wwPDB EM Validation Summary Report (i)

Nov 20, 2022 – 06:18 PM EST

PDB ID	:	7TYR
EMDB ID	:	EMD-26192
Title	:	Cryo-EM structure of the basal state of the Artemis:DNA-PKcs complex (see COMPND 13/14)
Authors	:	Watanabe, G.; Lieber, M.R.; Williams, D.R.
Deposited on	:	2022-02-14
Resolution	:	3.33 Å(reported)
Based on initial model	:	5LUQ

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:

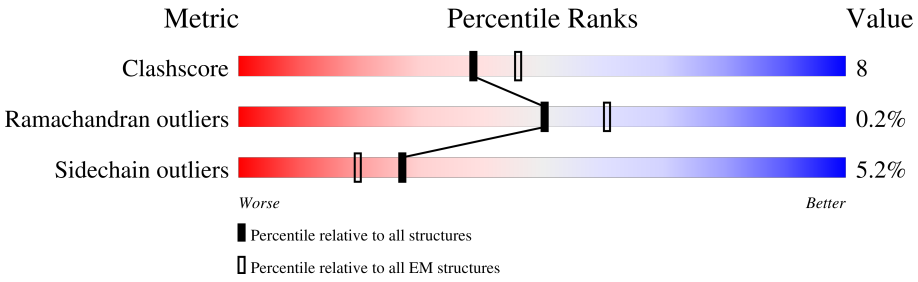
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

1 Overall quality at a glance (i)

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

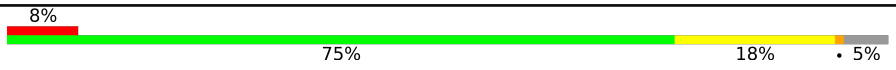
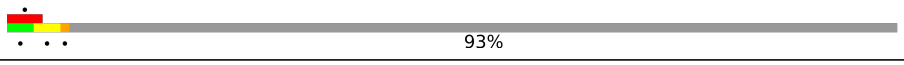
The reported resolution of this entry is 3.33 Å.

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	4128	
2	C	707	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 31610 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	3905	31173	19935	5272	5764	202	1	0

- Molecule 2 is a protein called Protein artemis.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
2	C	52	437	279	76	82	0	0

There are 15 discrepancies between the modelled and reference sequences:

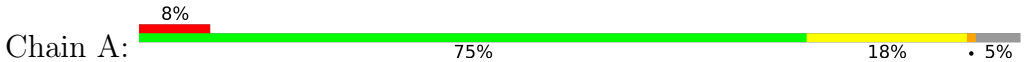
Chain	Residue	Modelled	Actual	Comment	Reference
C	693	GLU	-	expression tag	UNP Q96SD1
C	694	ASN	-	expression tag	UNP Q96SD1
C	695	LEU	-	expression tag	UNP Q96SD1
C	696	TYR	-	expression tag	UNP Q96SD1
C	697	PHE	-	expression tag	UNP Q96SD1
C	698	GLN	-	expression tag	UNP Q96SD1
C	699	GLY	-	expression tag	UNP Q96SD1
C	700	HIS	-	expression tag	UNP Q96SD1
C	701	HIS	-	expression tag	UNP Q96SD1
C	702	HIS	-	expression tag	UNP Q96SD1
C	703	HIS	-	expression tag	UNP Q96SD1
C	704	HIS	-	expression tag	UNP Q96SD1
C	705	HIS	-	expression tag	UNP Q96SD1
C	706	HIS	-	expression tag	UNP Q96SD1
C	707	HIS	-	expression tag	UNP Q96SD1

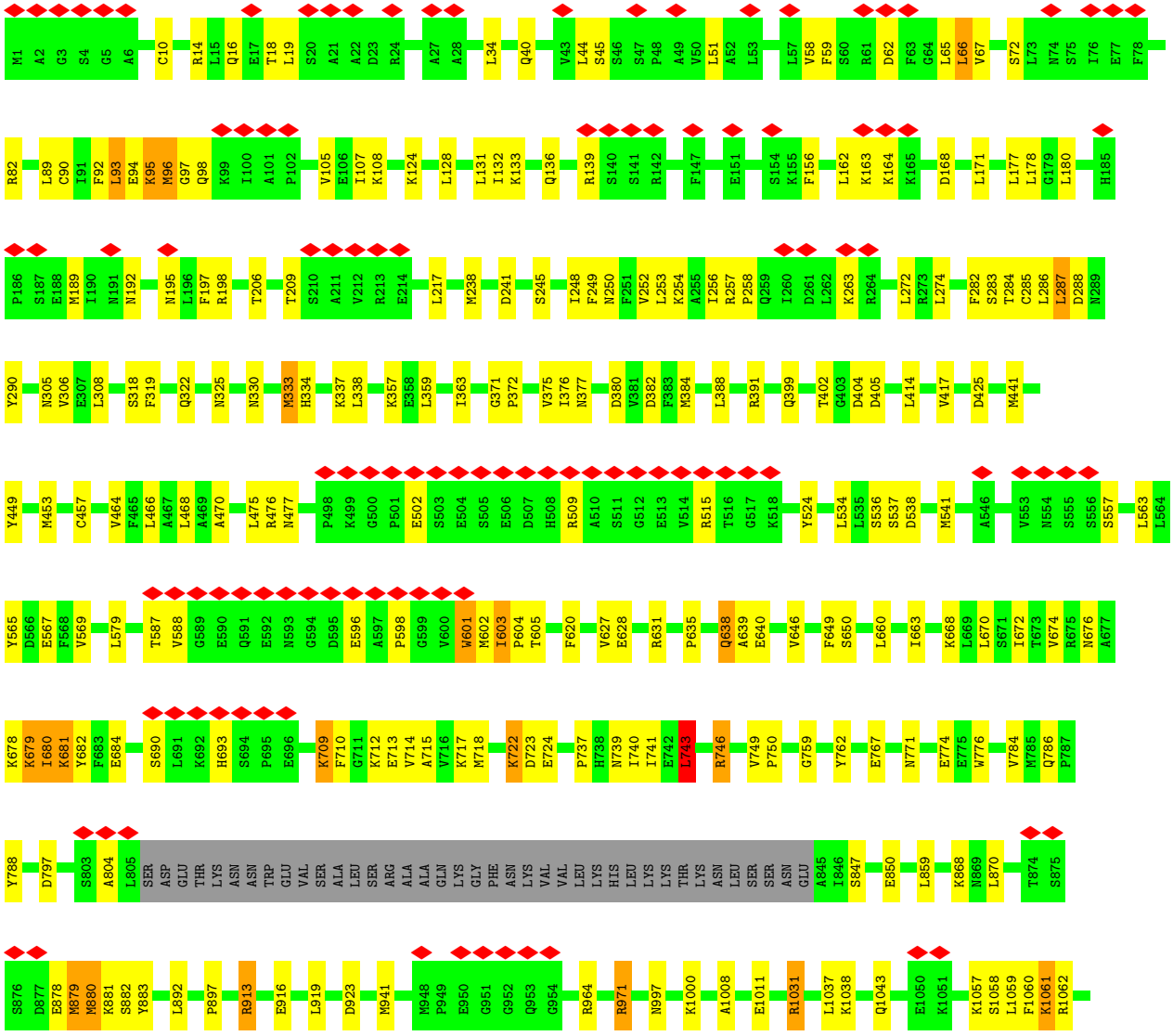
3 Residue-property plots (i)

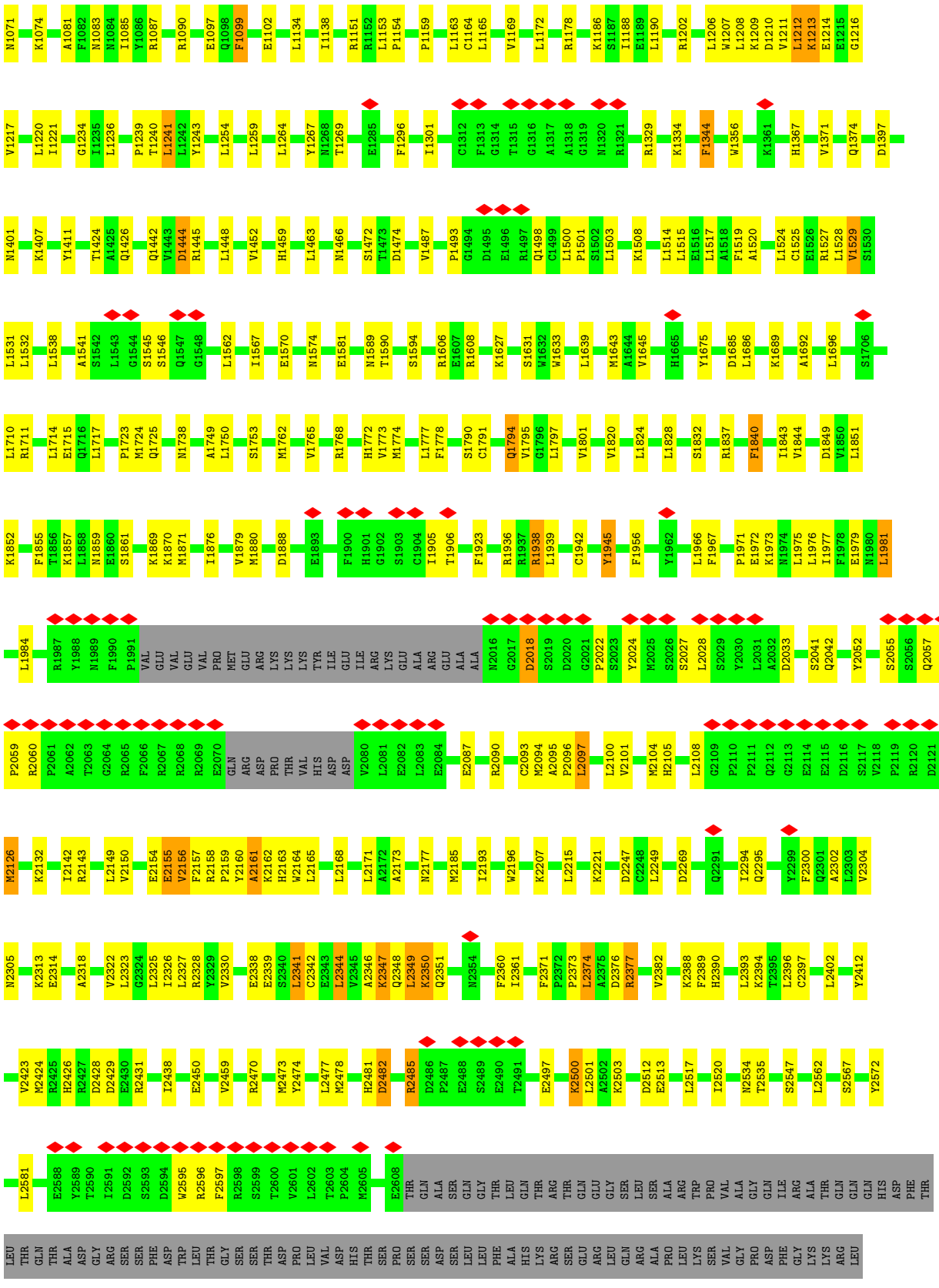
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion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

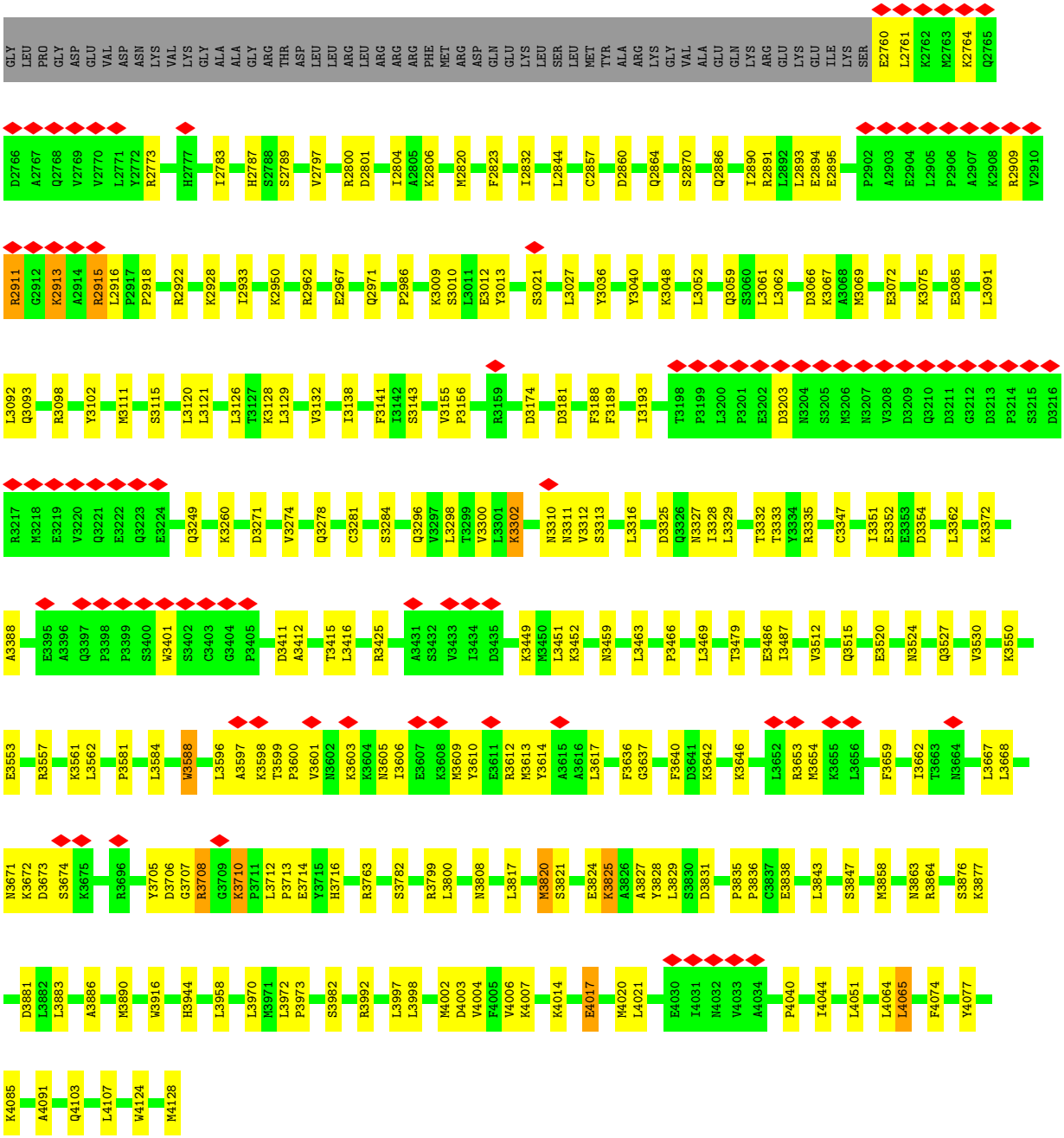
- Molecule 1: DNA-dependent protein kinase catalytic subunit

Chain A:



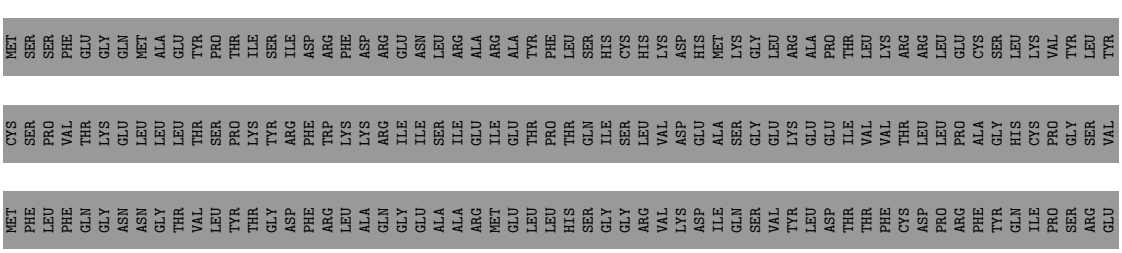


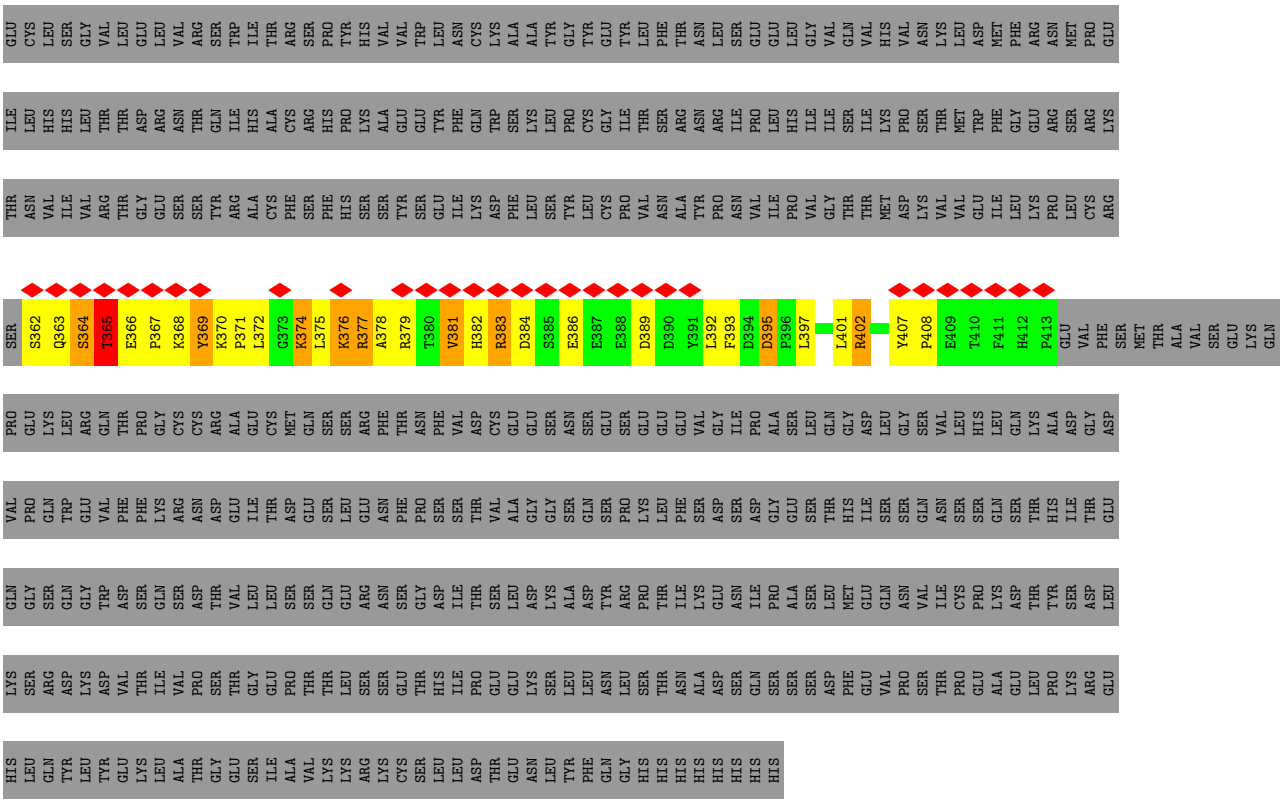

N1071 K1074 A1081 F1082 N1083 N1084 I1085 Y1086 R1087 R1090 E1097 Q1098 F1099 E1102 L1134 I1138 R1151 R1152 L1153 P1154 P1159 L1163 C1164 L1165 V1169 L1172 R1178 K1186 S1187 I1188 E1189 L1190 R1202 L1206 W1207 L1208 K1209 D1210 V1211 L1212 K1213 E1214 E1215 G1216
V1217 L1220 I1221 G1234 I1235 L1236 P1239 T1240 L1241 L1242 Y1243 L1254 L1259 L1264 Y1267 N1268 T1269 E1285 F1296 I1301 C1312 F1313 G1314 T1315 G1316 A1317 A1318 G1319 N1320 R1321 R1329 K1334 F1344 W1356 K1361 H1367 V1371 Q1374 D1397
N1401 K1407 Y1411 T1424 A1425 Q1426 Q1442 V1443 D1444 R1445 L1448 V1452 H1459 L1463 N1466 S1472 T1473 D1474 V1487 P1493 G1494 D1495 E1496 R1497 Q1498 C1499 L1500 P1501 S1502 L1503 K1508 L1514 L1515 E1516 L1517 A1518 F1519 A1520 L1524 C1525 E1526 R1527 L1528 V1529 S1530
L1531 L1532 L1538 A1541 S1542 L1543 G1544 S1545 S1546 Q1547 G1548 L1562 I1567 E1570 N1574 E1581 N1589 T1590 S1594 R1606 E1607 R1608 K1627 S1631 W1632 W1633 L1639 M1643 A1644 V1645 H1665 Y1675 D1685 L1686 K1689 A1692 L1696 S1706
L1710 R1711 L1714 E1715 Q1716 L1717 P1723 M1724 Q1725 N1738 A1749 L1750 S1753 M1762 V1765 R1768 H1772 V1773 M1774 L1777 F1778 S1790 C1791 Q1794 V1795 G1796 L1797 V1801 V1820 L1824 L1828 S1832 R1837 F1840 I1843 V1844 D1849 V1850 L1851
K1852 F1855 T1856 K1857 L1858 N1859 E1860 S1861 K1869 K1870 M1871 I1876 V1879 M1880 D1888 E1893 F1900 H1901 G1902 S1903 C1904 I1905 T1906 F1923 R1936 R1937 R1938 L1939 C1942 Y1945 F1956 Y1962 L1966 F1967 P1971 E1972 K1973 N1974 L1975 L1976 I1977 F1978 E1979 N1980 L1981
L1984 R1987 Y1988 N1989 F1990 P1991 VAL GLU VAL GLU VAL PRO MET GLU ARG LYS LYS LYS TYR ILE GLU ILE ARG LYS GLU ALA ARG GLU ALA ALA N2016 G2017 D2018 S2019 D2020 G2021 P2022 S2023 Y2024 M2025 S2026 S2027 L2028 S2029 Y2030 L2031 A2032 D2033 S2041 Q2042 Y2052 S2055 S2056 Q2057 D2058
P2059 R2060 P2061 A2062 T2063 G2064 R2065 F2066 R2067 R2068 R2069 E2070 GLN ARG ASP PRO THR VAL HIS ASP V2080 L2081 E2082 L2083 E2084 E2087 R2090 C2093 M2094 A2095 P2096 L2097 L2100 V2101 M2104 H2105 L2108 G2109 P2110 P2111 Q2112 Q2113 E2114 E2115 D2116 S2117 V2118 P2119 R2120 D2121 L2122 P2123
M2126 K2132 I2142 R2143 L2149 V2150 E2154 E2155 V2156 F2157 R2158 P2159 Y2160 A2161 K2162 H2163 W2164 L2165 L2168 L2171 A2172 A2173 N2177 M2185 I2193 W2196 K2207 L2215 K2221 D2247 C2248 L2249 D2269 Q2291 I2294 Q2295 Y2299 F2300 Q2301 A2302 L2303 V2304
N2305 K2313 E2314 A2318 V2322 L2323 G2324 L2325 I2326 L2327 R2328 Y2329 V2330 E2338 E2339 S2340 L2341 C2342 E2343 L2344 V2345 A2346 K2347 Q2348 L2349 K2350 Q2351 N2354 F2360 I2361 F2371 P2372 P2373 L2374 A2375 D2376 R2377 V2382 K2388 F2389 H2390 L2393 K2394 T2395 L2396 C2397 L2402 Y2412
V2423 M2424 R2425 H2426 R2427 D2428 D2429 E2430 R2431 I2438 E2450 V2459 R2470 M2473 Y2474 L2477 M2478 H2481 D2482 R2485 D2486 P2487 E2488 S2489 E2490 T2491 E2497 K2500 L2501 A2502 K2503 D2512 E2513 L2517 I2520 W2534 T2535 S2547 L2562 S2567 Y2572
L2581 E2588 Y2589 T2590 I2591 D2592 S2593 D2594 W2595 R2596 F2597 R2598 S2599 T2600 V2601 L2602 T2603 P2604 M2605 E2608 THR GLN ALA SER GLN GLY THR LEU GLN THR ARG THR GLN GLU GLY SER LEU GLN ARG ALA ARG PRO LEU LYS SER VAL GLY PRO ASP PHE GLY LYS LYS ARG LEU
LEU THR GLN THR ALA ASP GLY ARG SER SER PHE ASP TRP LEU THR GLY SER SER THR ASP PRO LEU VAL ASP HIS THR SER PRO SER SER ASP SER SER LEU LEU PHE ALA HIS LYS ARG SER GLU ARG LEU GLN ARG ALA PRO LEU LYS SER VAL GLY PRO ASP PHE GLY LYS LYS ARG LEU



• Molecule 2: Protein artemis

Chain C: 93%




SER S362 Q363 S364 T365 E366 P367 K368 Y369 K370 P371 L372 G373 K374 L375 K376 R377 A378 R379 T380 V381 H382 R383 D384 S385 E386 E387 E388 D389 D390 Y391 L392 F393 D394 D395 P396 L397 L401 R402 Y407 P408 E409 T410 F411 H412 P413 GLU

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	103485	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 ($e^-/Å^2$)	60.0	Depositor
Minimum defocus (nm)	750	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	46296	Depositor
Image detector	GATAN K3 BIOQUANTUM (6k x 4k)	Depositor
Maximum map value	2.449	Depositor
Minimum map value	-1.114	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.049	Depositor
Recommended contour level	0.3	Depositor
Map size (Å)	552.96, 552.96, 552.96	wwPDB
Map dimensions	512, 512, 512	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.08, 1.08, 1.08	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	0/31831	0.51	0/43042
2	C	0.56	0/452	0.67	0/614
All	All	0.29	0/32283	0.51	0/43656

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	31173	0	31434	479	0
2	C	437	0	424	35	0
All	All	31610	0	31858	501	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 501 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)
2:C:381:VAL:HG13	2:C:382:HIS:ND1	1.35	1.35
2:C:381:VAL:CG1	2:C:382:HIS:ND1	2.17	1.07

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:C:365:THR:HG23	2:C:367:PRO:HD2	1.38	1.03
1:A:2911:ARG:CZ	1:A:2913:LYS:HE3	1.93	0.98
2:C:369:TYR:HB2	2:C:371:PRO:HD3	1.51	0.92

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	3896/4128 (94%)	3699 (95%)	191 (5%)	6 (0%)	47 78
2	C	50/707 (7%)	38 (76%)	11 (22%)	1 (2%)	7 34
All	All	3946/4835 (82%)	3737 (95%)	202 (5%)	7 (0%)	50 78

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	882	SER
1	A	2161	ALA
1	A	2162	LYS
1	A	3716	HIS
1	A	682	TYR

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	3474/3671 (95%)	3312 (95%)	162 (5%)	26 60
2	C	50/647 (8%)	29 (58%)	21 (42%)	0 0
All	All	3524/4318 (82%)	3341 (95%)	183 (5%)	27 56

5 of 183 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2482	ASP
1	A	3588	TRP
1	A	2503	LYS
1	A	2962	ARG
1	A	3782	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	2481	HIS
1	A	2977	ASN
1	A	1238	GLN
1	A	1611	GLN
1	A	1859	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)

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)

There are no chain breaks in this entry.

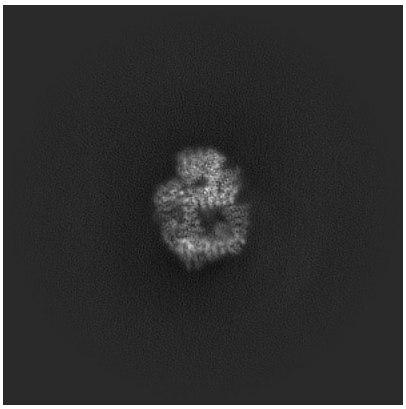
6 Map visualisation

This section contains visualisations of the EMDB entry EMD-26192. These allow visual inspection of the internal detail of the map and identification of artifacts.

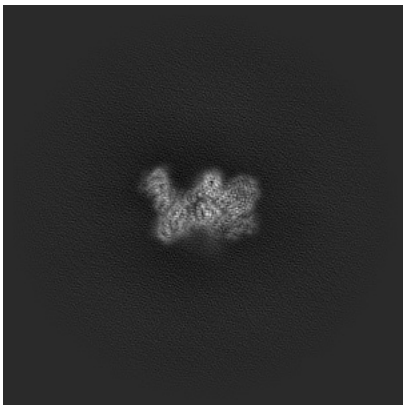
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections

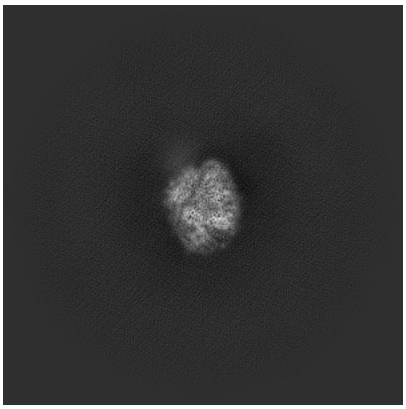
6.1.1 Primary map



X

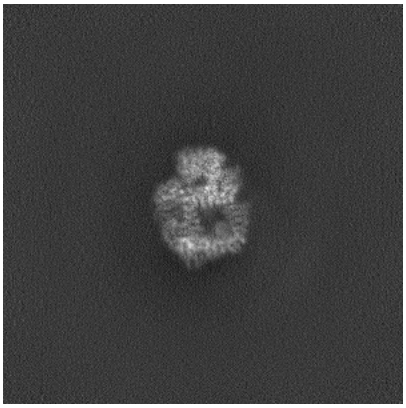


Y

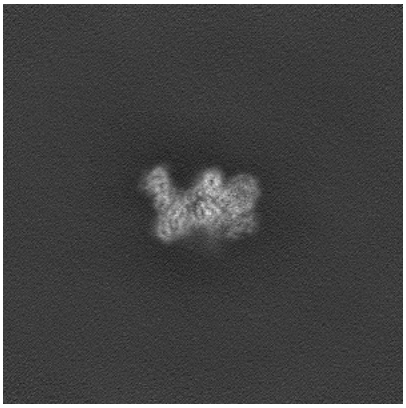


Z

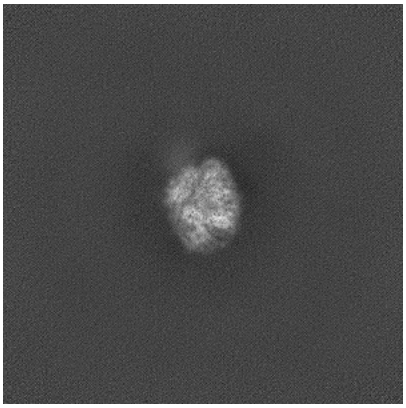
6.1.2 Raw map



X



Y

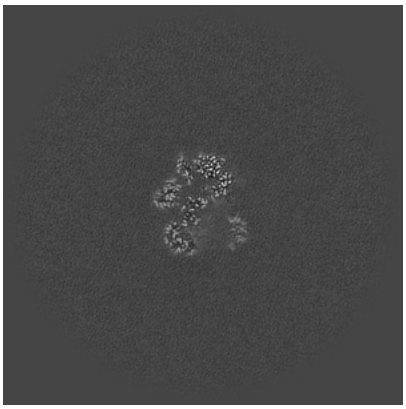


Z

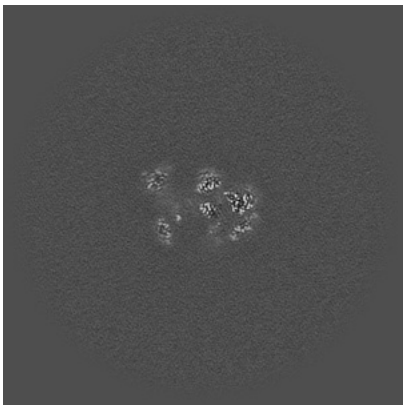
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

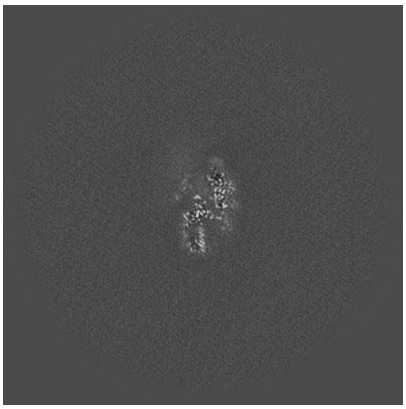
6.2.1 Primary map



X Index: 256

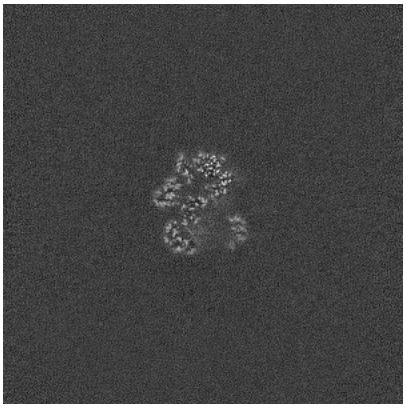


Y Index: 256

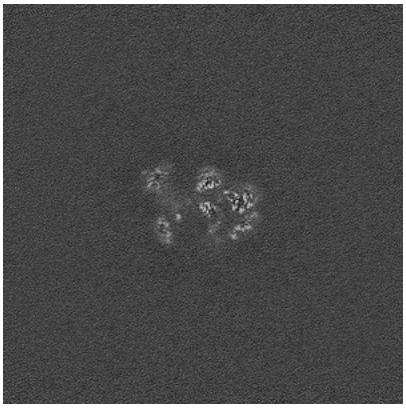


Z Index: 256

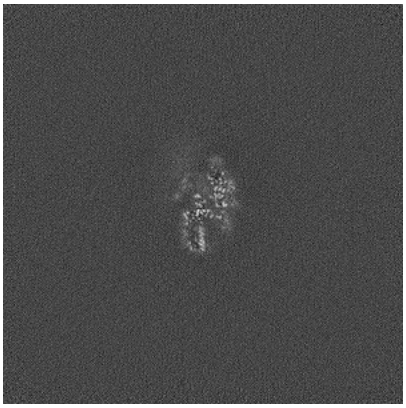
6.2.2 Raw map



X Index: 256



Y Index: 256

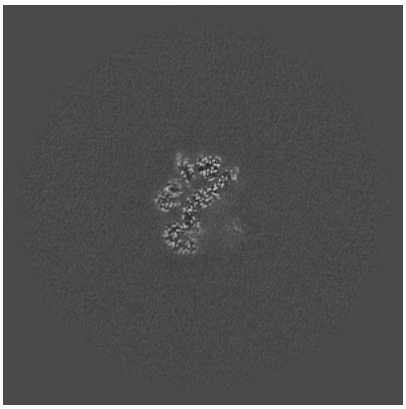


Z Index: 256

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

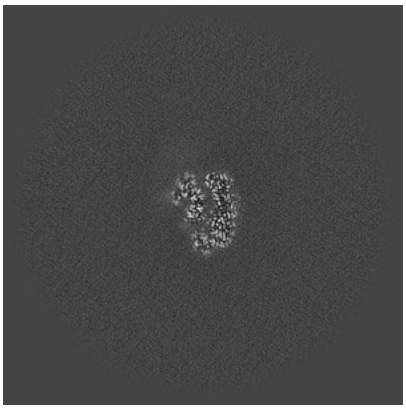
6.3 Largest variance slices (i)

6.3.1 Primary map



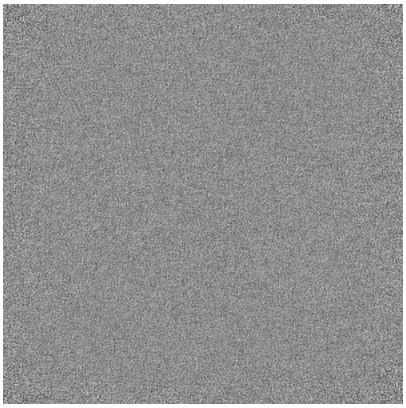
X Index: 252

Y Index: 238

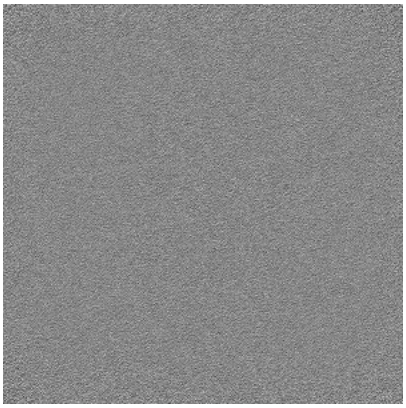


Z Index: 265

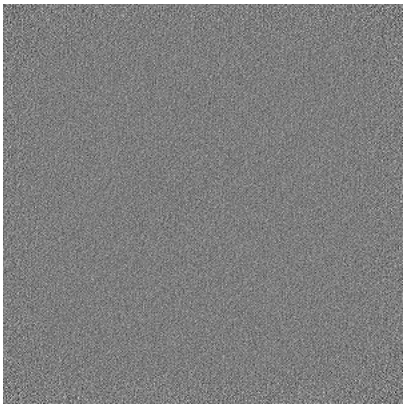
6.3.2 Raw map



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 surface views (i)

6.4.1 Primary map

X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.3. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.4.2 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.

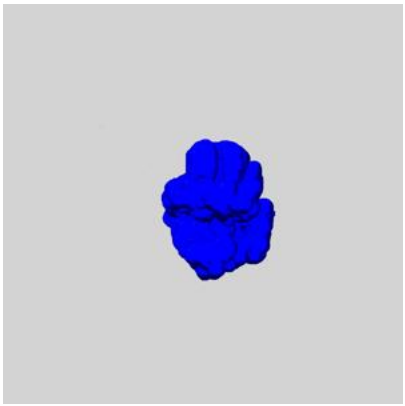
6.5 Mask visualisation (i)

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

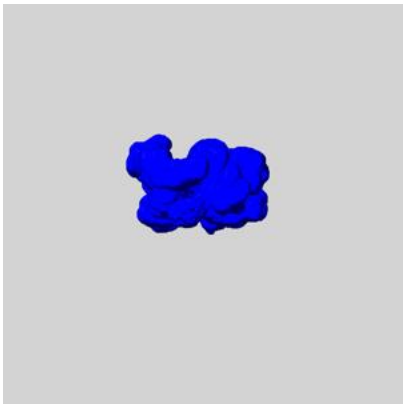
A mask typically either:

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

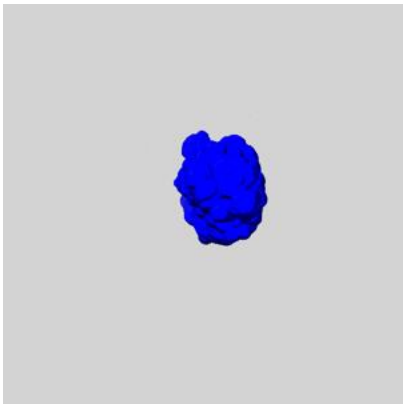
6.5.1 emd_26192_msk_1.map (i)



X



Y

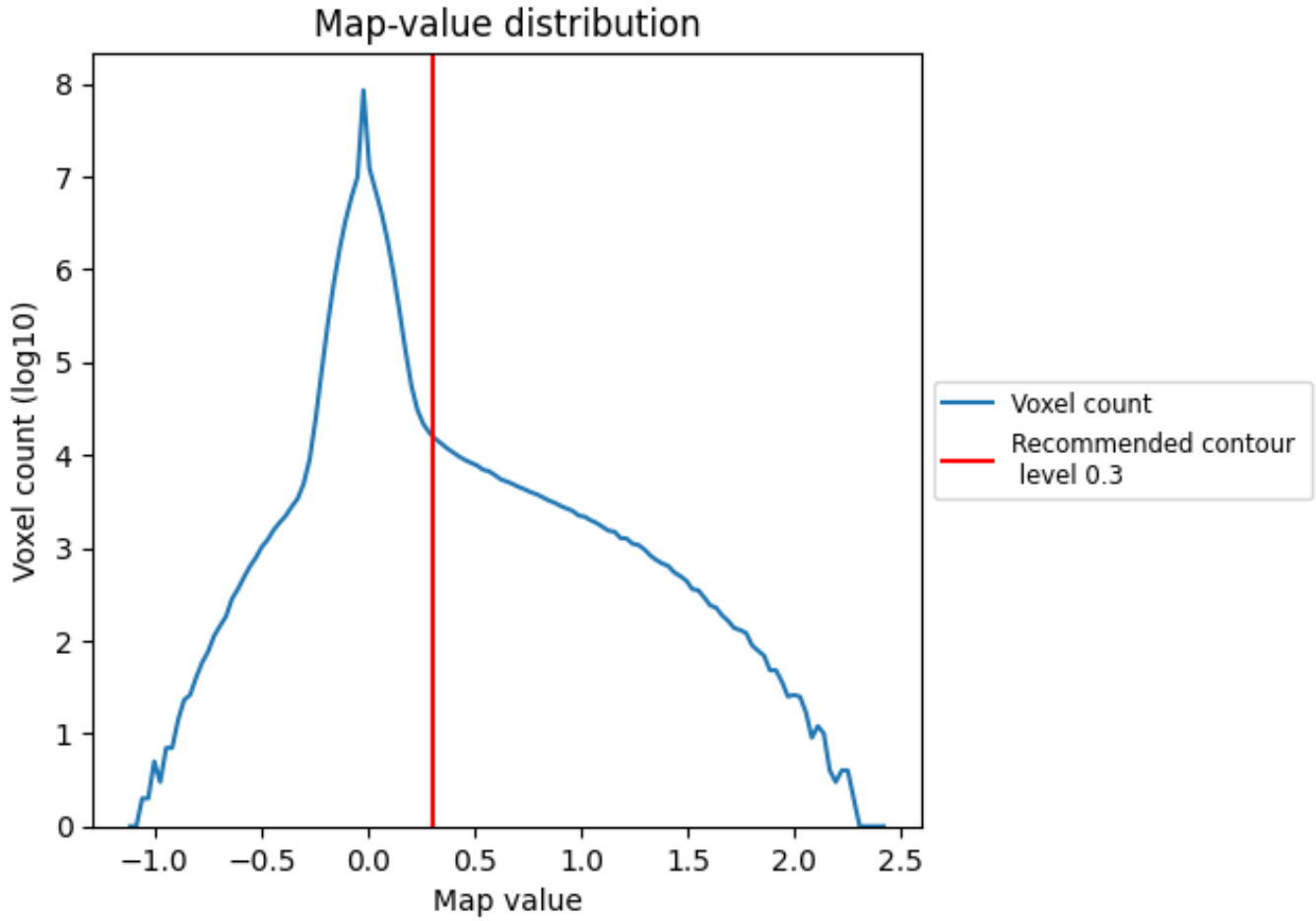


Z

7 Map analysis

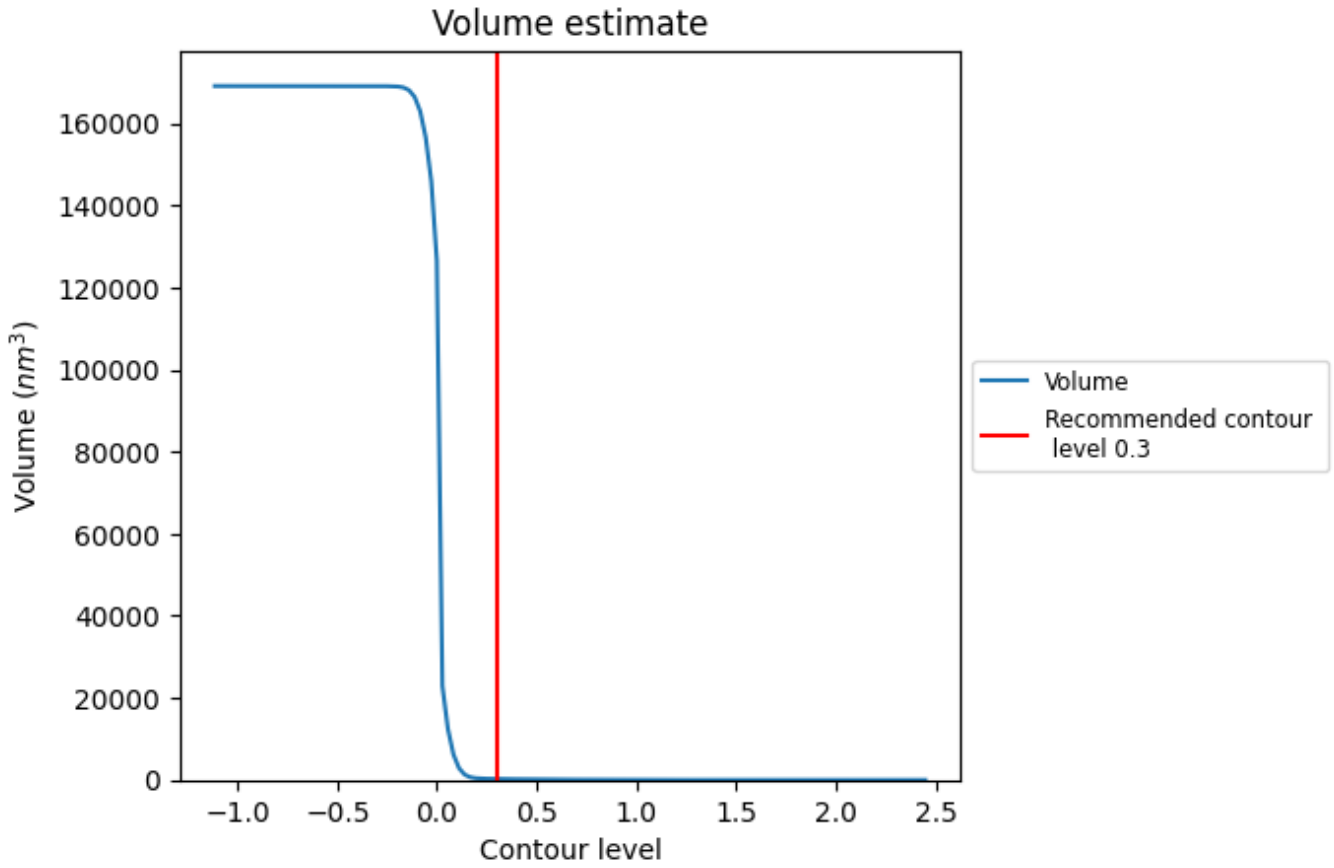
This section contains the results of statistical analysis of the map.

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.

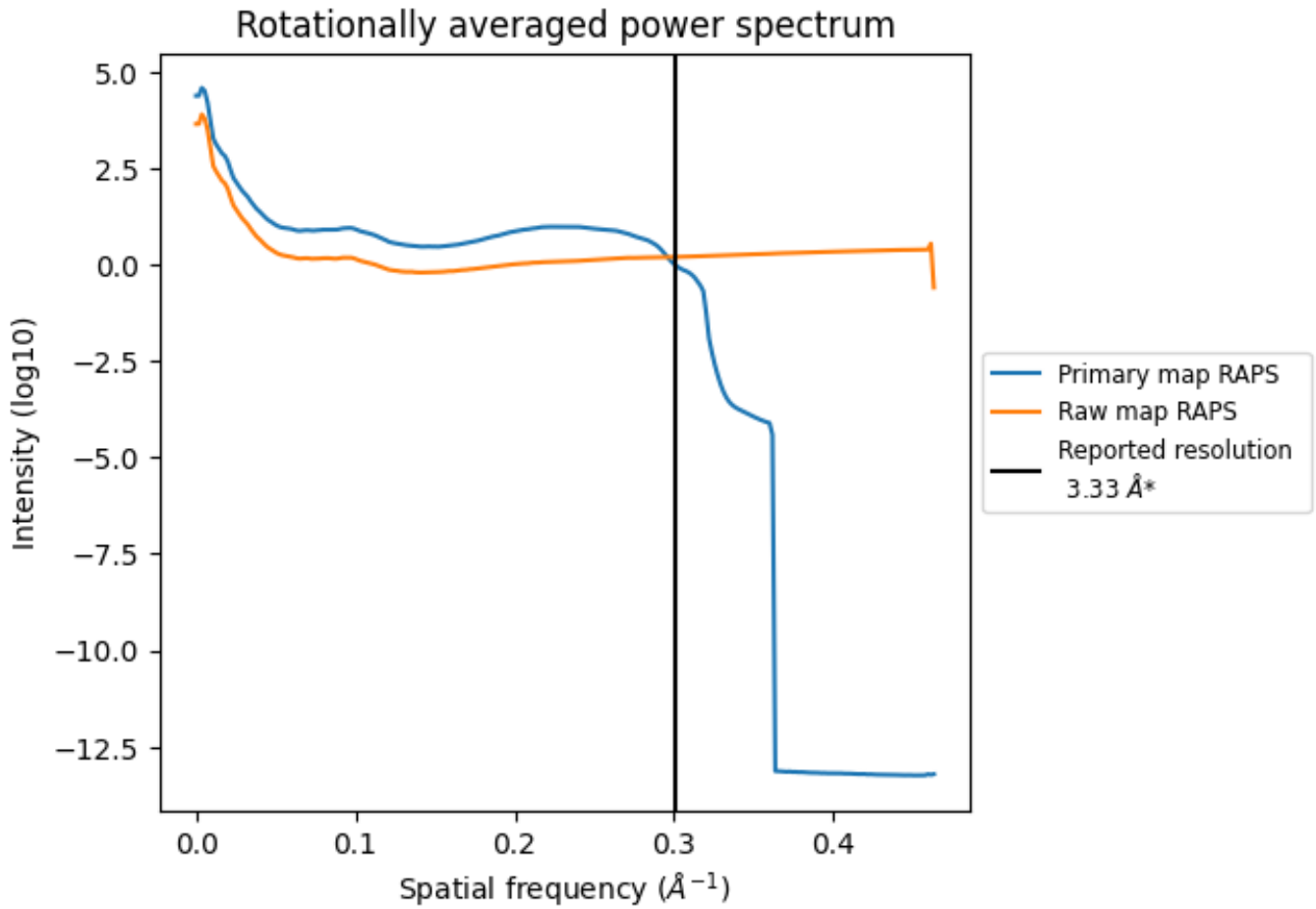
7.2 Volume estimate



The volume at the recommended contour level is 240 nm^3; this corresponds to an approximate mass of 217 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)

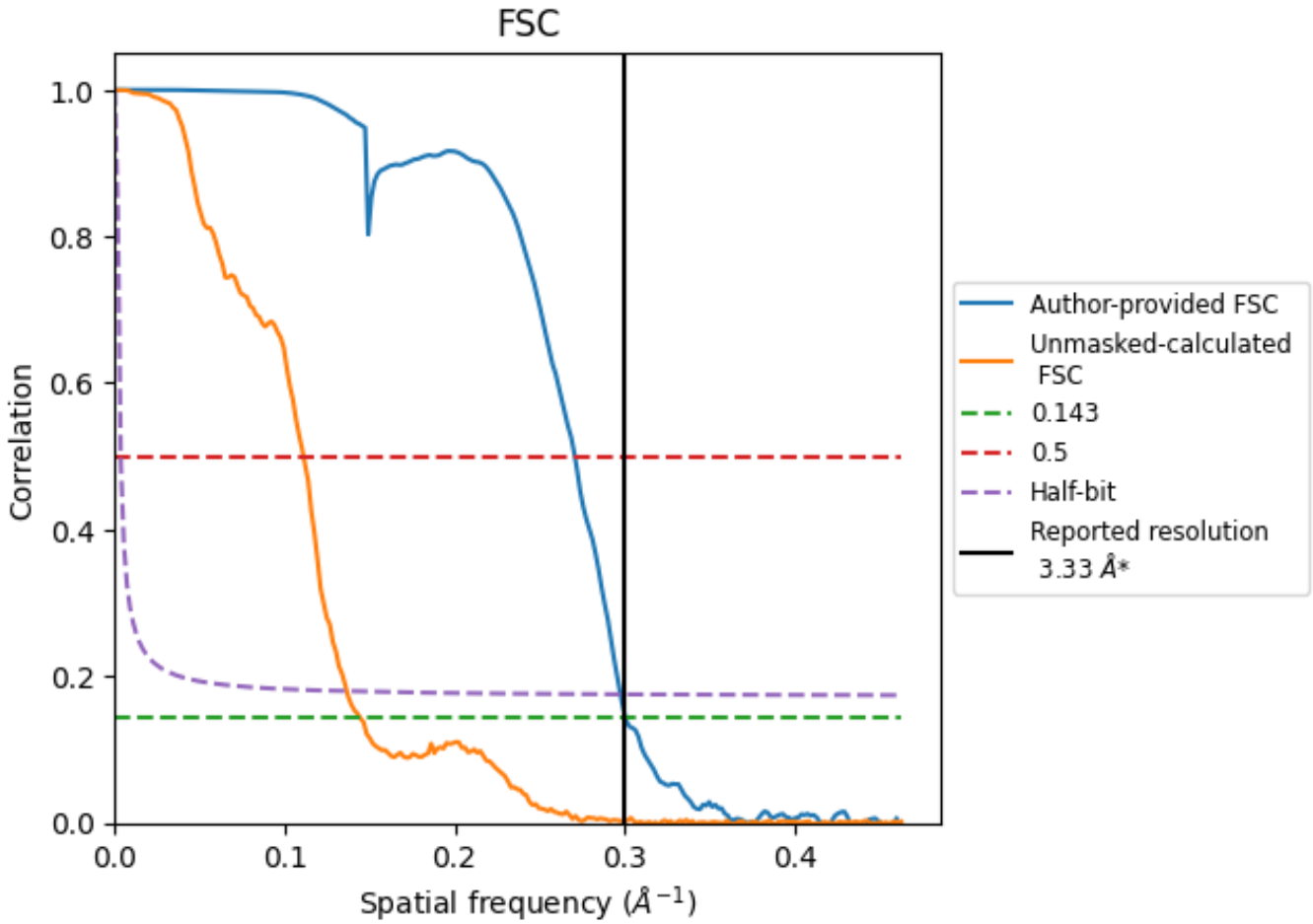


*Reported resolution corresponds to spatial frequency of 0.300 $Å^{-1}$

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 0.300 Å^{-1}

8.2 Resolution estimates (i)

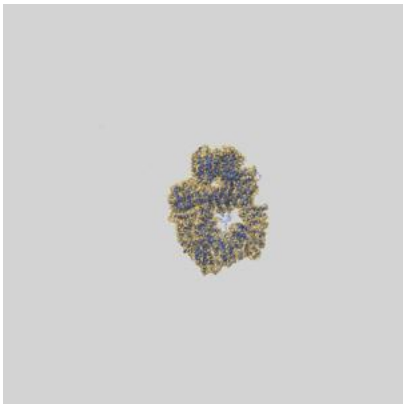
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.33	-	-
Author-provided FSC curve	3.33	3.69	3.36
Unmasked-calculated*	6.93	9.00	7.32

*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 6.93 differs from the reported value 3.33 by more than 10 %

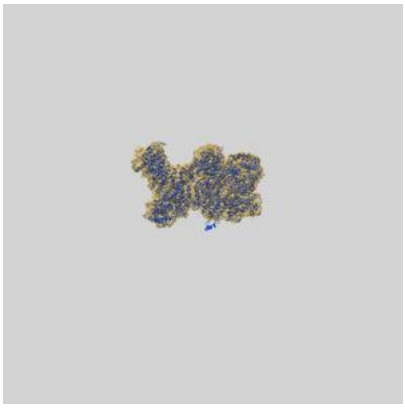
9 Map-model fit

This section contains information regarding the fit between EMDB map EMD-26192 and PDB model 7TYR. Per-residue inclusion information can be found in section 3 on page 4.

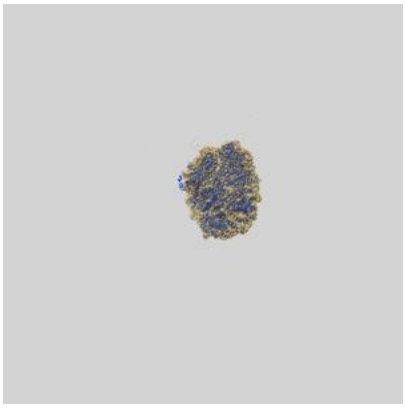
9.1 Map-model overlay



X



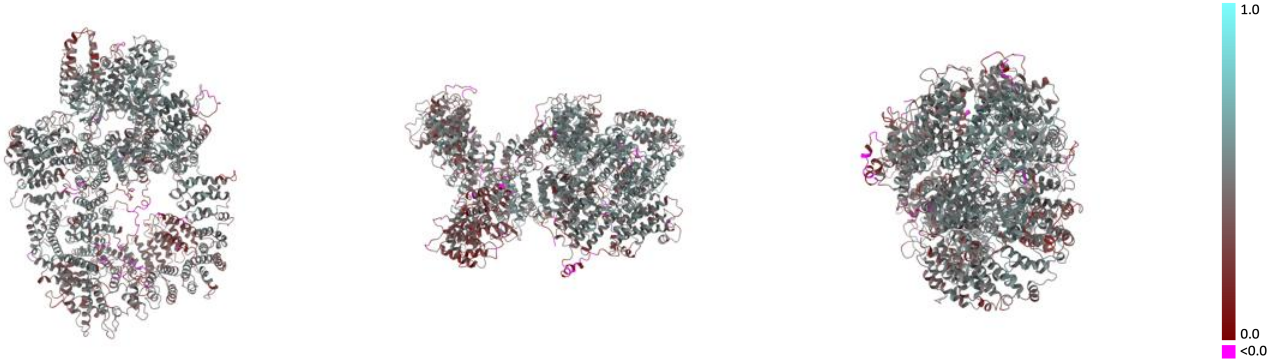
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.3 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.

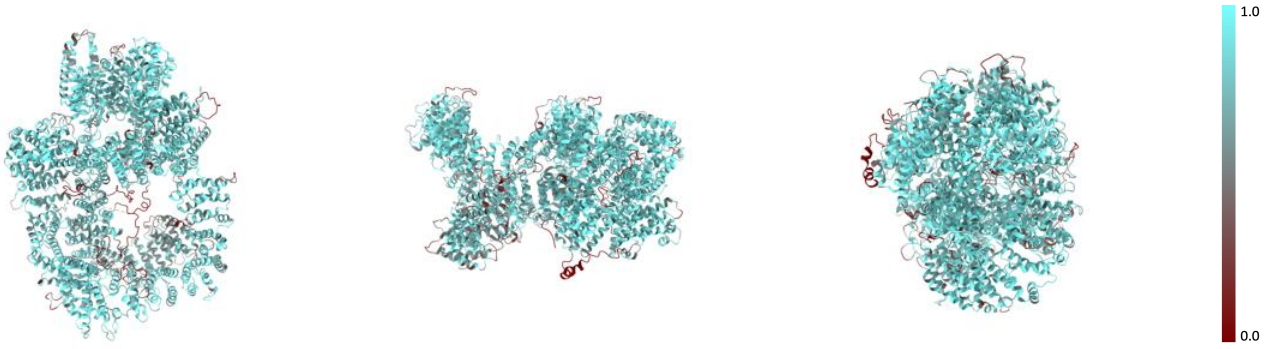
9.2 Q-score mapped to coordinate model (i)



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.

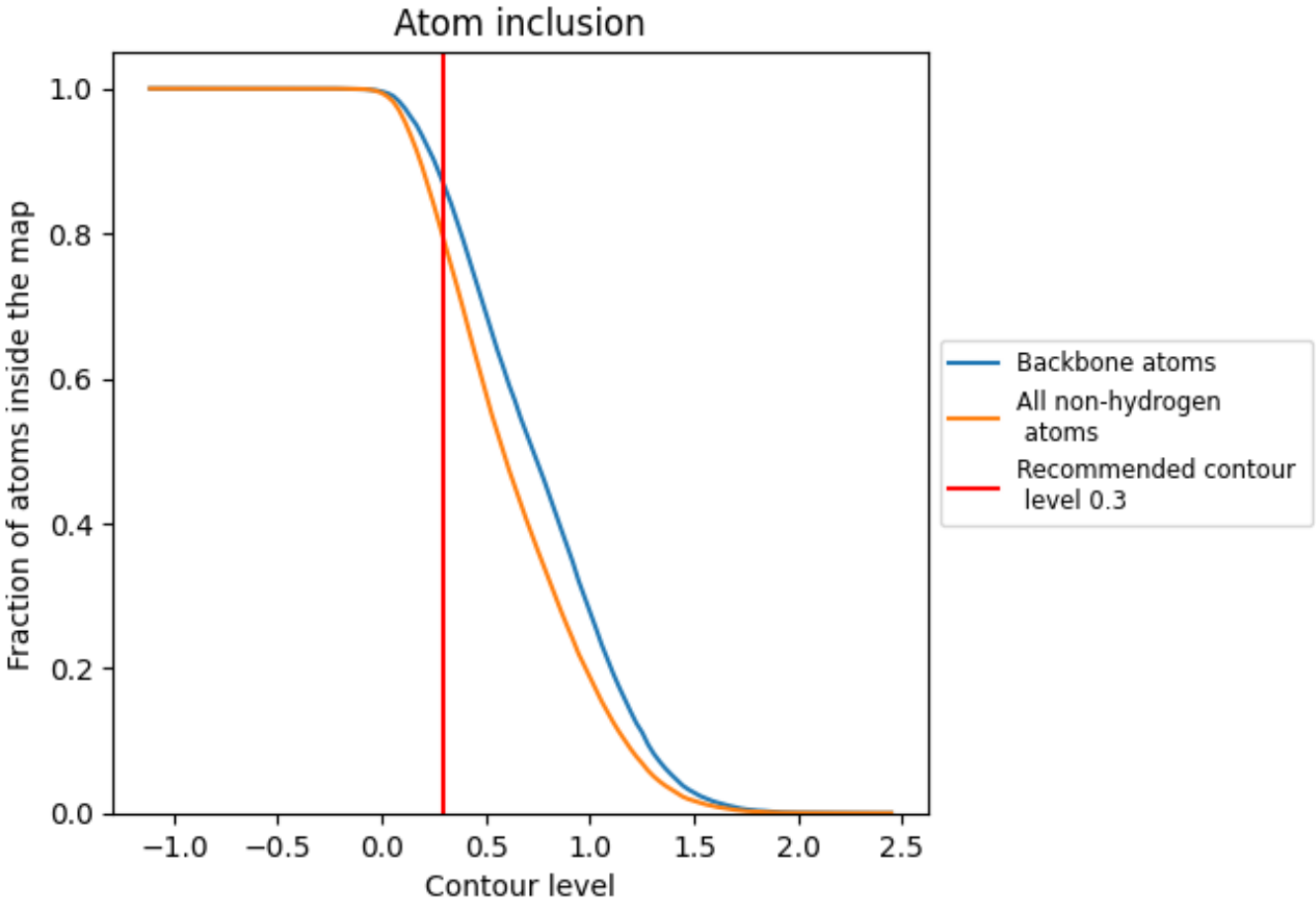
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.3).

9.4 Atom inclusion (i)



At the recommended contour level, 87% of all backbone atoms, 79% 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 (0.3) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7893	0.4330
A	0.7958	0.4360
C	0.3216	0.2170