wwPDB EM Validation Summary Report (i)

Apr 22, 2024 – 11:32 pm BST

PDB ID	:	6ZH8
EMDB ID	:	EMD-11216
Title	:	Cryo-EM structure of DNA-PKcs:DNA
Authors	:	Chaplin, A.K.; Hardwick, S.W.; Chirgadze, D.Y.; Blundell, T.L.
Deposited on	:	2020-06-21
Resolution	:	4.14 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 (i) 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 (i)) were used in the production of this report:

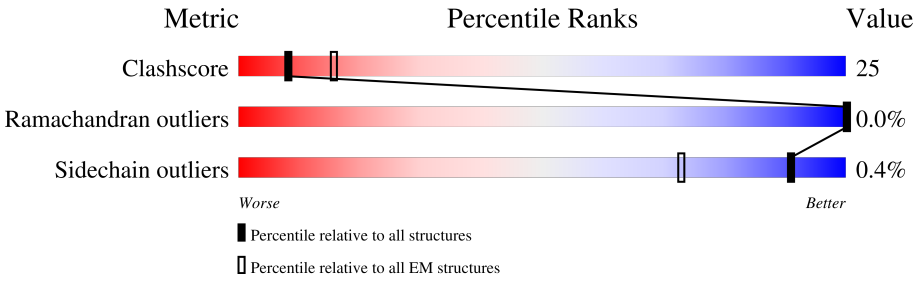
EMDB validation analysis	:	0.0.1.dev92
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 4.14 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	4156	48% 40% 11%
2	B	11	27% 73%
3	C	8	25% 75%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 29506 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,DNA-PKcs.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	3694	29120	18668	4926	5333	193	0	0

- Molecule 2 is a DNA chain called DNA (5'-D(P*AP*GP*TP*TP*TP*TP*TP*AP*GP*TP *T)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	B	11	222	106	34	71	11	0	0

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

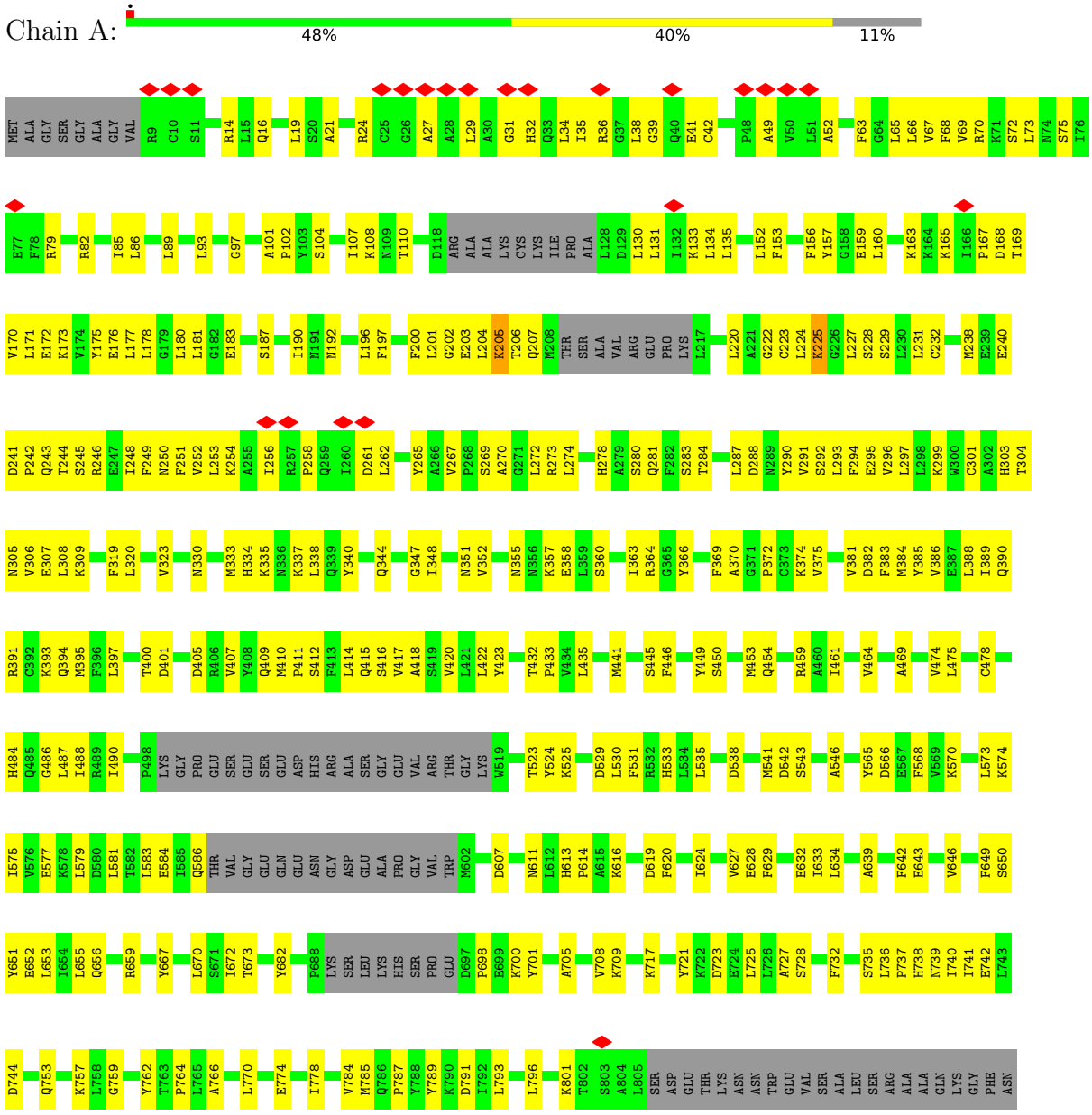
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
3	C	8	164	78	35	43	8	0	0

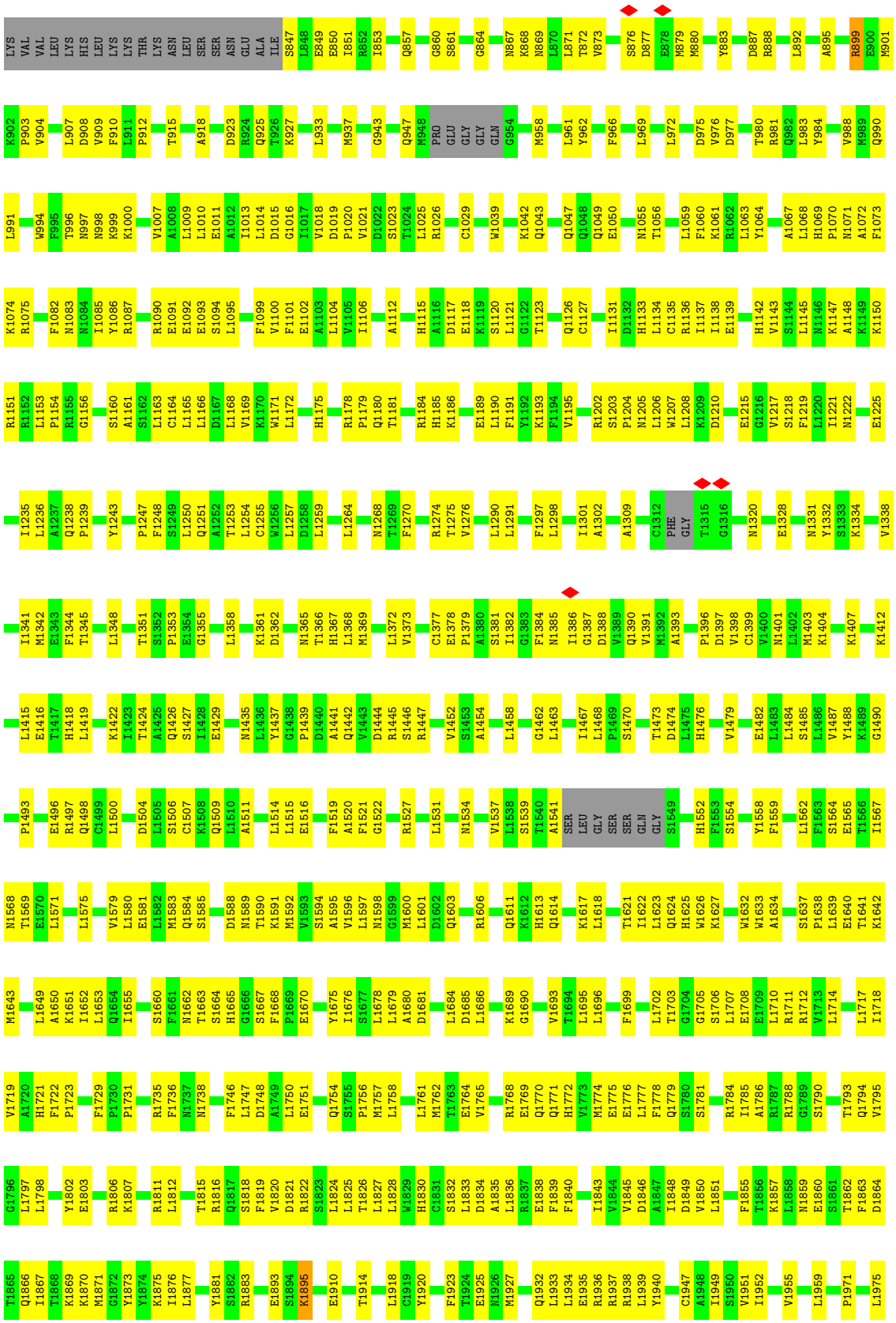
3 Residue-property plots (i)

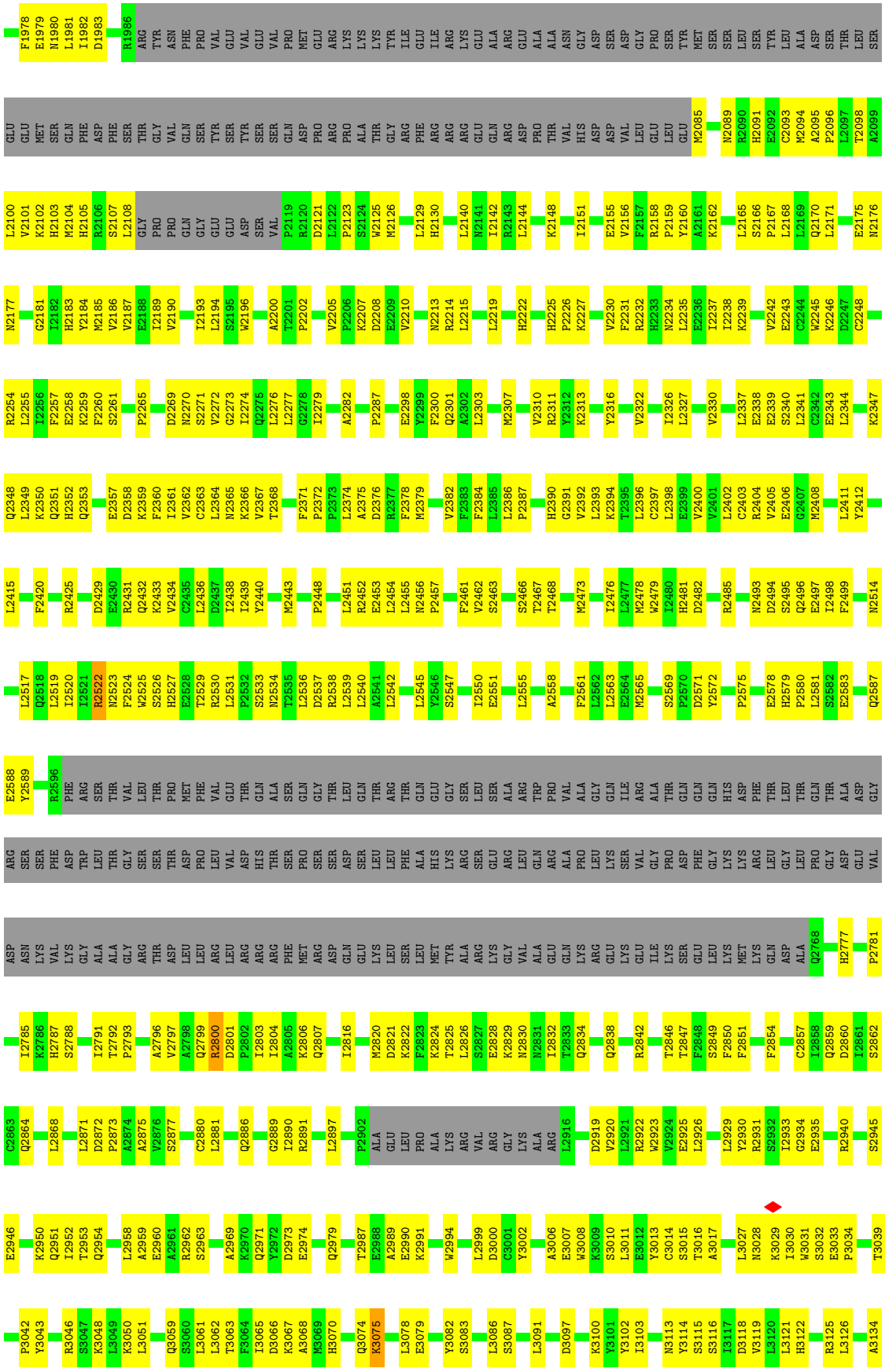
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion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

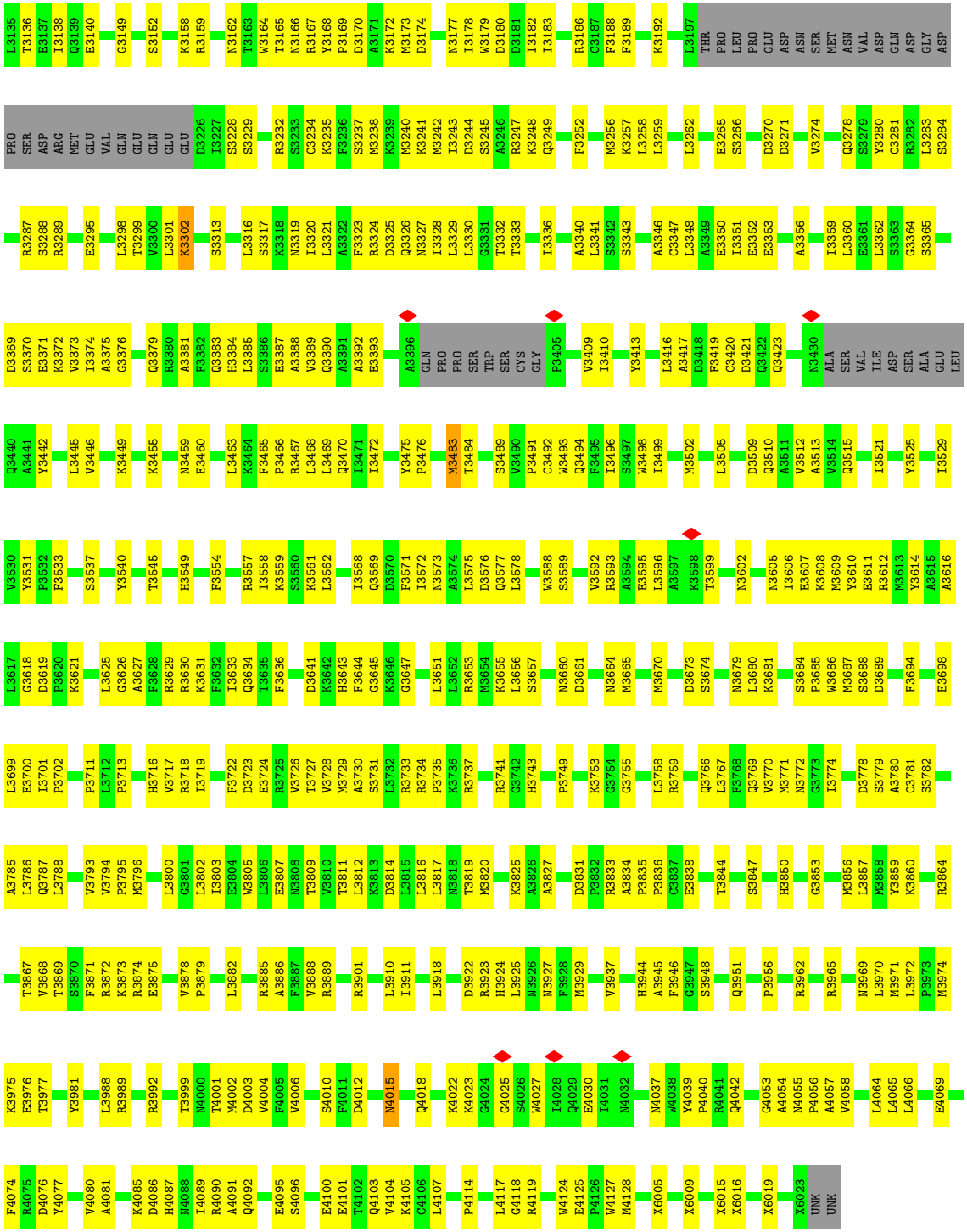
- Molecule 1: DNA-dependent protein kinase catalytic subunit,DNA-PKcs

Chain A:




LYS VAL VAL LEU LYS HIS LEU LYS LYS THR LYS ASN LEU SER SER ASN GLU ALA ILE S847 L848 E849 E850 I851 R852 I853 Q857 G860 S861 G864 N867 K868 N869 L870 L871 T872 V873 S876 D877 E878 M879 M880 Y883 D887 R888 L892 A895 R899 E900 M901
K902 P903 V904 L907 D908 V909 F910 L911 P912 T915 A918 D923 R924 Q925 T926 K927 L933 M937 G943 Q947 M948 PRO GLU GLY GLY GLN G954 M958 L961 Y962 F966 L969 L972 D975 V976 D977 T980 R981 Q982 L983 Y984 V988 M989 Q990
L991 W994 F996 T996 N997 N998 K999 K1000 V1007 A1008 L1009 L1010 E1011 A1012 I1013 L1014 D1015 G1016 I1017 V1018 D1019 P1020 V1021 D1022 S1023 T1024 L1025 R1026 C1029 W1039 K1042 Q1043 Q1047 Q1048 Q1049 E1050 N1055 T1056 L1059 F1060 K1061 R1062 L1063 Y1064 A1067 L1068 H1069 P1070 N1071 A1072 F1073
K1074 R1075 F1082 N1083 N1084 I1085 Y1086 R1087 R1090 E1091 E1092 E1093 S1094 L1095 F1099 V1100 F1101 E1102 A1103 L1104 V1105 I1106 A1112 H1115 A1116 D1117 E1118 K1119 S1120 L1121 G1122 T1123 Q1126 C1127 I1131 D1132 H1133 L1134 C1135 R1136 I1137 I1138 E1139 H1142 V1143 S1144 L1145 N1146 K1147 A1148 K1149 K1150
R1151 R1152 L1153 P1154 R1155 G1156 S1160 A1161 S1162 L1163 C1164 L1165 L1166 D1167 L1168 V1169 K1170 W1171 L1172 H1175 R1178 P1179 Q1180 T1181 R1184 H1185 K1186 E1189 L1190 F1191 Y1192 K1193 F1194 V1195 R1202 S1203 P1204 N1205 L1206 W1207 L1208 K1209 D1210 E1215 G1216 V1217 S1218 F1219 L1220 I1221 N1222 E1225
I1235 L1236 A1237 Q1238 P1239 Y1243 P1247 F1248 S1249 L1250 Q1251 A1252 T1253 L1254 C1255 W1256 L1257 D1258 L1259 L1264 N1268 T1269 F1270 R1274 T1275 V1276 L1290 L1291 F1297 L1298 I1301 A1302 A1309 C1312 PHE GLY T1315 G1316 N1320 E1328 N1331 Y1332 S1333 K1334 V1338
I1341 M1342 E1343 F1344 T1345 L1348 T1351 S1352 P1353 E1354 G1355 L1358 K1361 D1362 N1365 T1366 H1367 L1368 M1369 L1372 V1373 C1377 E1378 P1379 A1380 S1381 I1382 C1383 F1384 N1385 I1386 G1387 D1388 V1389 Q1390 V1391 M1392 A1393 P1396 D1397 V1398 C1399 N1400 N1401 L1402 M1403 K1404 K1407 K1412
L1415 E1416 T1417 H1418 L1419 K1422 I1423 T1424 A1425 Q1426 S1427 I1428 E1429 N1435 L1436 Y1437 G1438 P1439 D1440 A1441 Q1442 V1443 D1444 R1445 S1446 R1447 V1452 S1453 A1454 L1458 G1462 L1463 I1467 L1468 P1469 S1470 T1473 D1474 L1475 H1476 V1479 E1482 L1483 L1484 S1485 L1486 V1487 Y1488 K1489 G1490
P1493 E1496 R1497 Q1498 C1499 L1500 D1504 L1505 S1506 C1507 K1508 Q1509 L1510 A1511 L1514 L1515 E1516 F1519 A1520 F1521 G1522 R1527 L1531 N1534 V1537 L1538 S1539 T1540 A1541 SER LEU GLY SER SER GLN GLY S1549 H1552 F1553 S1554 Y1558 F1559 L1562 F1563 S1564 E1565 T1566 I1567
N1568 T1569 E1570 L1571 L1575 V1579 L1580 E1581 L1582 M1583 Q1584 S1585 D1588 N1589 T1590 K1591 M1592 V1593 S1594 A1595 V1596 L1597 N1598 G1599 M1600 L1601 D1602 Q1603 R1606 Q1611 K1612 H1613 Q1614 K1617 L1618 T1621 I1622 L1623 Q1624 H1625 W1626 K1627 W1632 W1633 A1634 S1637 P1638 L1639 E1640 T1641 K1642
M1643 L1649 A1650 K1651 I1652 L1653 Q1654 I1655 S1660 F1661 N1662 T1663 S1664 H1665 G1666 S1667 F1668 P1669 E1670 Y1675 I1676 S1677 L1678 L1679 A1680 D1681 L1684 D1685 L1686 K1689 G1690 V1693 T1694 L1695 L1696 F1699 L1702 T1703 G1704 G1705 S1706 L1707 E1708 E1709 L1710 R1711 R1712 V1713 L1714 L1717 I1718
V1719 A1720 H1721 F1722 P1723 F1729 P1730 P1731 R1735 F1736 N1737 N1738 F1746 L1747 D1748 A1749 L1750 E1751 Q1754 S1755 P1756 M1757 L1758 L1761 M1762 T1763 E1764 V1765 R1768 E1769 Q1770 Q1771 H1772 V1773 M1774 E1775 E1776 L1777 F1778 Q1779 S1780 S1781 R1784 I1785 A1786 R1787 R1788 G1789 S1790 T1793 Q1794 V1795
G1796 L1797 L1798 Y1802 E1803 R1806 K1807 R1811 L1812 T1815 R1816 Q1817 S1818 F1819 V1820 D1821 R1822 S1823 L1824 L1825 T1826 L1827 L1828 W1829 H1830 G1831 S1832 L1833 D1834 A1835 L1836 R1837 E1838 F1839 F1840 I1843 V1844 V1845 D1846 A1847 I1848 D1849 V1850 L1851 F1855 T1856 K1857 L1858 N1859 E1860 S1861 T1862 F1863 D1864
T1865 Q1866 I1867 T1868 K1869 K1870 M1871 G1872 Y1873 Y1874 K1875 I1876 L1877 Y1881 S1882 R1883 E1893 S1894 K1895 E1910 T1914 L1918 C1919 Y1920 F1923 T1924 E1925 N1926 M1927 Q1932 L1933 L1934 E1935 R1936 R1937 R1938 L1939 Y1940 C1947 A1948 I1949 S1950 V1951 I1952 V1955 L1959 P1971 L1975





• Molecule 2: DNA (5'-D(P*AP*GP*TP*TP*TP*TP*TP*AP*GP*TP*T)-3')

Chain B: 27% 73%

• Molecule 3: DNA (5'-D(P*AP*CP*TP*AP*AP*AP*AP*A)-3')

Chain C:

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	31881	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 ($e^-/Å^2$)	52.97	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	0.445	Depositor
Minimum map value	-0.076	Depositor
Average map value	0.002	Depositor
Map value standard deviation	0.022	Depositor
Recommended contour level	0.125	Depositor
Map size (Å)	356.99997, 356.99997, 356.99997	wwPDB
Map dimensions	340, 340, 340	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.05, 1.05, 1.05	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	0.32	0/29582	0.47	0/40029
2	B	0.60	0/247	1.15	0/378
3	C	0.64	0/185	0.87	0/282
All	All	0.33	0/30014	0.49	0/40689

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	29120	0	29078	1353	0
2	B	222	0	117	10	0
3	C	164	0	87	6	0
All	All	29506	0	29282	1369	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 25.

The worst 5 of 1369 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:3472:ILE:HG21	1:A:3483:MET:CE	1.65	1.25

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:3472:ILE:HG21	1:A:3483:MET:HE1	1.25	1.11
1:A:3472:ILE:CG2	1:A:3483:MET:CE	2.32	1.06
1:A:3472:ILE:CG2	1:A:3483:MET:HE1	1.91	0.98
1:A:1351:THR:HG22	1:A:1353:PRO:HD2	1.48	0.95

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	3634/4156 (87%)	3246 (89%)	387 (11%)	1 (0%)	100 100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1895	LYS

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	3172/3671 (86%)	3158 (100%)	14 (0%)	91 94

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3075	LYS
1	A	3166	ASN
1	A	4015	ASN
1	A	3483	MET
1	A	3653	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	2352	HIS
1	A	2951	GLN
1	A	3772	ASN
1	A	2807	GLN
1	A	3379	GLN

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)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	4128:MET	C	5011:UNK	N	92.87
1	A	5016:UNK	C	6004:UNK	N	47.75

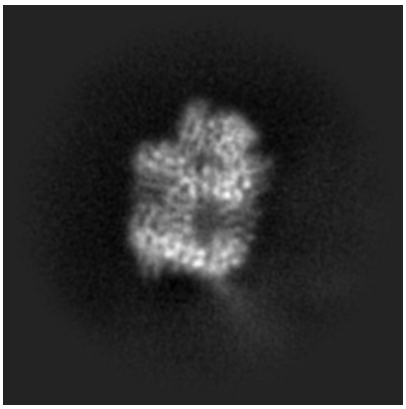
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-11216. These allow visual inspection of the internal detail of the map and identification of artifacts.

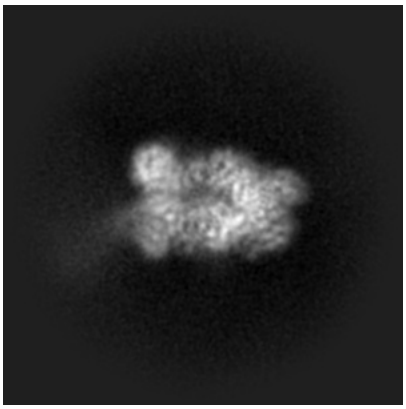
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections

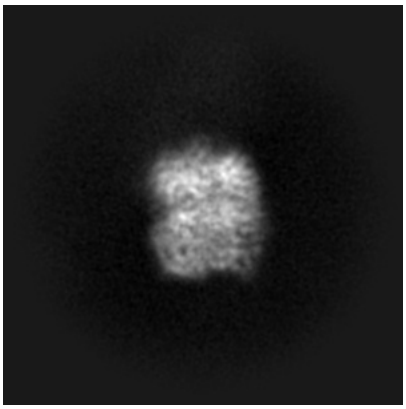
6.1.1 Primary map



X

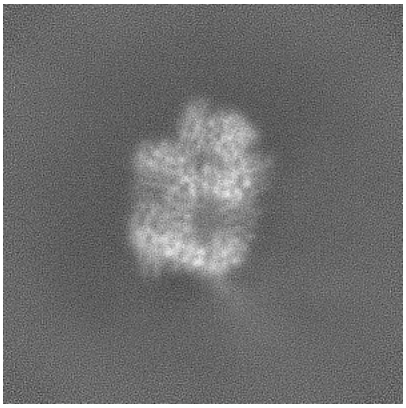


Y

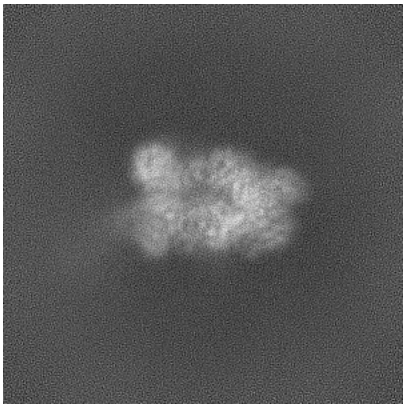


Z

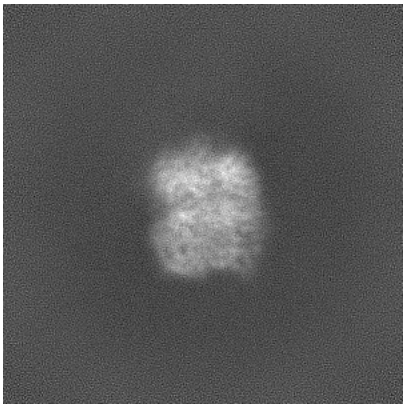
6.1.2 Raw map



X



Y

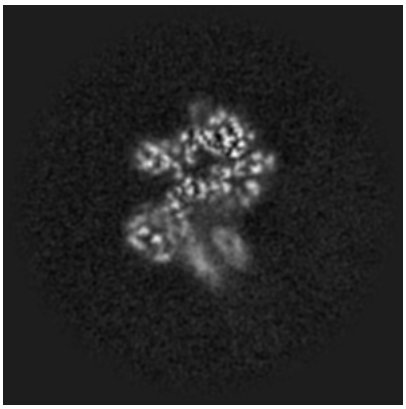


Z

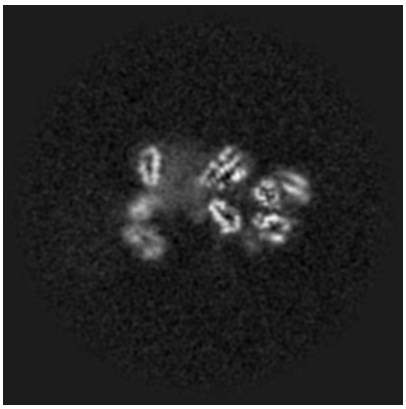
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

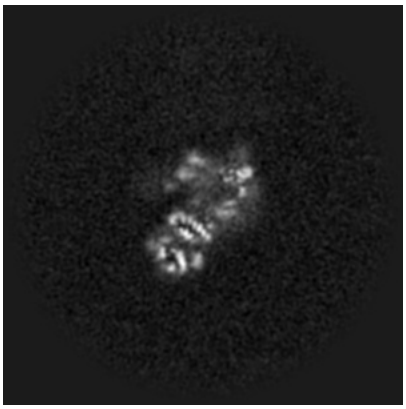
6.2.1 Primary map



X Index: 170

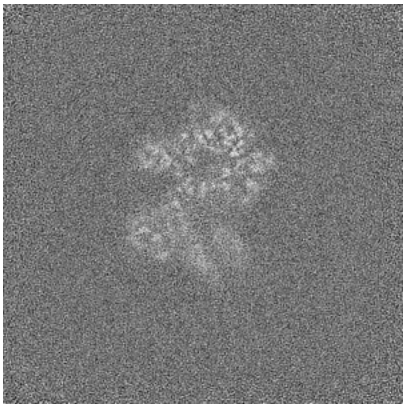


Y Index: 170

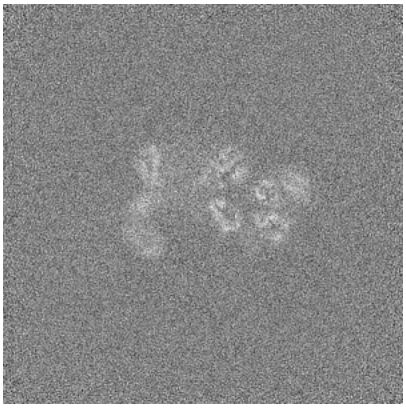


Z Index: 170

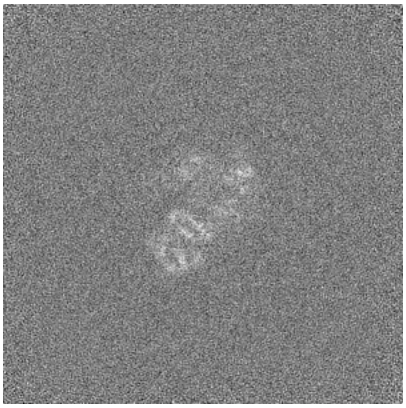
6.2.2 Raw map



X Index: 170



Y Index: 170

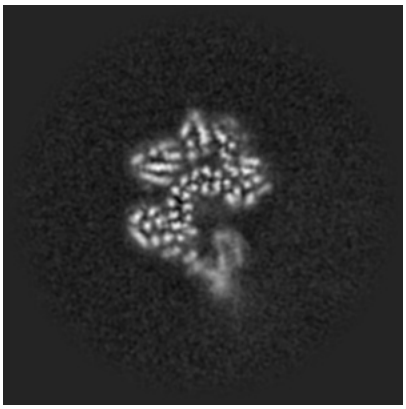


Z Index: 170

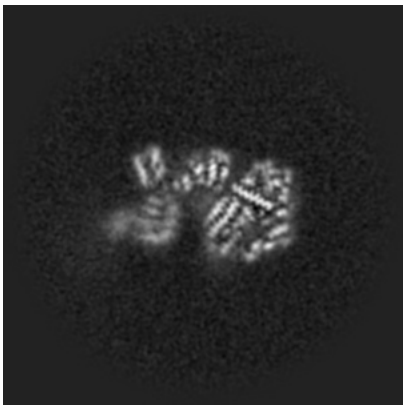
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

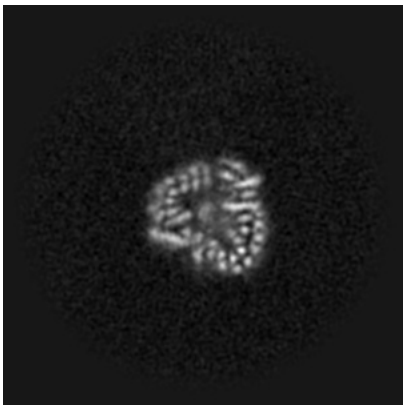
6.3.1 Primary map



X Index: 158

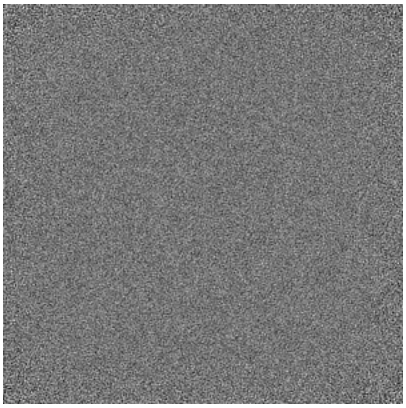


Y Index: 190

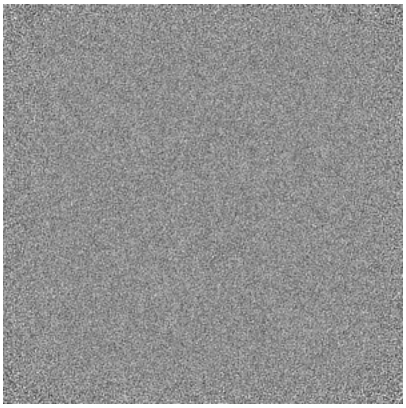


Z Index: 129

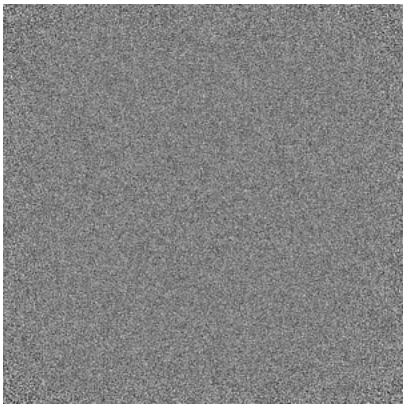
6.3.2 Raw map



X Index: 0



Y Index: 0

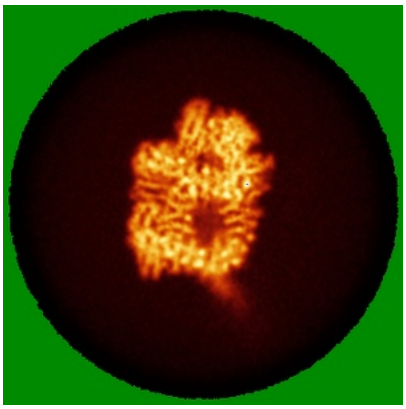


Z Index: 0

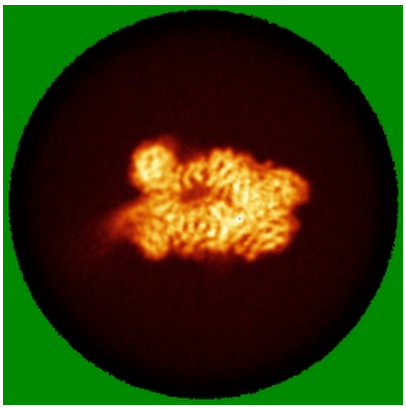
The images above show the largest variance slices of the map in three orthogonal directions.

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

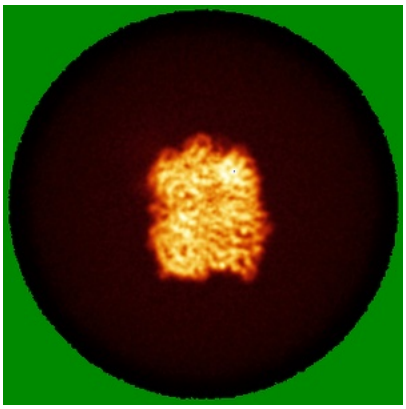
6.4.1 Primary map



X

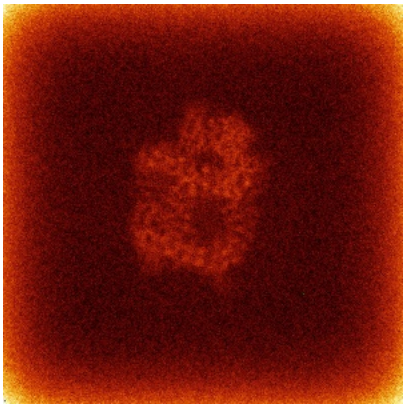


Y

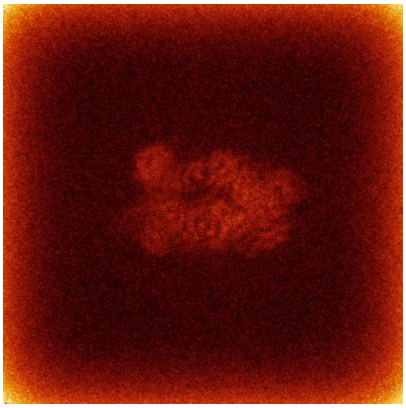


Z

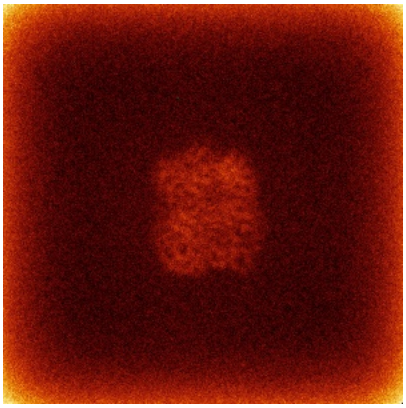
6.4.2 Raw map



X



Y

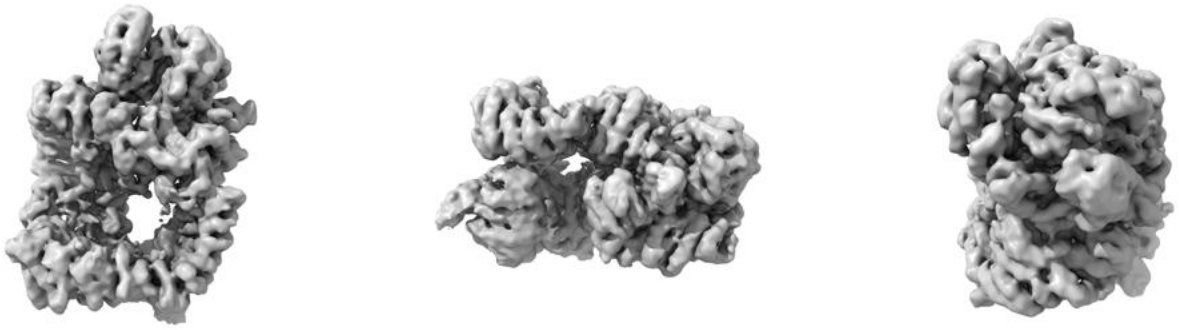


Z

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)

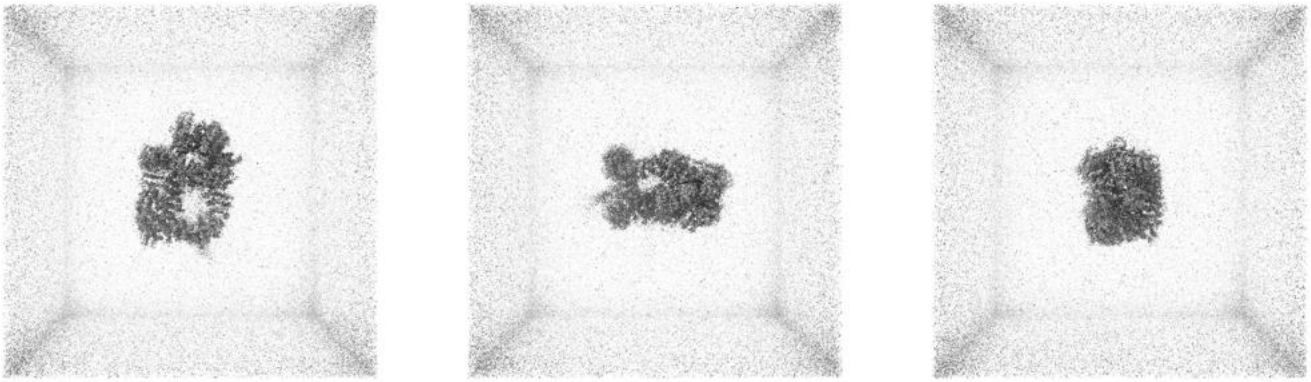
6.5.1 Primary map



X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.125. 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.

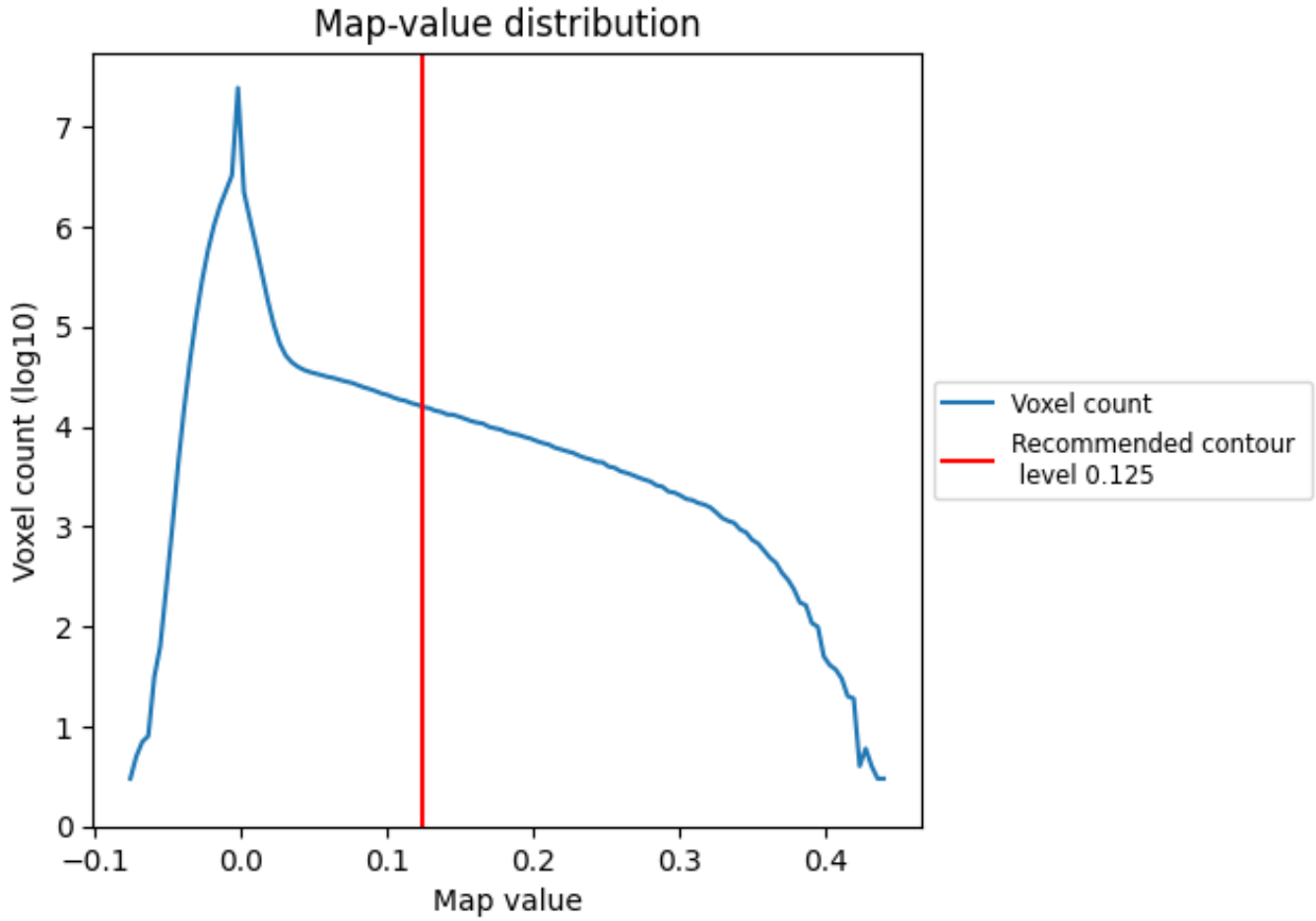
6.6 Mask visualisation (i)

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

7 Map analysis

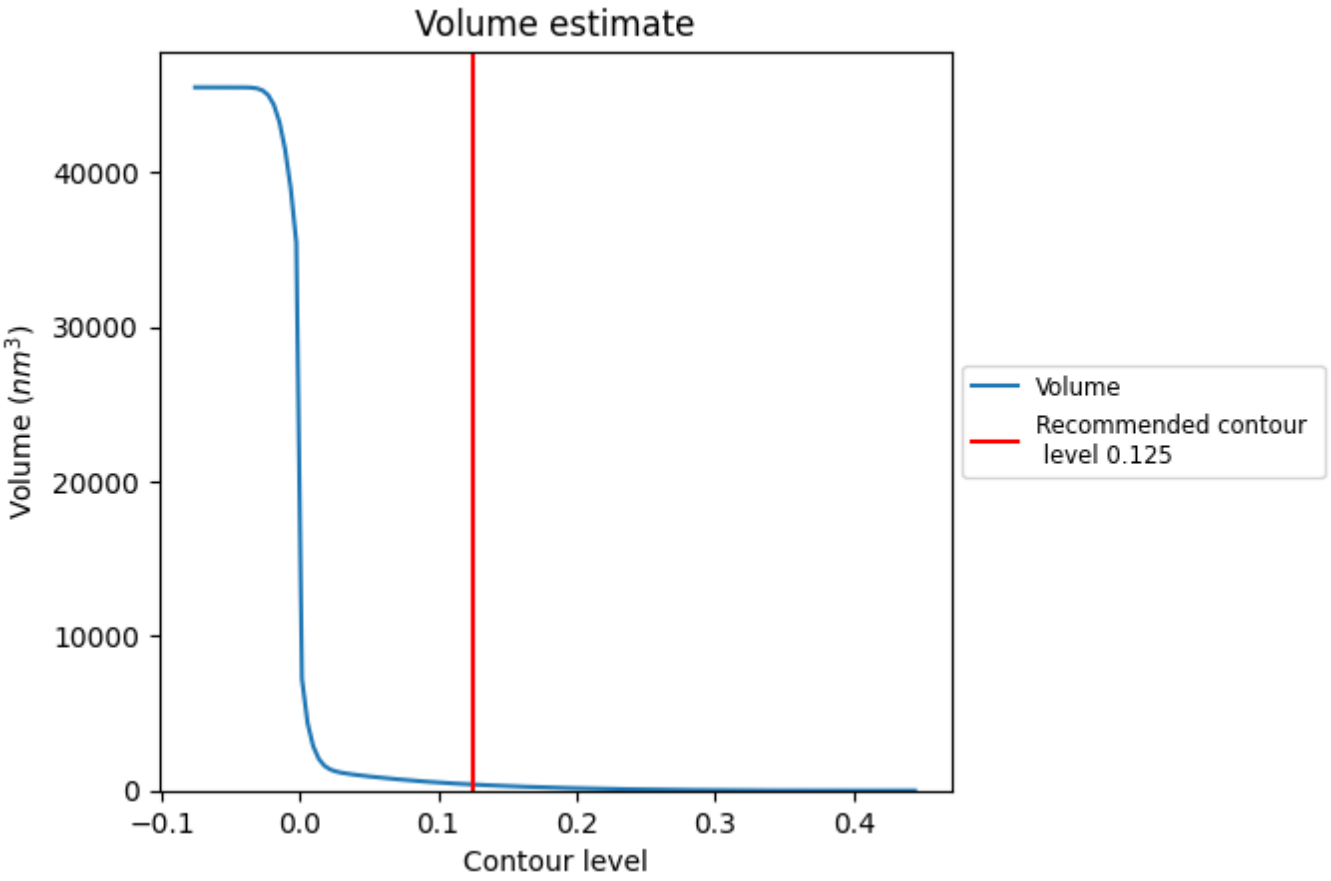
This section contains the results of statistical analysis of the map.

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.

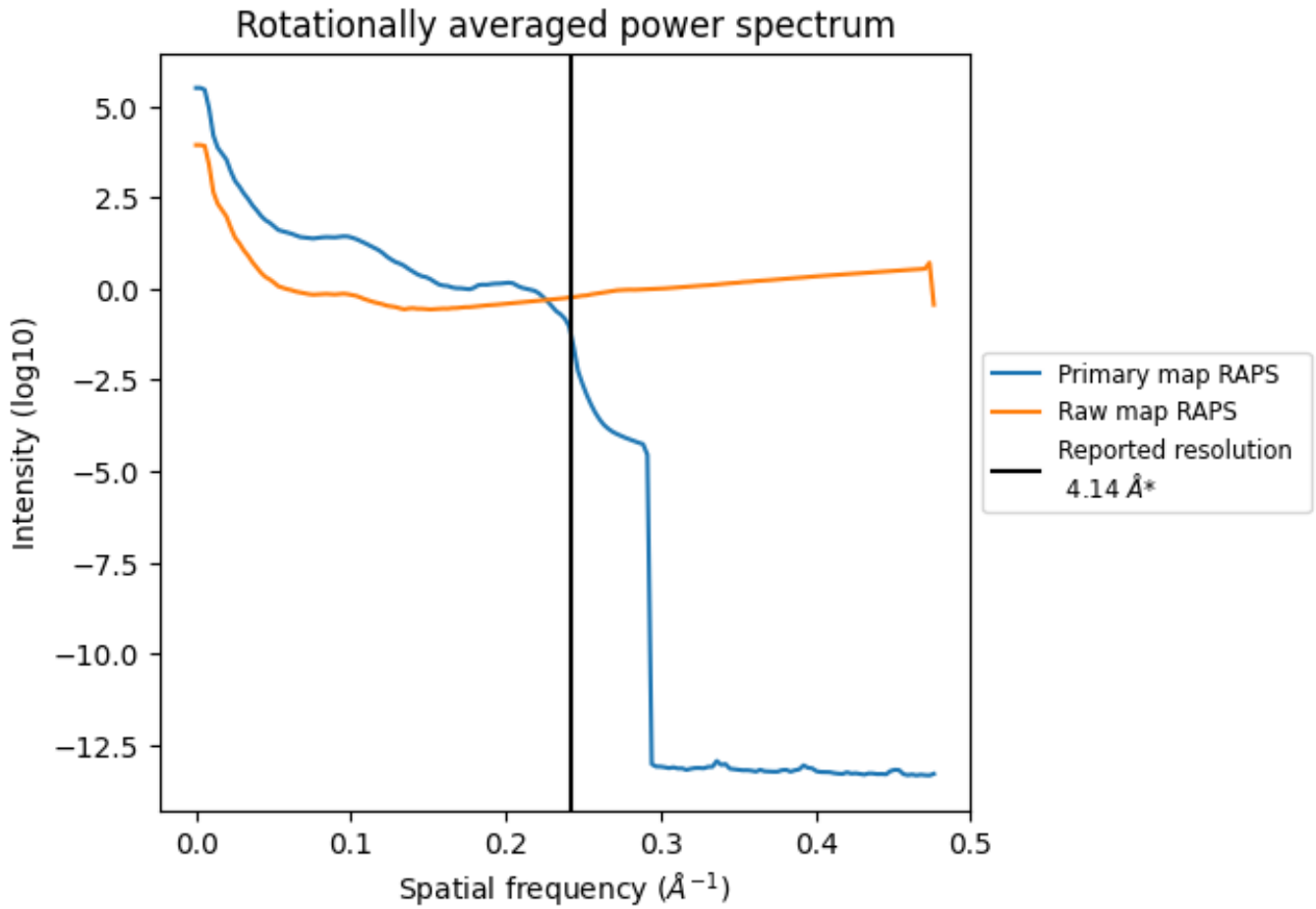
7.2 Volume estimate (i)



The volume at the recommended contour level is 392 nm^3; this corresponds to an approximate mass of 354 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)

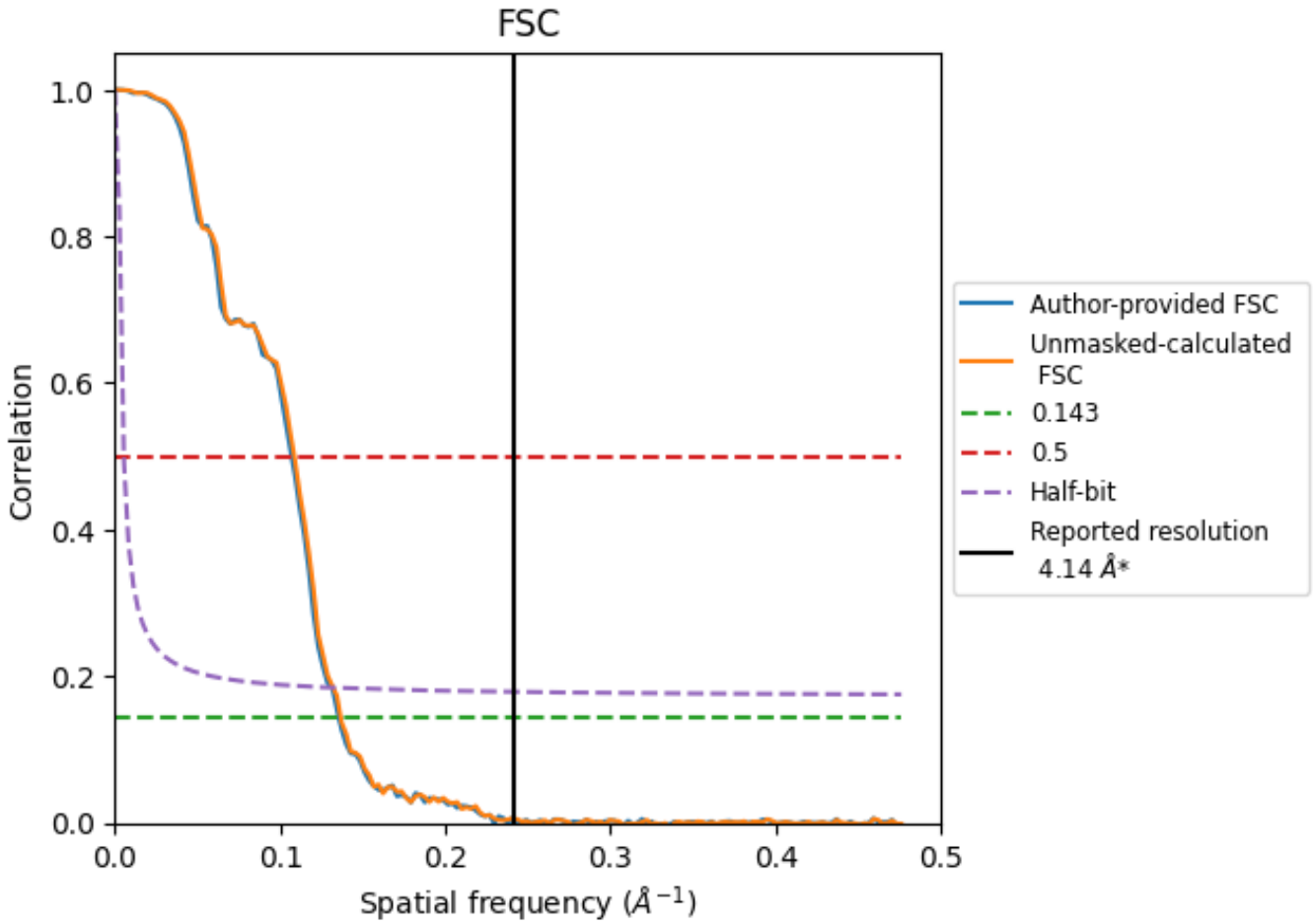


*Reported resolution corresponds to spatial frequency of 0.242 $Å^{-1}$

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 0.242 $Å^{-1}$

8.2 Resolution estimates (i)

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	4.14	-	-
Author-provided FSC curve	7.37	9.34	7.59
Unmasked-calculated*	7.31	9.20	7.56

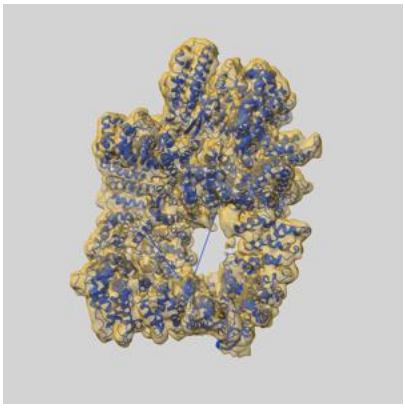
*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps. The value from author-provided FSC intersecting FSC 0.143 CUT-OFF 7.37 differs from the reported value 4.14 by more than 10 %

The value from deposited half-maps intersecting FSC 0.143 CUT-OFF 7.31 differs from the reported value 4.14 by more than 10 %

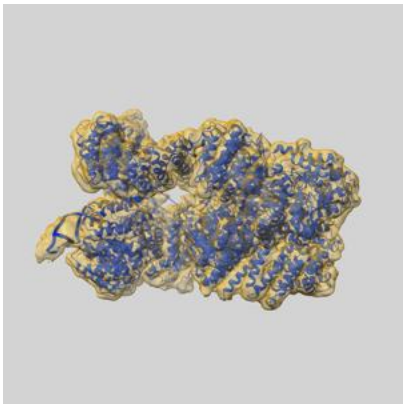
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-11216 and PDB model 6ZH8. Per-residue inclusion information can be found in section 3 on page 4.

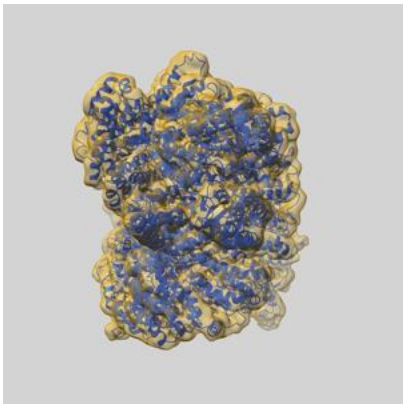
9.1 Map-model overlay



X



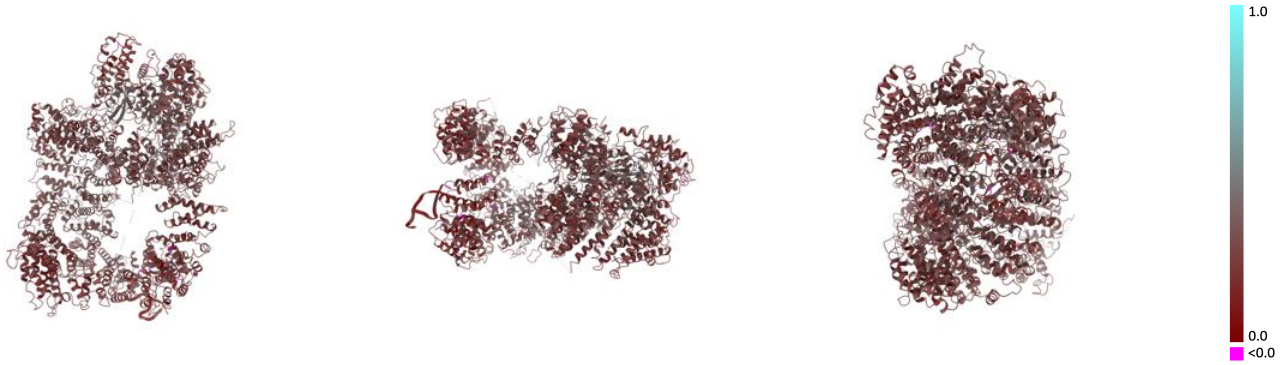
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.125 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.

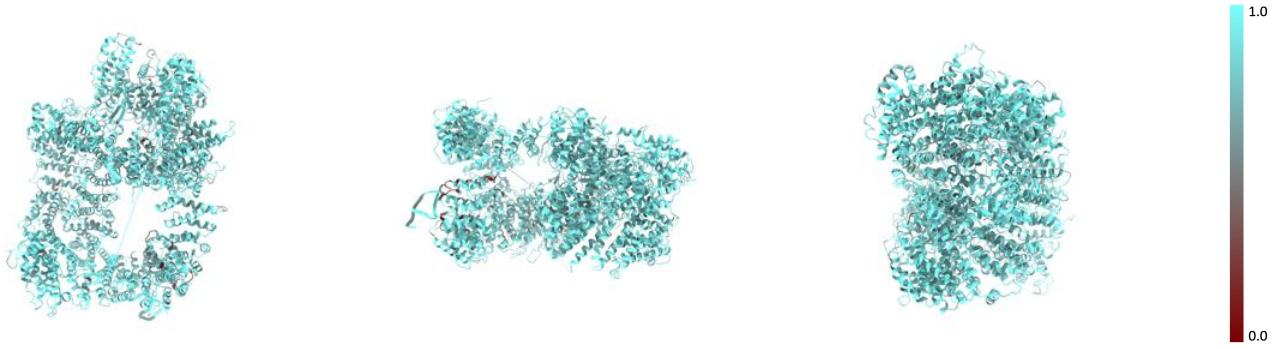
9.2 Q-score mapped to coordinate model (i)



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.

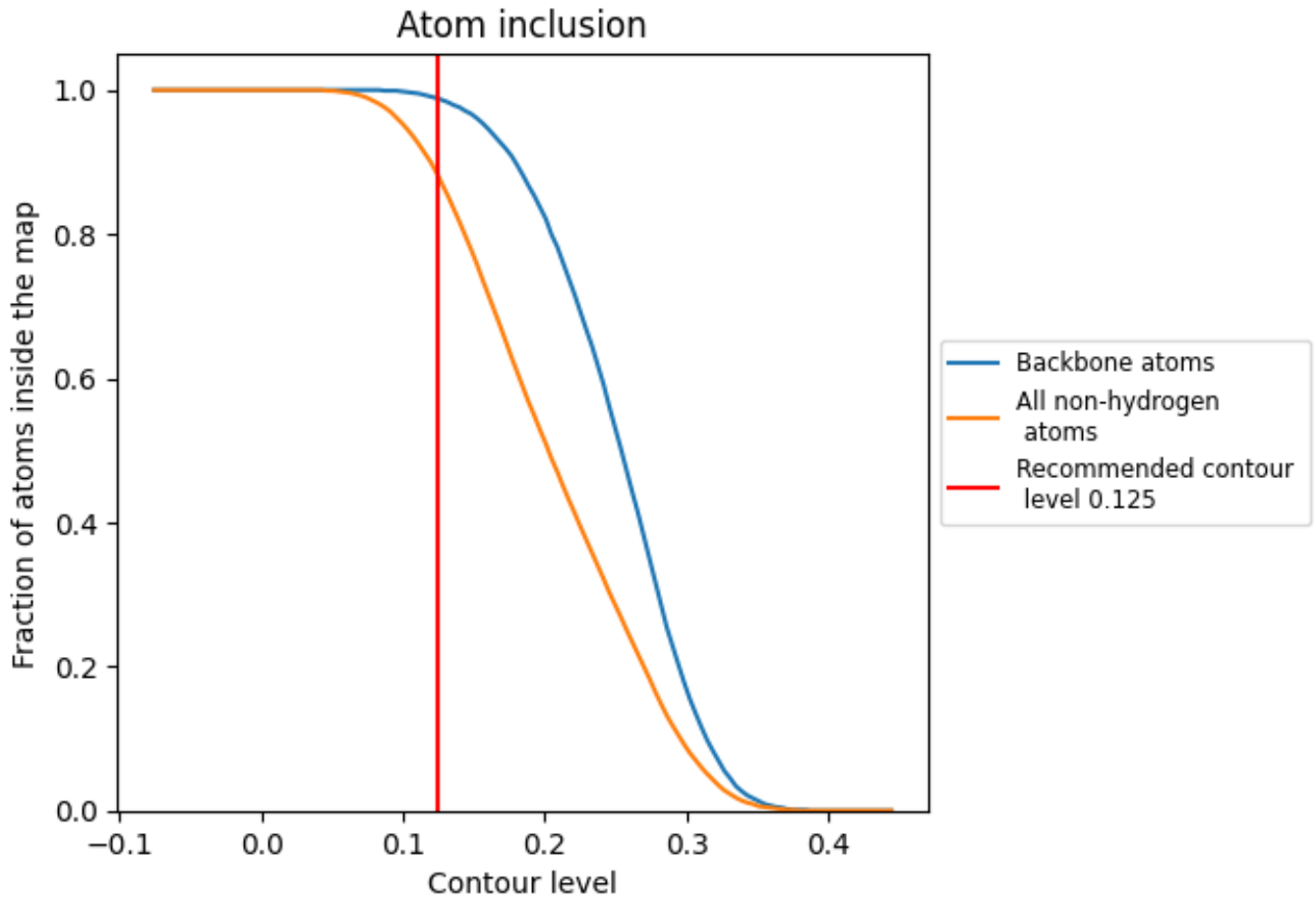
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.125).

9.4 Atom inclusion (i)



At the recommended contour level, 99% of all backbone atoms, 88% of all non-hydrogen atoms, are inside the map.

9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (0.125) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8800	0.2770
A	0.8800	0.2780
B	0.8290	0.2110
C	0.9760	0.2030