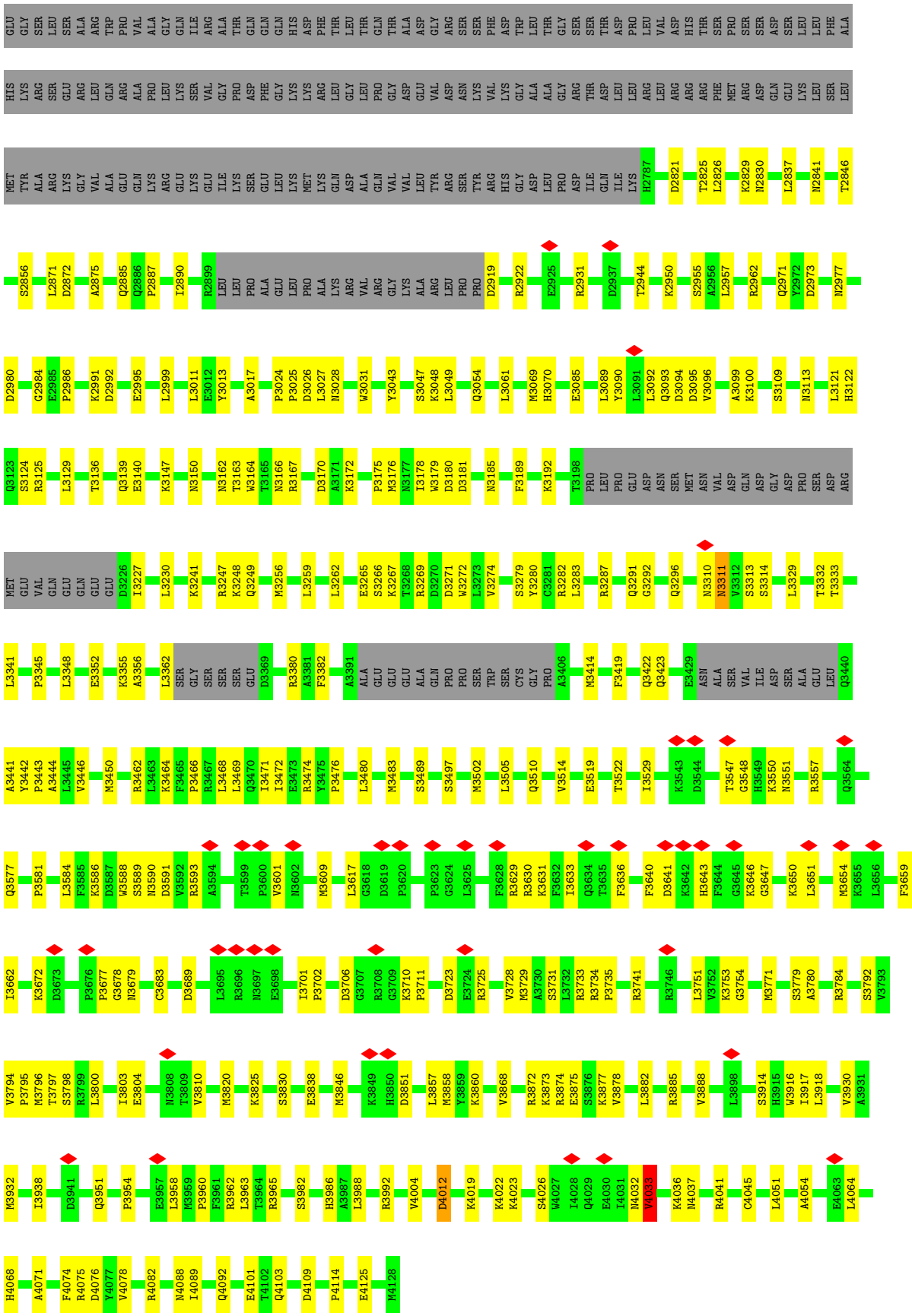
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion 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: DNA-dependent protein kinase catalytic subunit

Chain A:

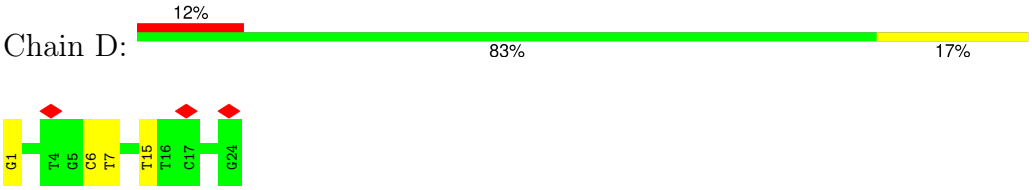



E1102 A1103 L1104 V1105 I1106 D1117 I1131 L1134 I1137 K1141 A1148 A1161 V1169 K1170 W1171 L1172 H1175 R1178 P1179 K1186 L1190 F1194 P1204 N1205 L1212 K1213 S1218 F1219 L1220 I1221 G1230 GLN PRO SER GLY ILE LEU ALA GLN PRO THR L1241
G1246 P1247 F1248 S1249 W1256 L1259 L1260 L1261 L1264 E1265 T1275 G1283 THR GLU ALA GLN S1288 S1289 M1303 HIS ASP ILE ILE ALA ALA GLU LYS CYS PHE GLY THR GLY ALA ALA GLY ASN ARG THR S1323 G1327 E1328 N1331 V1337 R1340 I1341 M1342 E1343 F1344 T1345 T1346
E1354 L1358 L1359 L1363 T1366 H1367 L1368 M1369 R1370 V1371 L1376 N1385 M1392 V1400 L1406 D1413 I1414 L1415 E1416 R1420 T1424 A1425 Q1426 S1427 L1436 D1440 V1443 D1444 R1445 K1456 H1459 R1460 A1461 N1466 I1467 L1468 P1469 S1472 T1473 D1474 L1475
L1483 P1493 GLY ASP GLU ARG Q1498 P1501 D1504 L1505 S1506 C1507 K1508 Q1509 L1510 L1514 L1515 E1516 F1519 A1520 E1526 R1527 L1528 V1529 L1538 A1541 SER LEU GLY SER SER GLN GLY SER VAL ILE HIS PHE SER HIS G1556 E1557 Y1558 F1559 Y1560 S1561 S1564 E1565 T1566
T1569 K1573 D1588 V1596 M1600 Q1603 E1607 Q1611 K1612 H1613 L1618 T1621 W1632 V1645 V1659 F1668 V1671 A1680 K1683 L1684 D1685 L1686 H1687 L1688 V1693 L1696 E1709 R1712 L1717 M1724 Q1725 S1726 R1727 T1733 N1737
V1740 D1741 C1742 M1743 K1744 L1747 L1759 R1768 Q1771 H1772 S1781 F1782 R1783 S1790 T1793 Q1794 V1795 G1796 L1797 L1798 R1806 Q1817 S1818 D1821 R1822 L1825 W1829 H1830 C1831 S1832 L1833 D1834 A1835 D1846 D1849 R1854 F1855 T1856 K1857 L1858 N1859 E1860 S1861 T1868
M1871 I1876 V1879 D1888 I1896 H1901 C1904 I1905 L1911 N1926 M1927 Q1932 L1933 L1934 E1935 R1936 R1937 R1938 N1946 I1949 I1952 L1959 E1969 K1970 P1971 N1974 L1975 L1976 I1977 F1978 E1979 N1980 L1981 I1982 D1983 R1986 R1987 Y1988 P1991 V1992 E1993 V1994
GLU VAL PRO MET GLU ARG LYS LYS LYS TYR ILE GLU ILE ARG LYS GLU ALA ARG GLU ALA ALA ASN GLY ASP SER ASP GLY PRO SER TYR MET SER SER LEU SER TYR LEU ALA ASP S2034 E2038 E2039 M2040 D2044 F2045 S2046 THR GLY VAL GLN SER TYR SER TYR SER SER GLN ASP
PRO ARG PRO ALA THR GLY ARG PHE ARG ARG ARG GLU GLN ARG ASP PRO THR VAL HIS ASP ASP VAL LEU GLU LEU GLU MET ASP E2087 H2091 E2092 L2097 L2100 H2103 R2106 S2107 L2108 GLY PRO PRO GLN GLY GLU GLU ASP SER VAL PRO R2120 D2121 L2122 M2126 H2130
L2140 N2141 I2142 R2143 L2144 F2145 K2148 N2152 T2153 E2154 E2155 R2158 W2164 L2165 S2166 P2167 Q2170 A2173 E2180 G2181 I2182 H2183 V2186 A2191 T2192 I2193 L2194 S2195 A2200 V2205 D2208 E2209 V2210 L2211 A2212 N2213 R2214 L2215 L2216 N2217 F2218 L2219 M2220 K2221 H2222 N2234
I2237 I2238 L2241 Y2253 R2254 F2257 E2258 S2261 D2264 P2265 N2266 G2278 A2282 L2285 P2290 G2293 I2294 Q2295 V2315 A2319 L2323 L2327 V2330 L2341 L2344 F2383 F2384 L2385 C2403 R2404 L2415 K2418 D2419 F2420 M2424 R2425
D2428 Q2432 C2435 K2441 M2442 L2446 K2447 P2448 V2458 V2459 E2460 H2464 P2465 S2466 T2467 T2468 C2469 R2470 E2471 N2475 I2476 L2477 I2480 Y2484 E2488 T2491 D2492 N2493 D2494 S2495 Q2496 E2497 K2500 I2507 H2527 E2528 T2529 R2530 L2531 R2538 L2539 L2540 A2541 L2542
L2545 Y2546 S2547 P2548 S2567 MET SER PRO ASP TYR PRO ASN PRO MET PHE GLU HIS PRO LEU SER GLU CYS GLU PHE GLN GLU TYR THR ILE ASP SER ASP TRP ARG PHE ARG SER THR VAL LEU THR PRO MET PHE VAL GLU THR GLN ALA SER GLN GLY THR LEU GLN THR ARG THR GLN



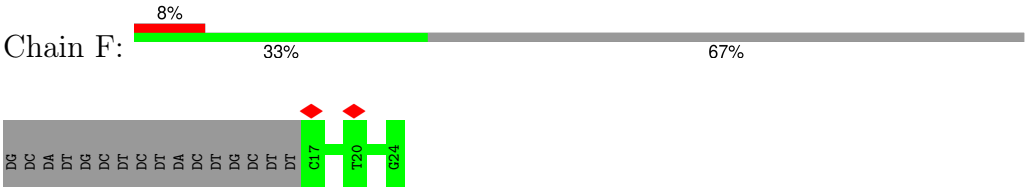
• Molecule 2: DNA (5'-D(*GP*CP*AP*TP*GP*CP*TP*CP*TP*AP*CP*TP*GP*CP*TP*TP*CP*GP*AP*TP*AP*TP*CP*G)-3')

Chain D:



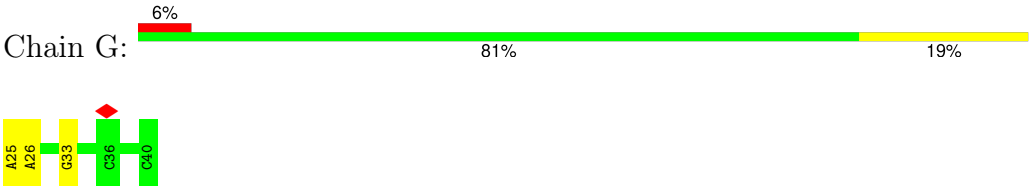
• Molecule 2: DNA (5'-D(*GP*CP*AP*TP*GP*CP*TP*CP*TP*AP*CP*TP*GP*CP*TP*TP*CP*GP*AP*TP*AP*TP*CP*G)-3')

Chain F:



• Molecule 3: DNA (5'-D(*AP*AP*GP*CP*AP*GP*TP*AP*GP*AP*GP*CP*AP*TP*GP*C)-3')

Chain G:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	62117	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 ($e^-/Å^2$)	45	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	0.059	Depositor
Minimum map value	-0.034	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.002	Depositor
Recommended contour level	0.012	Depositor
Map size (Å)	408.31998, 408.31998, 408.31998	wwPDB
Map dimensions	352, 352, 352	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.16, 1.16, 1.16	Depositor

5 Model quality

5.1 Standard geometry

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	0.29	0/28742	0.65	20/38884 (0.1%)
2	D	0.30	0/540	0.58	0/831
2	F	0.28	0/183	0.49	0/280
3	G	0.31	0/372	0.58	0/573
All	All	0.29	0/29837	0.65	20/40568 (0.0%)

There are no bond length outliers.

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	3311	ASN	N-CA-C	-22.78	82.19	112.03
1	A	1469	PRO	N-CA-C	-11.82	88.12	112.47
1	A	4033	VAL	N-CA-C	-7.62	105.88	113.20
1	A	1019	ASP	CA-C-N	7.41	129.10	119.84
1	A	1019	ASP	C-N-CA	7.41	129.10	119.84

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	28180	0	28380	493	0
2	D	484	0	274	3	0
2	F	164	0	91	0	0
3	G	330	0	180	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	29158	0	28925	495	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 495 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)
1:A:1687:HIS:CE1	1:A:1688:LEU:HD12	1.75	1.22
1:A:1687:HIS:CE1	1:A:1688:LEU:CD1	2.33	1.11
1:A:1687:HIS:ND1	1:A:1688:LEU:HD12	1.73	1.02
1:A:3647:GLY:O	1:A:3651:LEU:HB2	1.70	0.92
1:A:2467:THR:O	1:A:2471:GLU:HB2	1.77	0.84

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	3527/4128 (85%)	3067 (87%)	449 (13%)	11 (0%)	37 72

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1020	PRO
1	A	399	GLN
1	A	1992	VAL
1	A	2183	HIS
1	A	2467	THR

5.3.2 Protein sidechains

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	3094/3671 (84%)	3074 (99%)	20 (1%)	84 88

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2944	THR
1	A	3601	VAL
1	A	4033	VAL
1	A	3918	LEU
1	A	870	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 58 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	2481	HIS
1	A	3944	HIS
1	A	3093	GLN
1	A	3924	HIS
1	A	3660	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

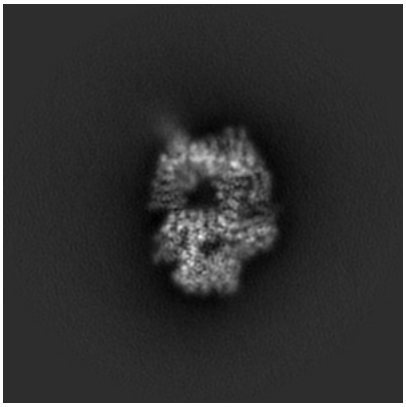
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-22622. These allow visual inspection of the internal detail of the map and identification of artifacts.

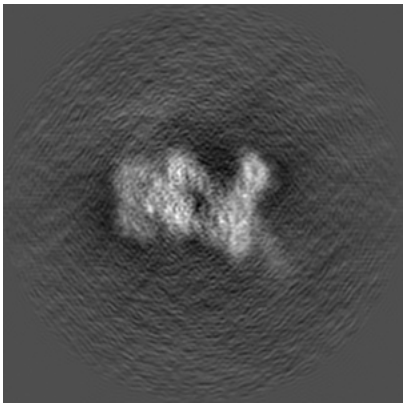
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections

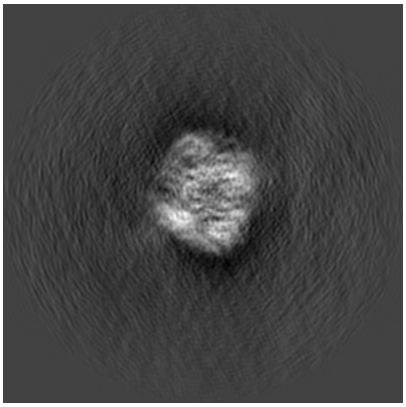
6.1.1 Primary map



X



Y

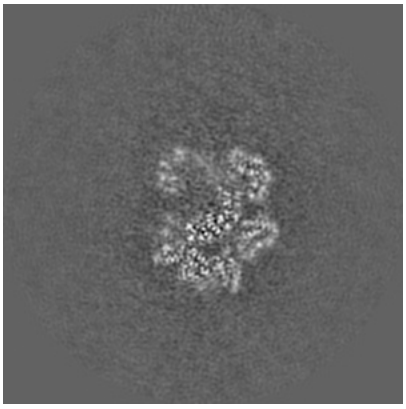


Z

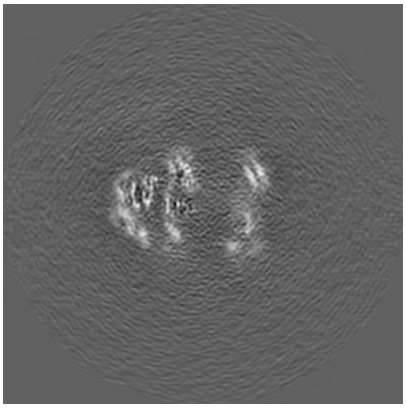
The images above show the map projected in three orthogonal directions.

6.2 Central slices

6.2.1 Primary map



X Index: 176



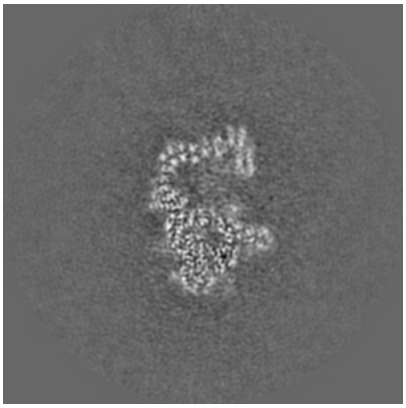
Y Index: 176

Z Index: 176

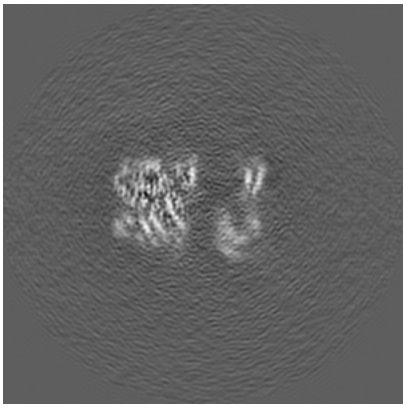
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

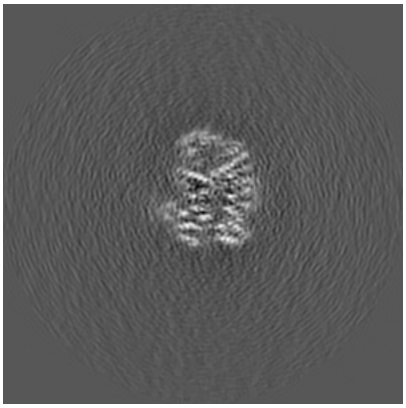
6.3.1 Primary map



X Index: 193



Y Index: 166

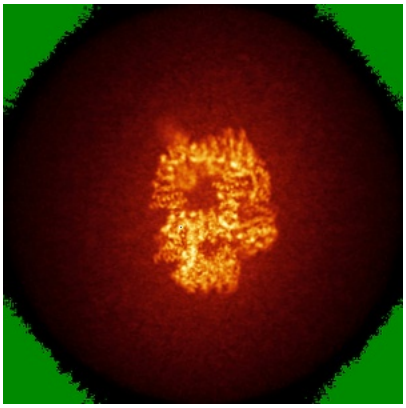


Z Index: 155

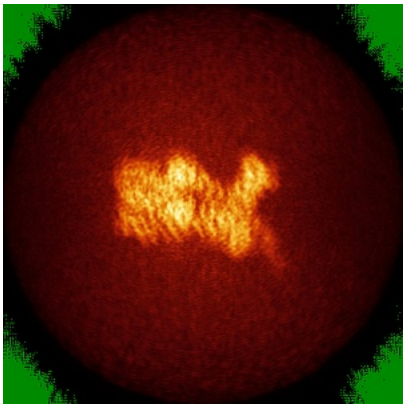
The images above show the largest variance slices of the map in three orthogonal directions.

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

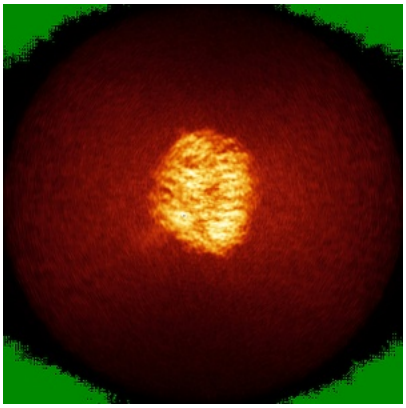
6.4.1 Primary map



X



Y

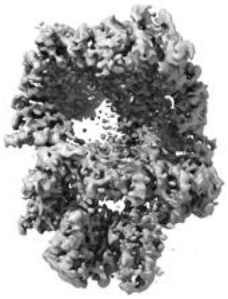


Z

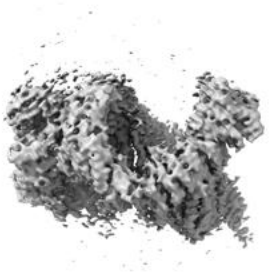
The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views (i)

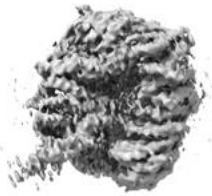
6.5.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.012. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

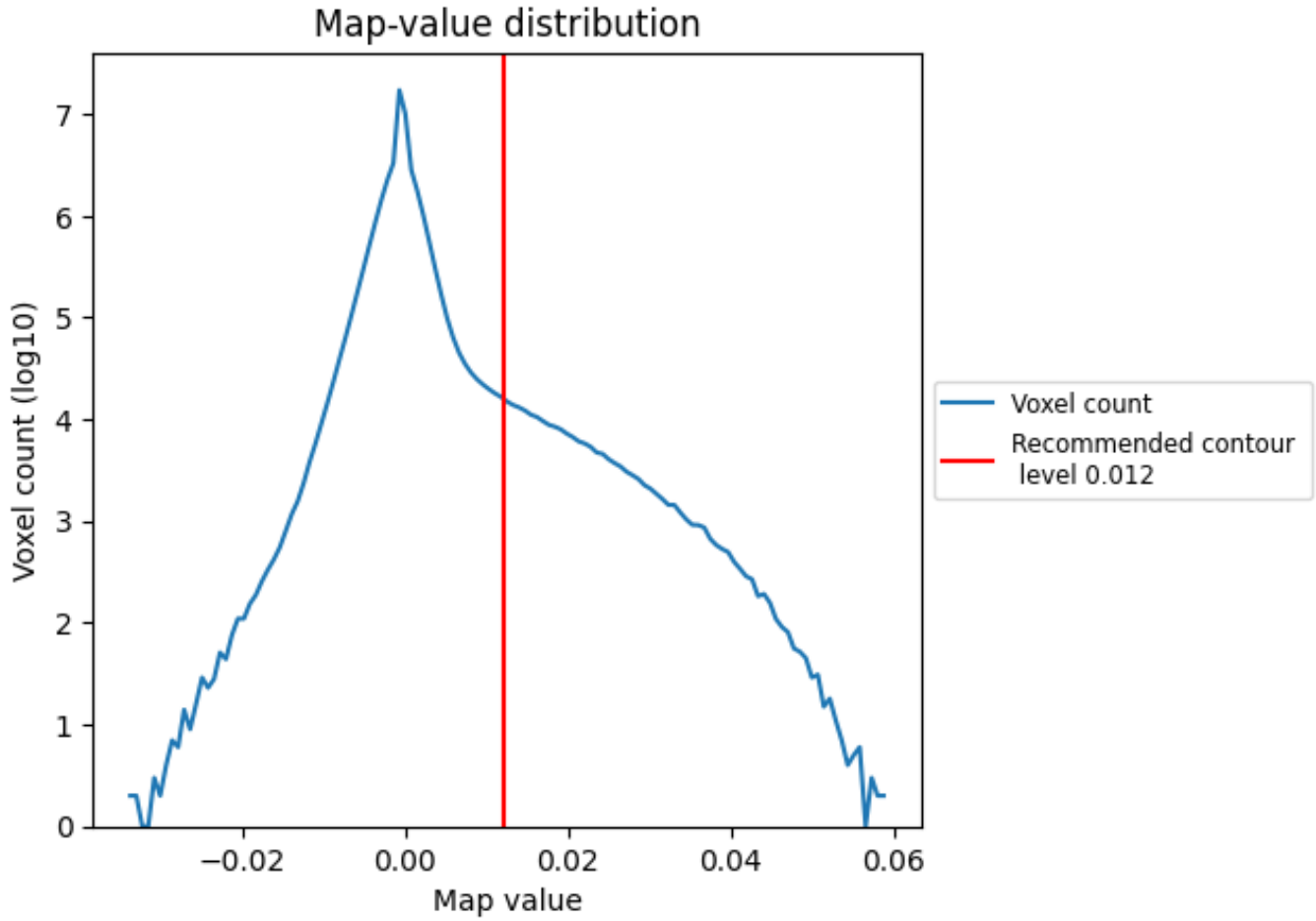
6.6 Mask visualisation (i)

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

7 Map analysis

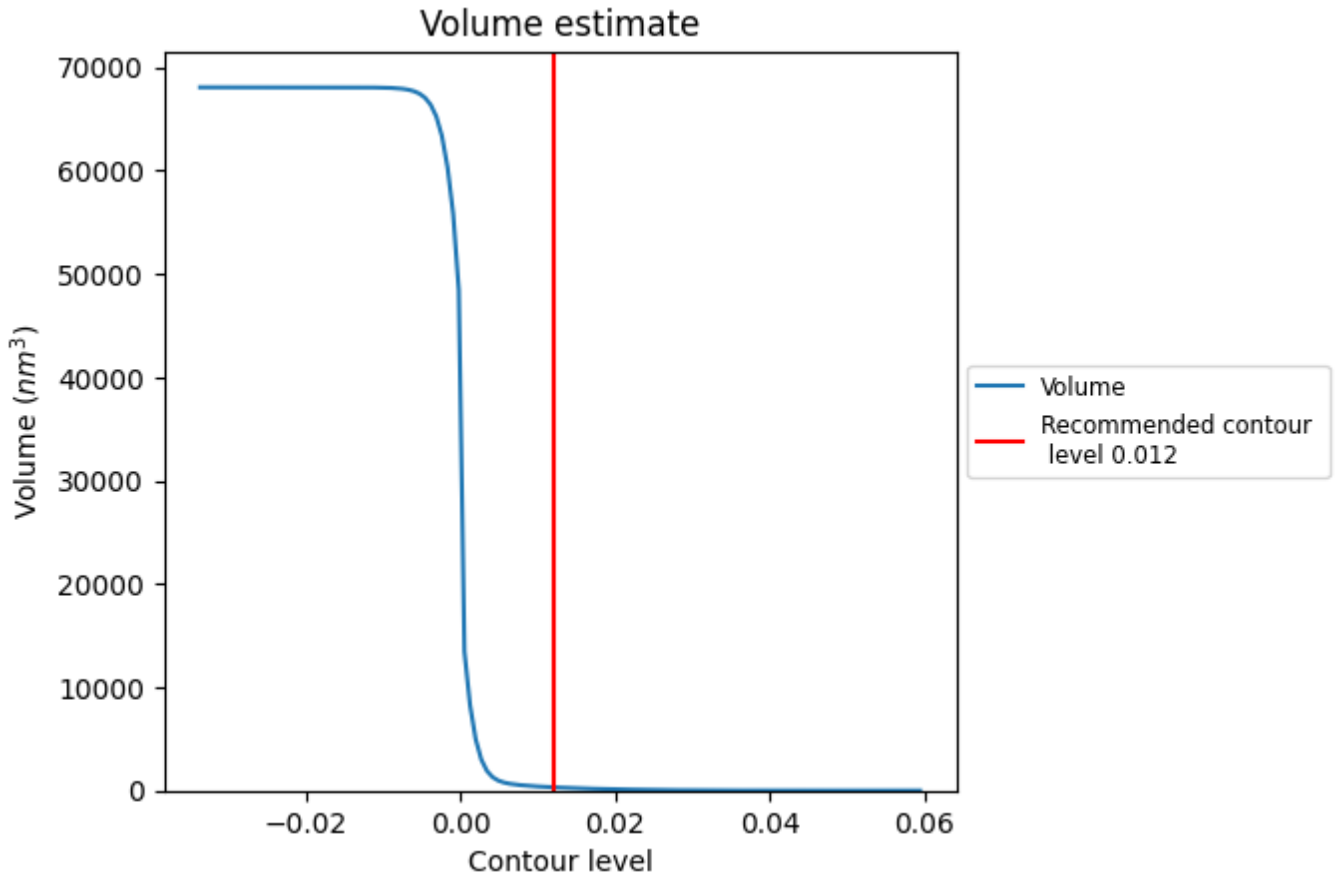
This section contains the results of statistical analysis of the map.

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.

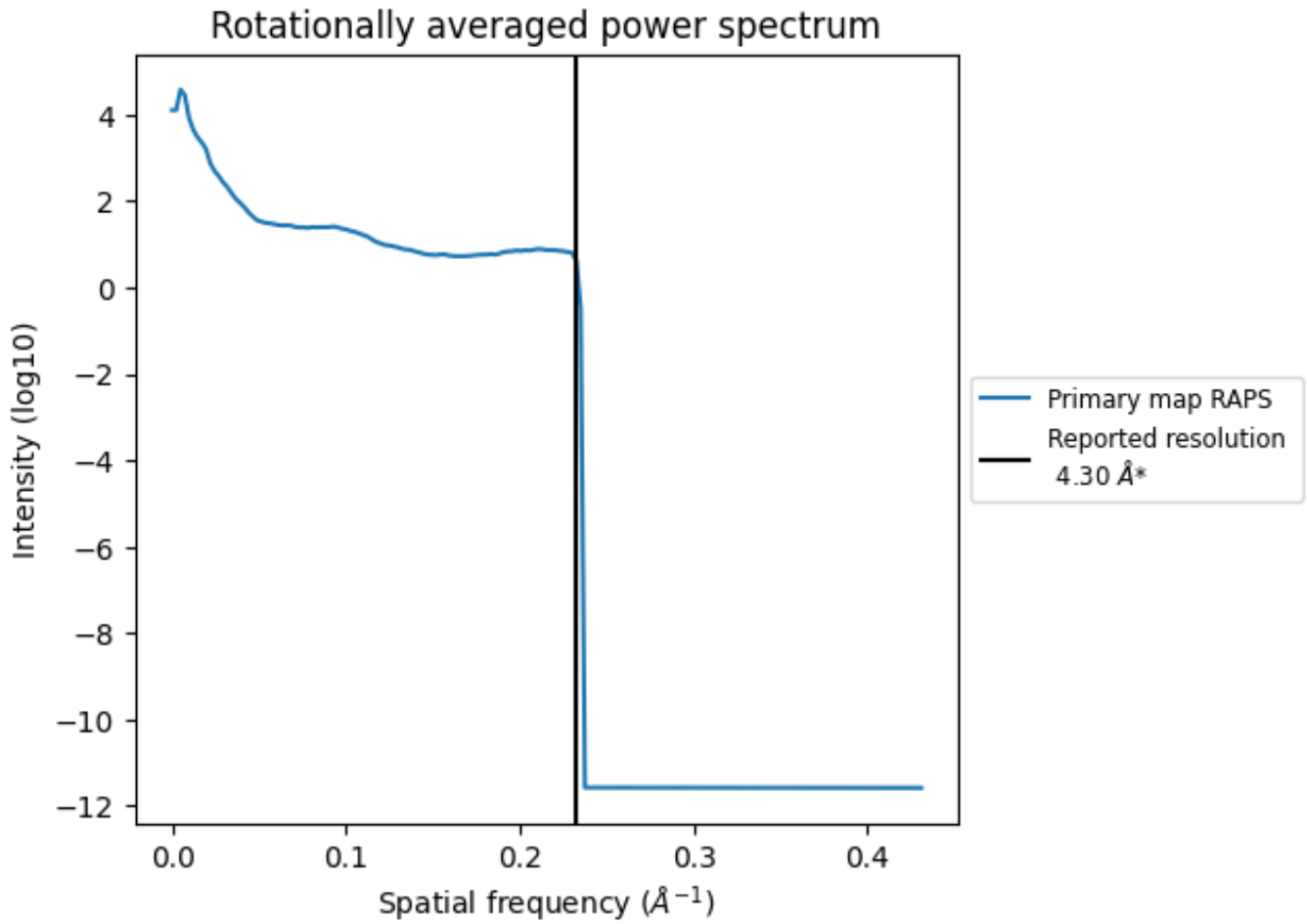
7.2 Volume estimate (i)



The volume at the recommended contour level is 317 nm^3; this corresponds to an approximate mass of 286 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 0.233 Å^{-1}

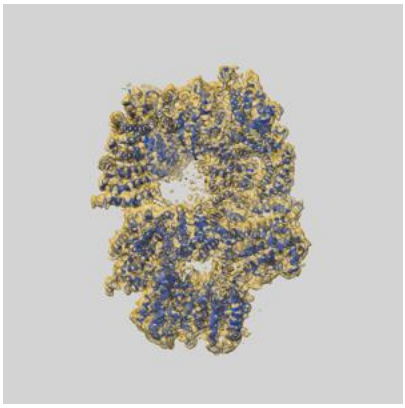
8 Fourier-Shell correlation

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

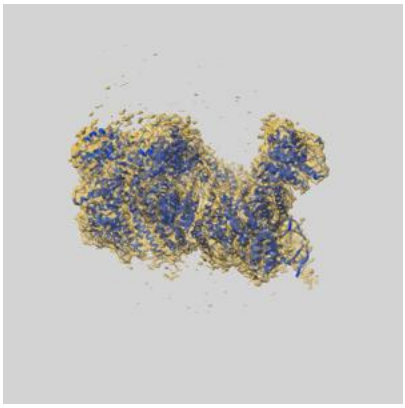
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-22622 and PDB model 7K19. Per-residue inclusion information can be found in section 3 on page 4.

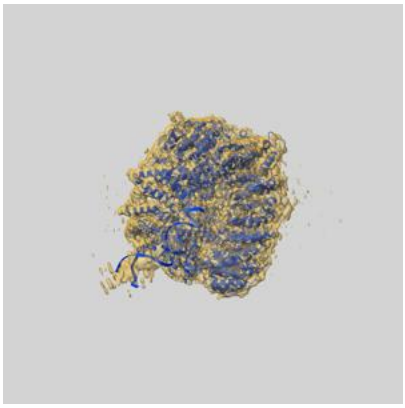
9.1 Map-model overlay



X



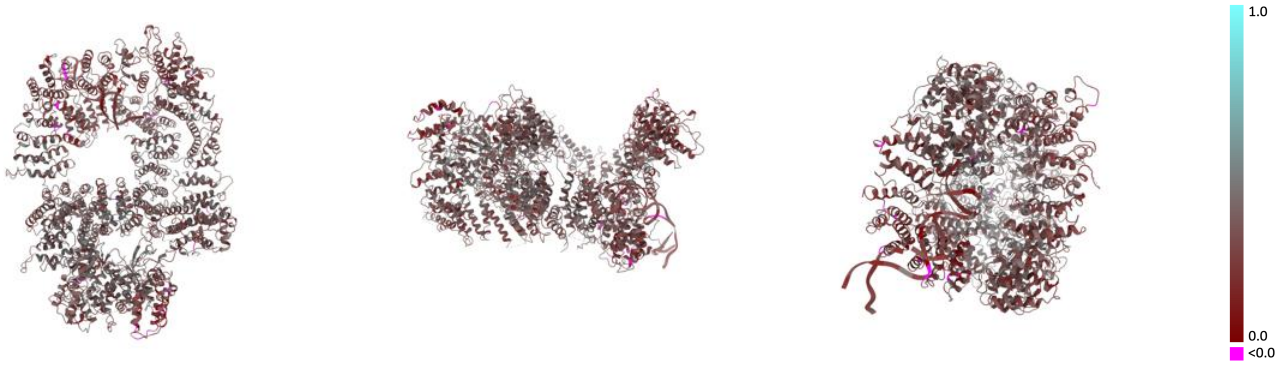
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.012 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.

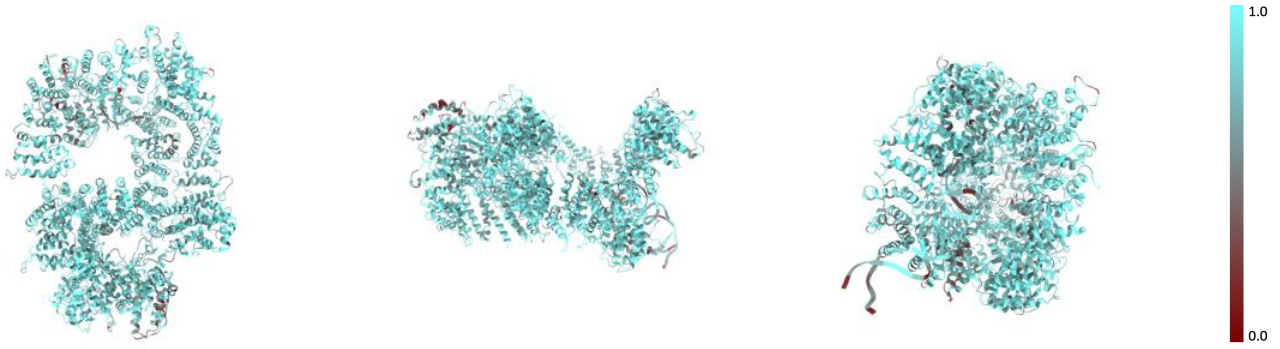
9.2 Q-score mapped to coordinate model (i)



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.

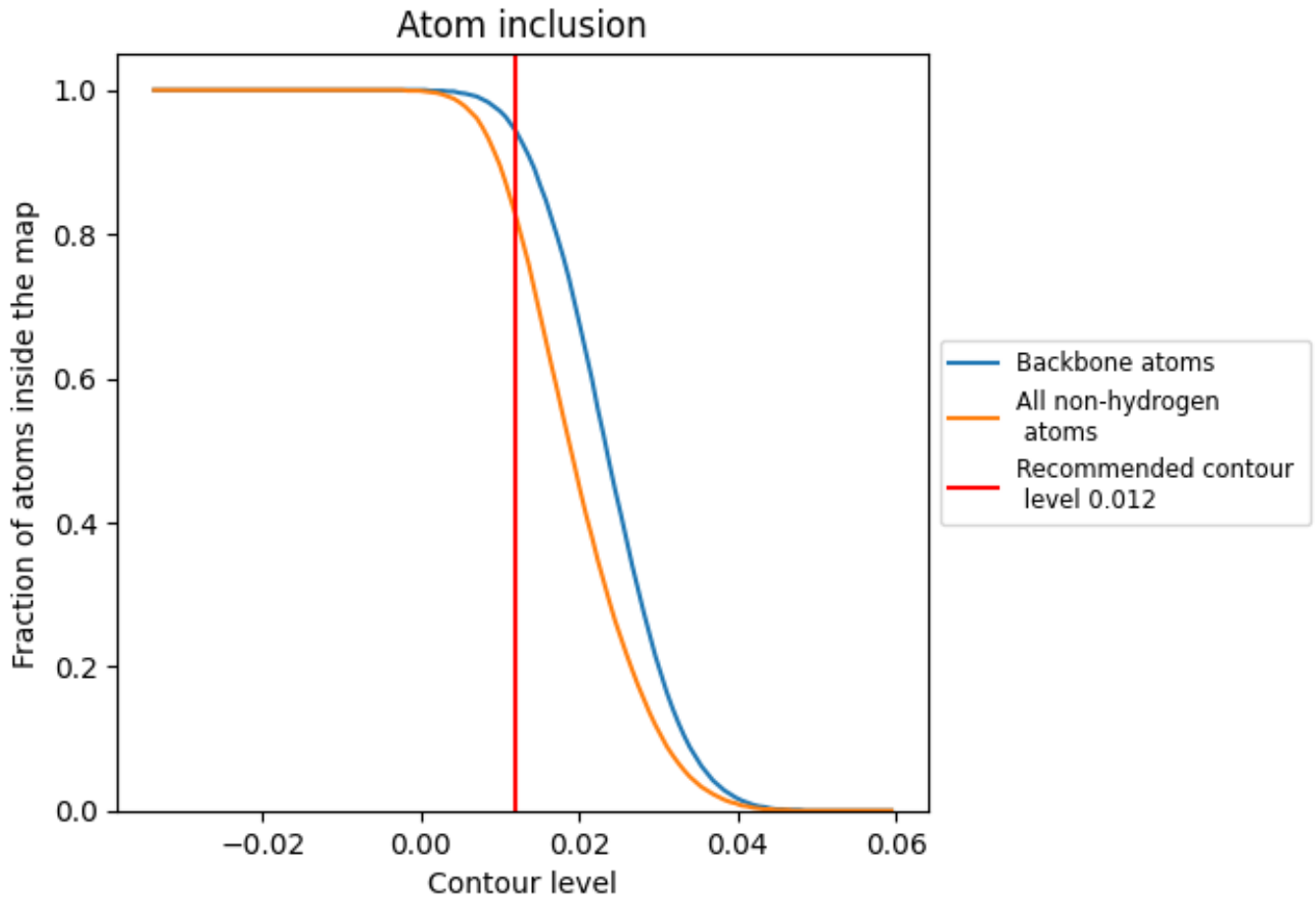
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.012).

9.4 Atom inclusion (i)



At the recommended contour level, 94% of all backbone atoms, 82% 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 (0.012) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8240	0.3390
A	0.8290	0.3430
D	0.6610	0.2000
F	0.5670	0.2560
G	0.7640	0.2420