wwPDB EM Validation Summary Report (i)

Nov 13, 2022 – 09:00 PM EST

PDB ID	:	6X9O
EMDB ID	:	EMD-22106
Title	:	High resolution cryoEM structure of huntingtin in complex with HAP40
Authors	:	Harding, R.J.; Deme, J.C.; Lea, S.M.; Arrowsmith, C.H.; Structural Genomics Consortium (SGC)
Deposited on	:	2020-06-03
Resolution	:	2.60 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:

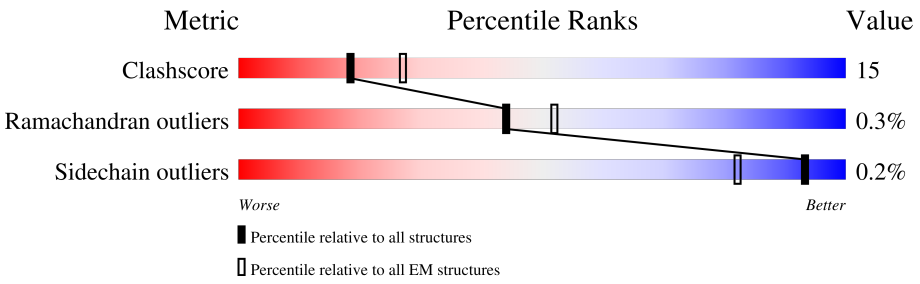
EMDB validation analysis	:	0.0.1.dev43
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 2.60 Å.

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



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

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

Mol	Chain	Length	Quality of chain
1	A	3156	8% 56% 21% 23%
2	B	389	• 51% 12% 37%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 20899 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 Huntingtin.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	2425	19063	12213	3276	3451	123	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	39	GLN	-	insertion	UNP P42858
A	40	GLN	-	insertion	UNP P42858
A	2311	HIS	TYR	variant	UNP P42858
A	2788	ILE	VAL	variant	UNP P42858
A	3145	GLY	-	expression tag	UNP P42858
A	3146	GLY	-	expression tag	UNP P42858
A	3147	SER	-	expression tag	UNP P42858
A	3148	GLY	-	expression tag	UNP P42858
A	3149	ASP	-	expression tag	UNP P42858
A	3150	TYR	-	expression tag	UNP P42858
A	3151	LYS	-	expression tag	UNP P42858
A	3152	ASP	-	expression tag	UNP P42858
A	3153	ASP	-	expression tag	UNP P42858
A	3154	ASP	-	expression tag	UNP P42858
A	3155	ASP	-	expression tag	UNP P42858
A	3156	LYS	-	expression tag	UNP P42858

- Molecule 2 is a protein called 40-kDa huntingtin-associated protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	B	244	1836	1166	327	336	7	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-17	MET	-	initiating methionine	UNP P23610

Continued on next page...

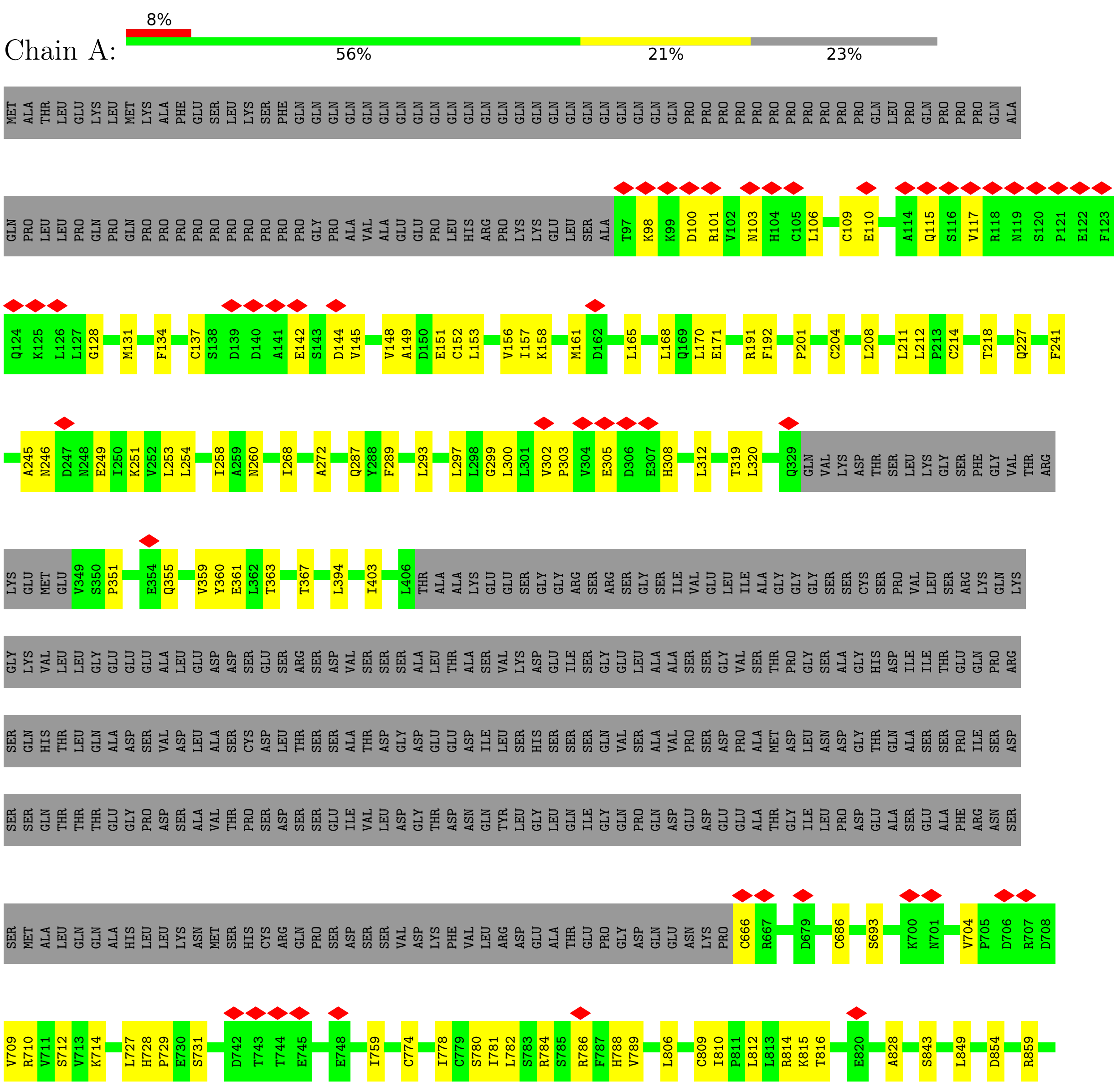
Continued from previous page...

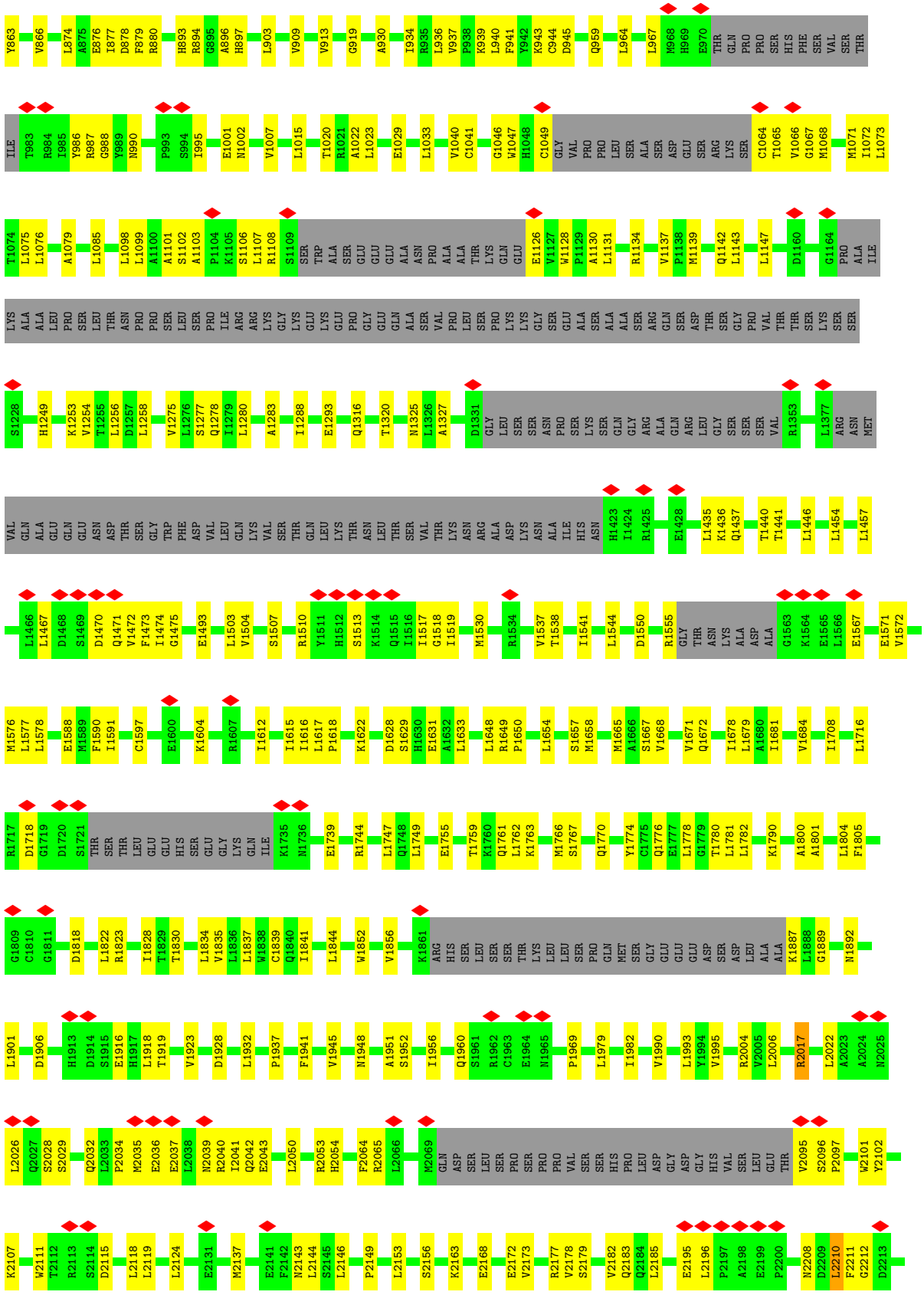
Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	HIS	-	expression tag	UNP P23610
B	-15	HIS	-	expression tag	UNP P23610
B	-14	HIS	-	expression tag	UNP P23610
B	-13	HIS	-	expression tag	UNP P23610
B	-12	HIS	-	expression tag	UNP P23610
B	-11	HIS	-	expression tag	UNP P23610
B	-10	SER	-	expression tag	UNP P23610
B	-9	SER	-	expression tag	UNP P23610
B	-8	GLY	-	expression tag	UNP P23610
B	-7	ARG	-	expression tag	UNP P23610
B	-6	GLU	-	expression tag	UNP P23610
B	-5	ASN	-	expression tag	UNP P23610
B	-4	LEU	-	expression tag	UNP P23610
B	-3	TYR	-	expression tag	UNP P23610
B	-2	PHE	-	expression tag	UNP P23610
B	-1	GLN	-	expression tag	UNP P23610
B	0	GLY	-	expression tag	UNP P23610

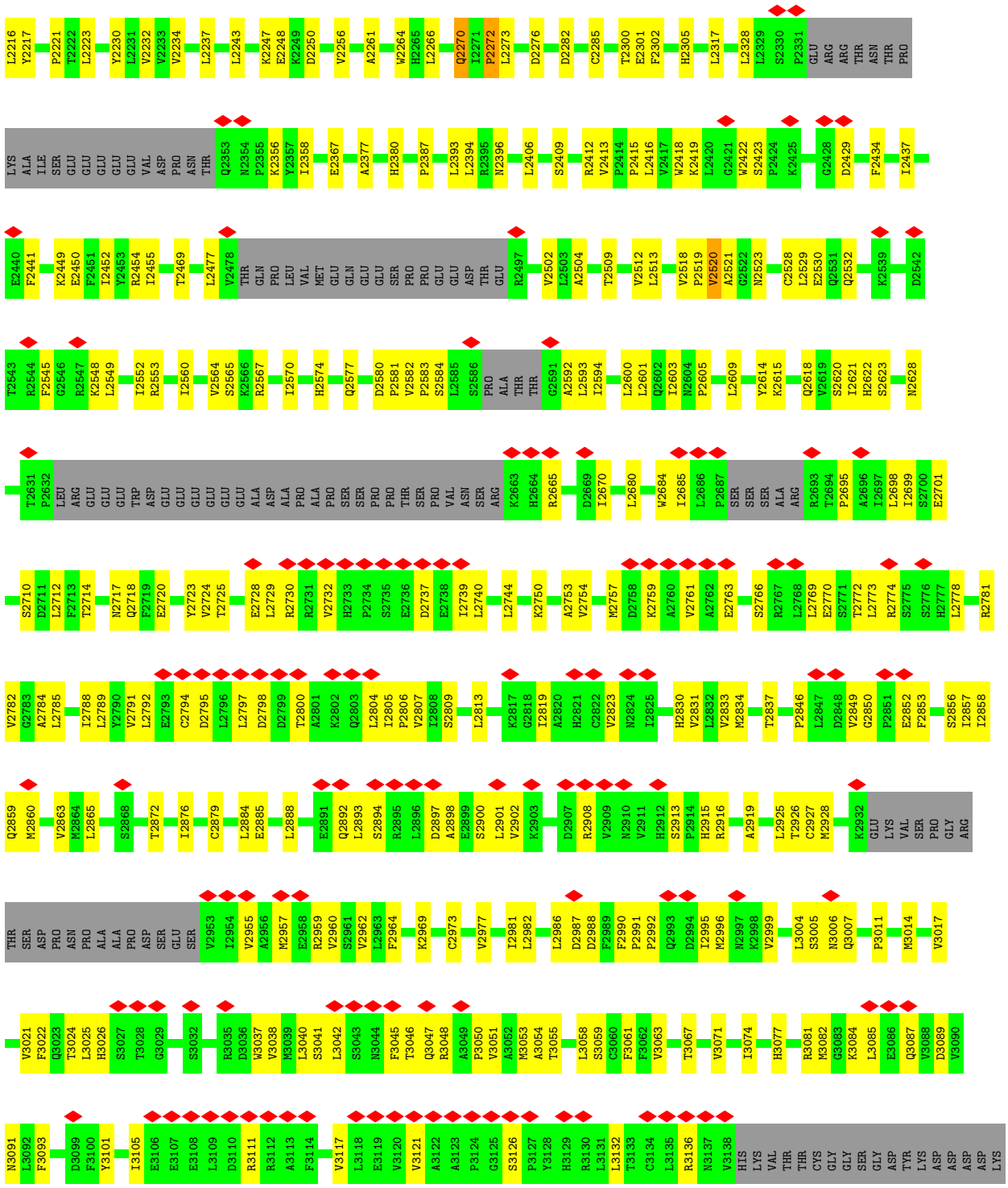
3 Residue-property plots ⓘ

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: Huntingtin

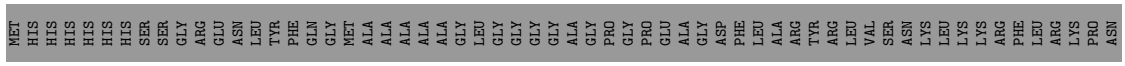


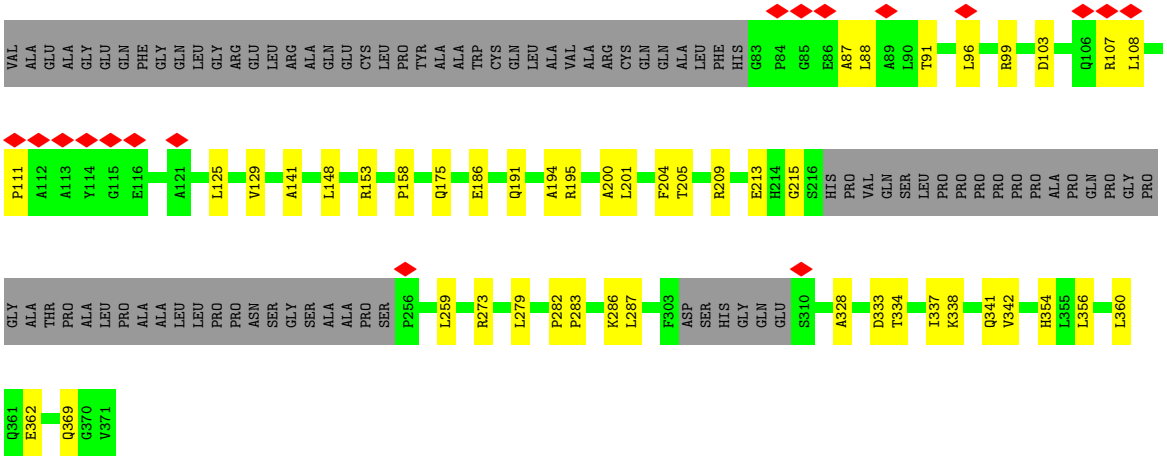

Y863 V866 L874 A875 E876 I877 D878 F879 R880 H893 R894 G895 A896 H897 L903 V909 V913 G919 A930 I934 R935 L936 V937 P938 K939 L940 F941 Y942 K943 C944 D945 Q959 L964 L967 M968 H969 E970 THR GLN PRO PRO SER HIS PHE SER VAL SER THR
ILE T983 R984 I985 Y986 R987 G988 Y989 N990 P993 S994 I995 E1001 N1002 V1007 L1015 T1020 R1021 A1022 L1023 E1029 L1033 V1040 C1041 G1046 W1047 H1048 C1049 GLY VAL PRO PRO LEU SER ALA SER ASP GLU SER ARG LYS SER C1064 T1065 V1066 G1067 M1068 M1071 I1072 L1073
T1074 L1075 L1076 A1079 L1085 L1098 L1099 A1100 A1101 S1102 A1103 P1104 K1105 S1106 L1107 R1108 S1109 SER TRP ALA SER GLU GLU GLU ALA ASN PRO ALA ALA THR LYS GLN GLU E1126 V1127 W1128 P1129 A1130 L1131 R1134 V1137 P1138 M1139 Q1142 L1143 L1147 D1160 G1164 PRO ALA ILE
LYS ALA ALA LEU PRO SER LEU THR ASN PRO PRO SER LEU SER PRO ILE ARG ARG LYS GLY LYS GLU GLU PRO GLY GLU GLU GLN ALA SER VAL PRO LEU SER PRO LYS LYS GLY SER GLU ALA SER ALA ALA SER ARG GLN SER ASP THR SER GLY PRO VAL THR THR SER LYS SER SER
S1228 H1249 K1253 V1254 T1255 L1256 D1257 L1258 V1275 L1276 S1277 Q1278 I1279 L1280 A1283 I1288 E1293 Q1316 T1320 N1325 L1326 A1327 D1331 GLY LEU SER SER ASN PRO SER LYS SER GLN GLY ARG ALA GLN ARG LEU GLY SER SER SER VAL R1353 L1377 ARG ASN MET
VAL GLN ALA GLU GLN GLU ASN ASP THR SER GLY TRP PHE ASP VAL LEU GLN LYS VAL SER THR GLN LEU LYS THR ASN LEU THR SER VAL THR LYS ASN ARG ALA ASP LYS ASN ALA ILE HIS ASN H1423 I1424 R1425 E1428 L1435 K1436 Q1437 T1440 T1441 L1446 L1454 L1457
L1466 L1467 D1468 S1469 D1470 Q1471 V1472 F1473 I1474 G1475 E1493 L1503 V1504 S1507 R1510 Y1511 H1512 S1513 K1514 Q1515 I1516 I1517 G1518 I1519 M1530 R1534 V1537 T1538 I1541 L1544 D1550 R1555 GLY THR ASN LYS ALA ASP ALA G1563 K1564 E1565 L1566 E1567 E1571 V1572
M1576 L1577 L1578 E1588 M1589 F1590 I1591 C1597 E1600 K1604 R1607 I1612 I1615 I1616 L1617 P1618 K1622 D1628 S1629 H1630 E1631 A1632 L1633 L1648 R1649 P1650 L1654 S1657 M1658 M1665 A1666 S1667 V1668 V1671 Q1672 I1678 L1679 A1680 I1681 V1684 I1708 L1716
R1717 D1718 G1719 D1720 S1721 THR SER THR LEU GLU GLU HIS SER GLU GLY LYS GLN ILE K1735 N1736 E1739 R1744 L1747 Q1748 L1749 E1755 T1759 K1760 Q1761 L1762 K1763 M1766 S1767 Q1770 Y1774 C1775 Q1776 E1777 L1778 G1779 T1780 L1781 L1782 K1790 A1800 A1801 L1804 F1805
G1809 C1810 G1811 D1818 L1822 R1823 I1828 T1829 T1830 L1834 V1835 L1836 L1837 W1838 C1839 Q1840 I1841 L1844 W1852 V1856 K1861 ARG HIS SER LEU SER SER THR LYS LEU LEU SER PRO GLN MET SER GLY GLU GLU GLU ASP SER ASP LEU ALA ALA K1887 L1888 G1889 N1892
L1901 D1906 H1913 D1914 S1915 E1916 H1917 L1918 T1919 V1923 D1928 L1932 P1937 F1941 V1945 N1948 A1951 S1952 I1956 Q1960 S1961 R1962 C1963 E1964 N1965 P1969 L1979 I1982 V1990 L1993 Y1994 V1995 R2004 V2005 L2006 R2017 L2022 A2023 A2024 N2025
L2026 Q2027 S2028 S2029 Q2032 L2033 P2034 M2035 E2036 E2037 L2038 N2039 R2040 I2041 Q2042 E2043 L2050 R2053 H2054 F2064 R2065 L2066 M2069 GLN ASP SER LEU SER SER PRO SER PRO PRO VAL SER SER HIS PRO LEU LEU ASP GLY ASP GLY HIS VAL SER LEU LEU GLU THR V2095 S2096 P2097 W2101 Y2102
K2107 W2111 T2112 R2113 S2114 D2115 L2118 L2119 L2124 E2131 M2137 E2141 F2142 N2143 L2144 S2145 L2146 P2149 L2153 S2156 K2163 E2168 E2172 V2173 R2177 V2178 S2179 V2182 Q2183 Q2184 L2185 E2195 L2196 P2197 A2198 E2199 P2200 N2208 D2209 L2210 F2211 G2212 D2213



- Molecule 2: 40-kDa huntingtin-associated protein

Chain B:




VAL ALA GLU ALA GLY GLU GLN PHE GLY GLN LEU LEU GLY ARG GLU LEU ARG ALA ALA GLN GLU CYS LEU PRO TYR ALA ALA TRP CYS GLN LEU ALA VAL ALA ARG CYS GLN GLN ALA LEU PHE HIS G83 P84 G85 E86 A87 L88 A89 L90 T91 L96 R99 D103 Q106 R107 L108
P111 A112 A113 Y114 G115 E116 A121 L125 V129 A141 L148 R153 P158 Q175 E186 Q191 A194 R195 A200 L201 F204 T205 R209 E213 H214 G215 S216 HIS PRO VAL GLN SER LEU PRO PRO PRO PRO PRO ALA PRO GLN PRO GLY PRO
GLY ALA THR PRO ALA LEU PRO ALA ALA LEU LEU PRO PRO ASN SER GLY SER ALA ALA PRO SER P256 L259 R273 L279 P282 P283 K286 L287 F303 ASP SER HIS GLY GLN GLU S310 A328 D333 T334 I337 K338 Q341 V342 H354 L355 L356 L360
Q361 E362 Q369 G370 V371

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	647468	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 ($e^-/Å^2$)	48	Depositor
Minimum defocus (nm)	1200	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	165000	Depositor
Image detector	GATAN K2 QUANTUM (4k x 4k)	Depositor
Maximum map value	0.246	Depositor
Minimum map value	-0.119	Depositor
Average map value	-0.000	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.035	Depositor
Map size (Å)	368.256, 368.256, 368.256	wwPDB
Map dimensions	448, 448, 448	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.822, 0.822, 0.822	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

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	0/19436	0.47	0/26395
2	B	0.30	0/1868	0.42	0/2539
All	All	0.33	0/21304	0.47	0/28934

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 (i)

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	19063	0	19505	588	0
2	B	1836	0	1877	42	0
All	All	20899	0	21382	615	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 15.

The worst 5 of 615 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:2211:PHE:HE1	1:A:2264:TRP:CB	1.39	1.33
2:B:279:LEU:HD12	2:B:362:GLU:OE1	1.42	1.15

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:98:LYS:HG2	1:A:101:ARG:NH2	1.61	1.15
1:A:2211:PHE:CE1	1:A:2264:TRP:CB	2.31	1.14
1:A:930:ALA:HB1	1:A:1033:LEU:HD12	1.30	1.12

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	2387/3156 (76%)	2226 (93%)	153 (6%)	8 (0%)	41	64
2	B	238/389 (61%)	226 (95%)	12 (5%)	0	100	100
All	All	2625/3545 (74%)	2452 (93%)	165 (6%)	8 (0%)	44	64

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2520	VAL
1	A	2521	ALA
1	A	2210	LEU
1	A	2272	PRO
1	A	3126	SER

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	2144/2778 (77%)	2139 (100%)	5 (0%)	93	98
2	B	185/290 (64%)	185 (100%)	0	100	100
All	All	2329/3068 (76%)	2324 (100%)	5 (0%)	93	98

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	109	CYS
1	A	1590	PHE
1	A	1774	TYR
1	A	2017	ARG
1	A	2545	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	2628	ASN
1	A	2270	GLN
1	A	2054	HIS
1	A	1761	GLN
1	A	2134	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.

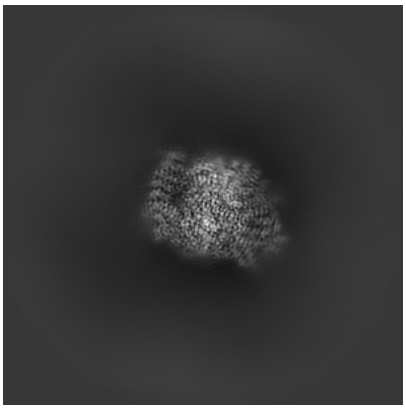
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-22106. These allow visual inspection of the internal detail of the map and identification of artifacts.

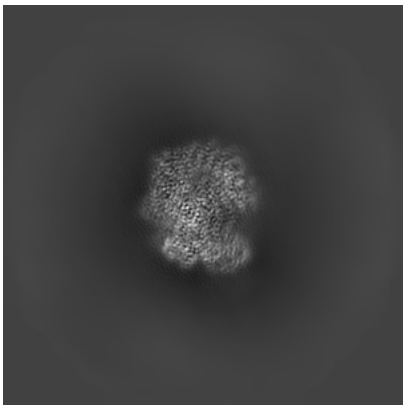
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections (i)

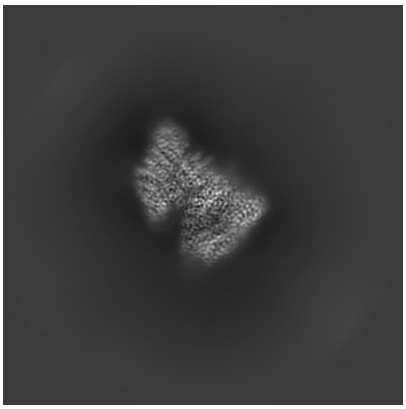
6.1.1 Primary map



X

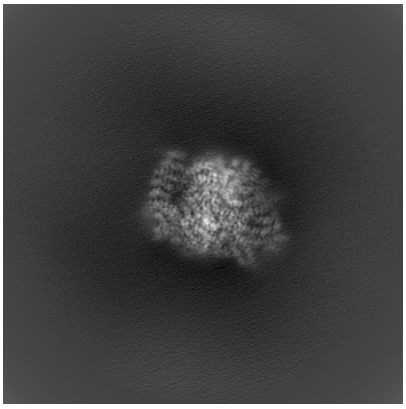


Y

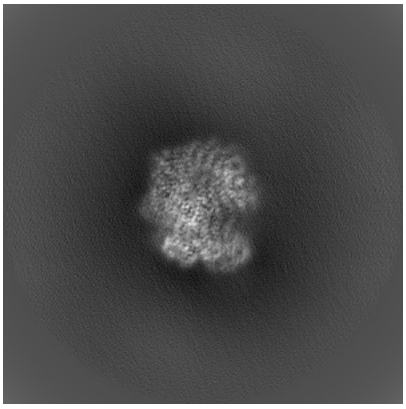


Z

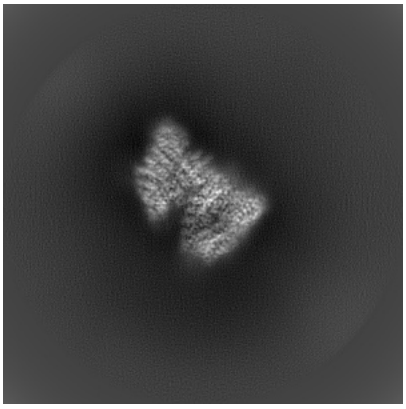
6.1.2 Raw map



X



Y

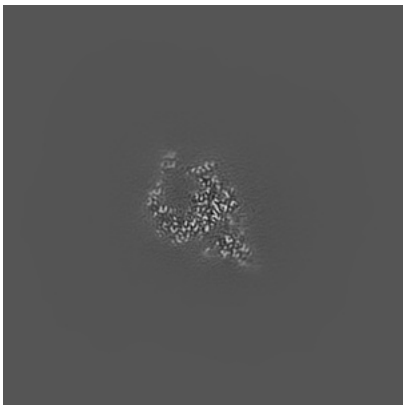


Z

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

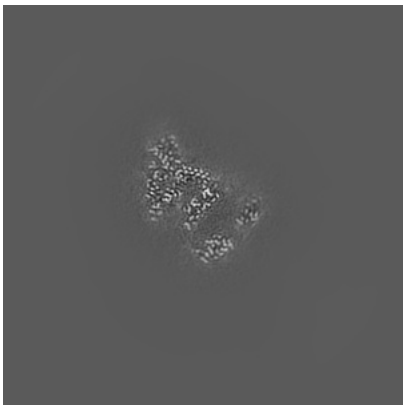
6.2 Central slices (i)

6.2.1 Primary map



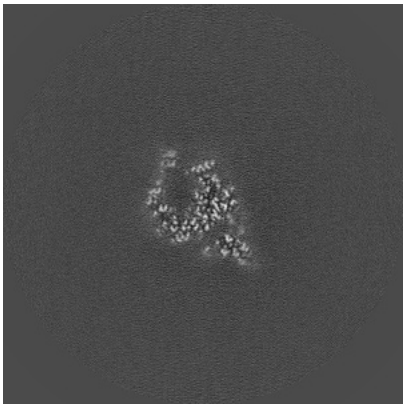
X Index: 224

Y Index: 224

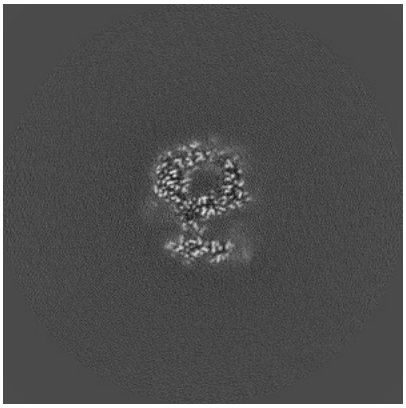


Z Index: 224

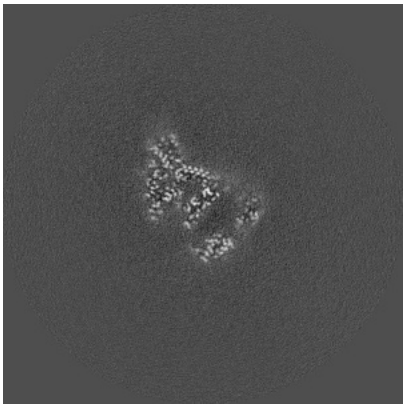
6.2.2 Raw map



X Index: 224



Y Index: 224

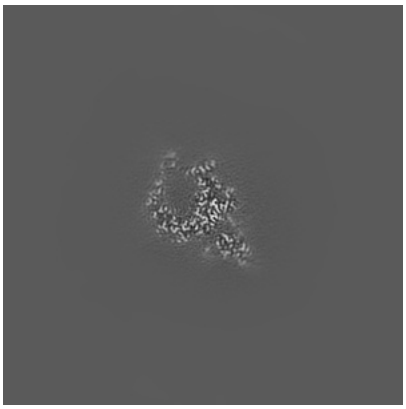


Z Index: 224

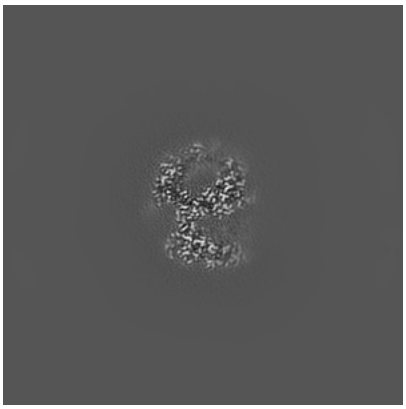
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

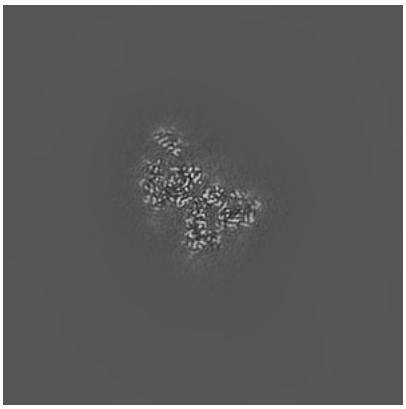
6.3.1 Primary map



X Index: 225

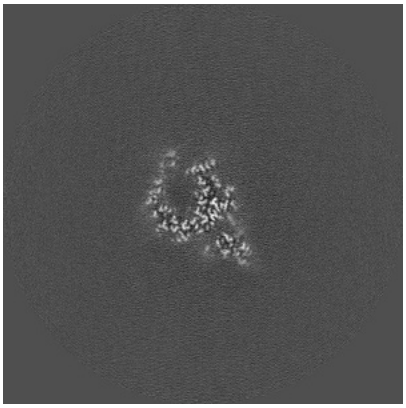


Y Index: 230

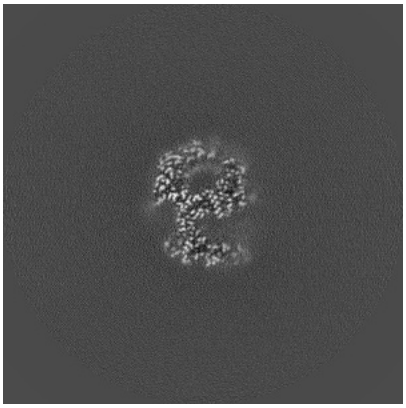


Z Index: 202

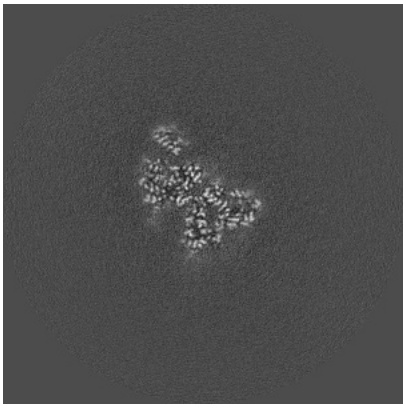
6.3.2 Raw map



X Index: 225



Y Index: 228

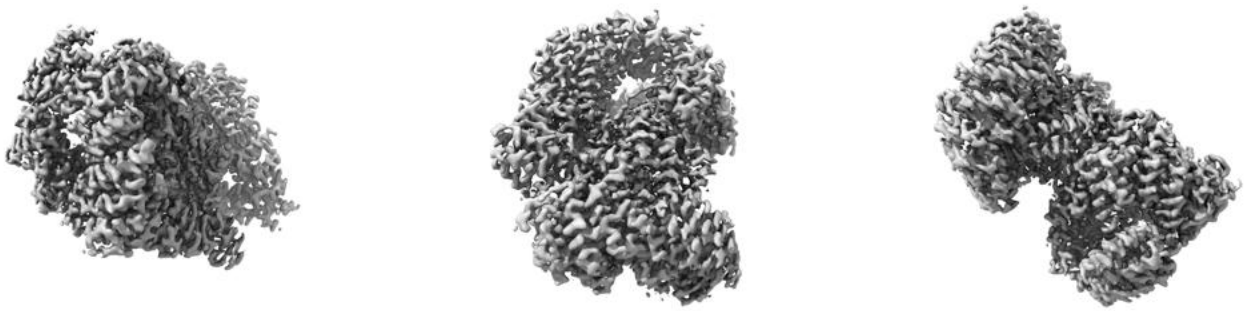


Z Index: 202

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

6.4 Orthogonal surface views (i)

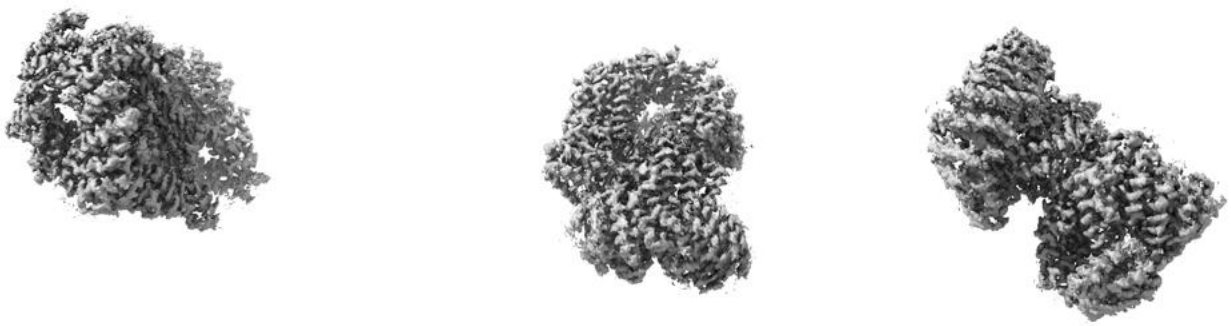
6.4.1 Primary map



X Y Z

The images above show the 3D surface view of the map at the recommended contour level 0.035. 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.

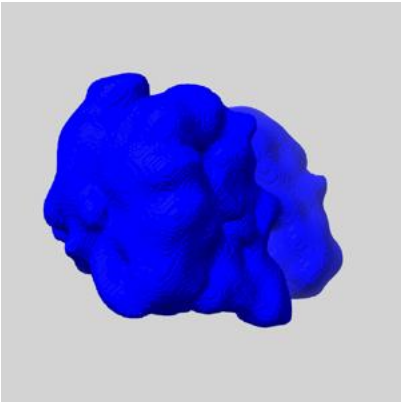
6.5 Mask visualisation (i)

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

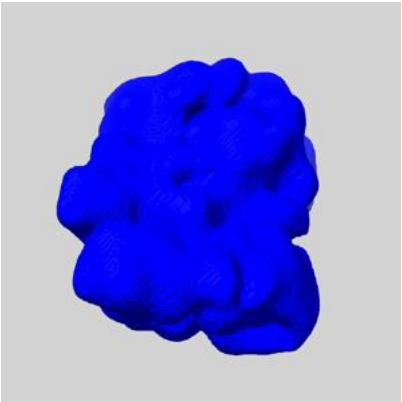
A mask typically either:

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

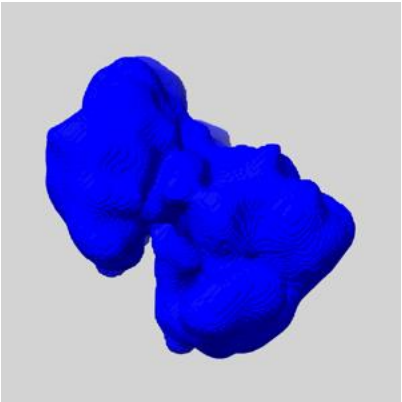
6.5.1 emd_22106_msk_1.map (i)



X



Y

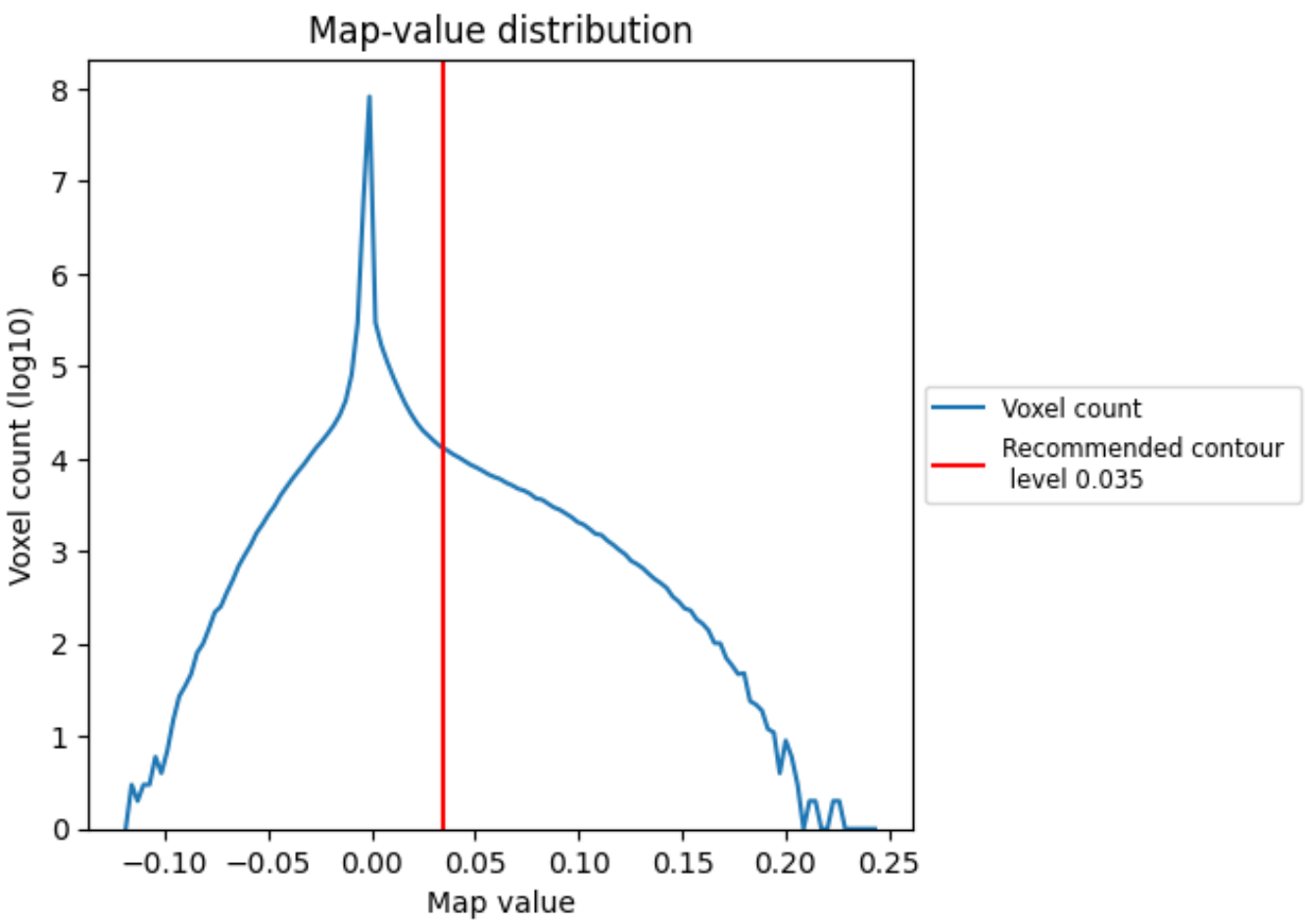


Z

7 Map analysis (i)

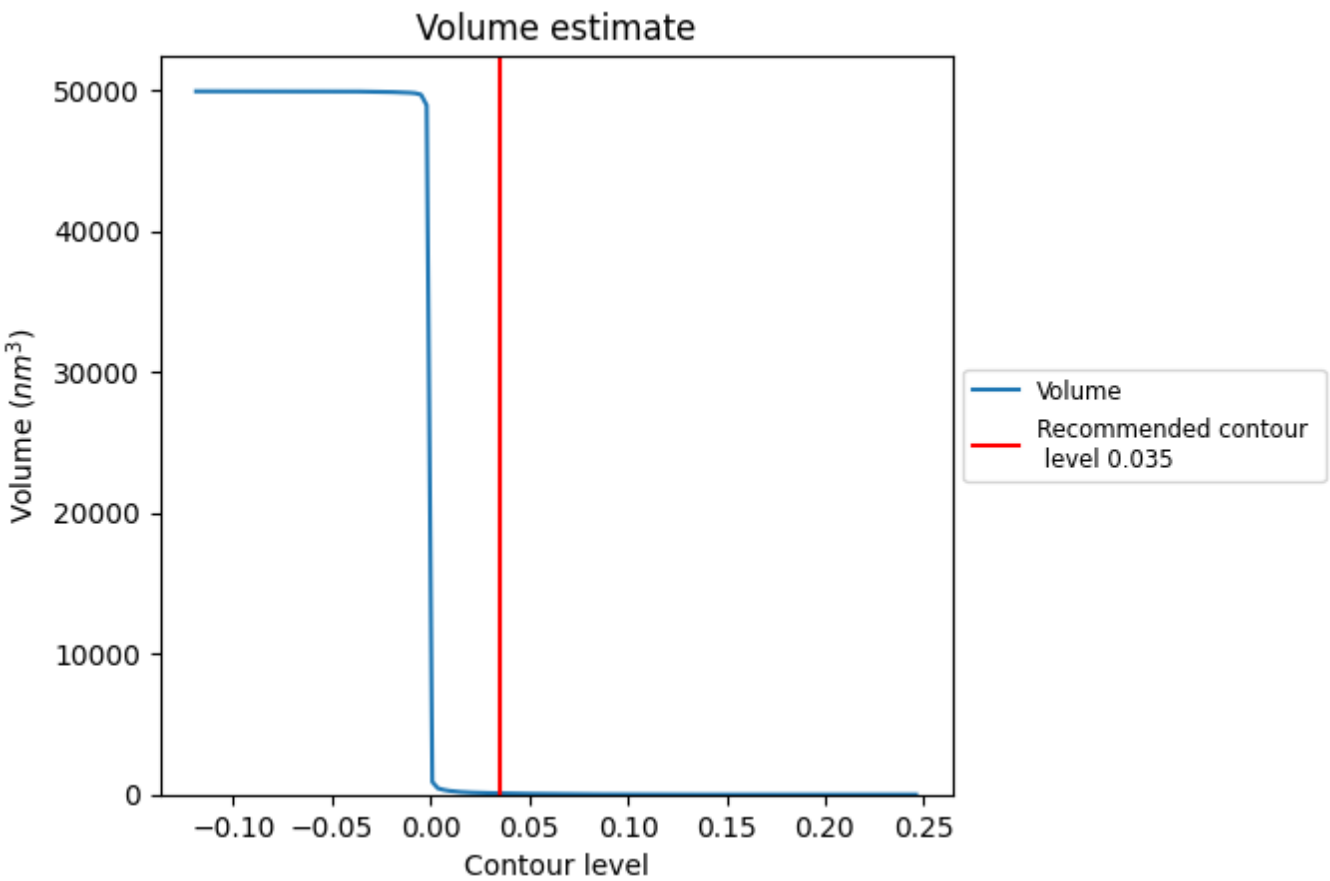
This section contains the results of statistical analysis of the map.

7.1 Map-value distribution (i)



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.

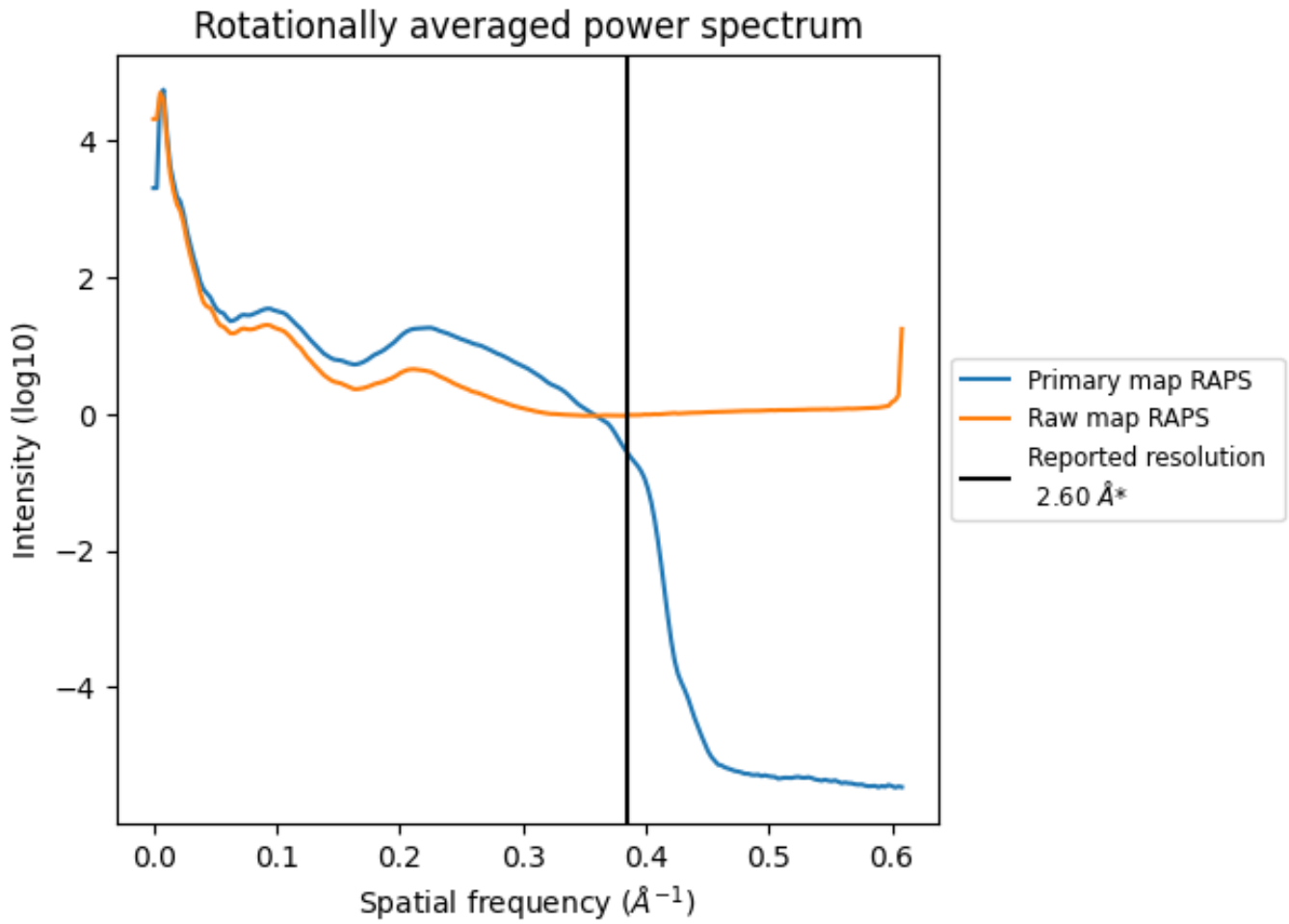
7.2 Volume estimate (i)



The volume at the recommended contour level is 87 nm^3; this corresponds to an approximate mass of 79 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)

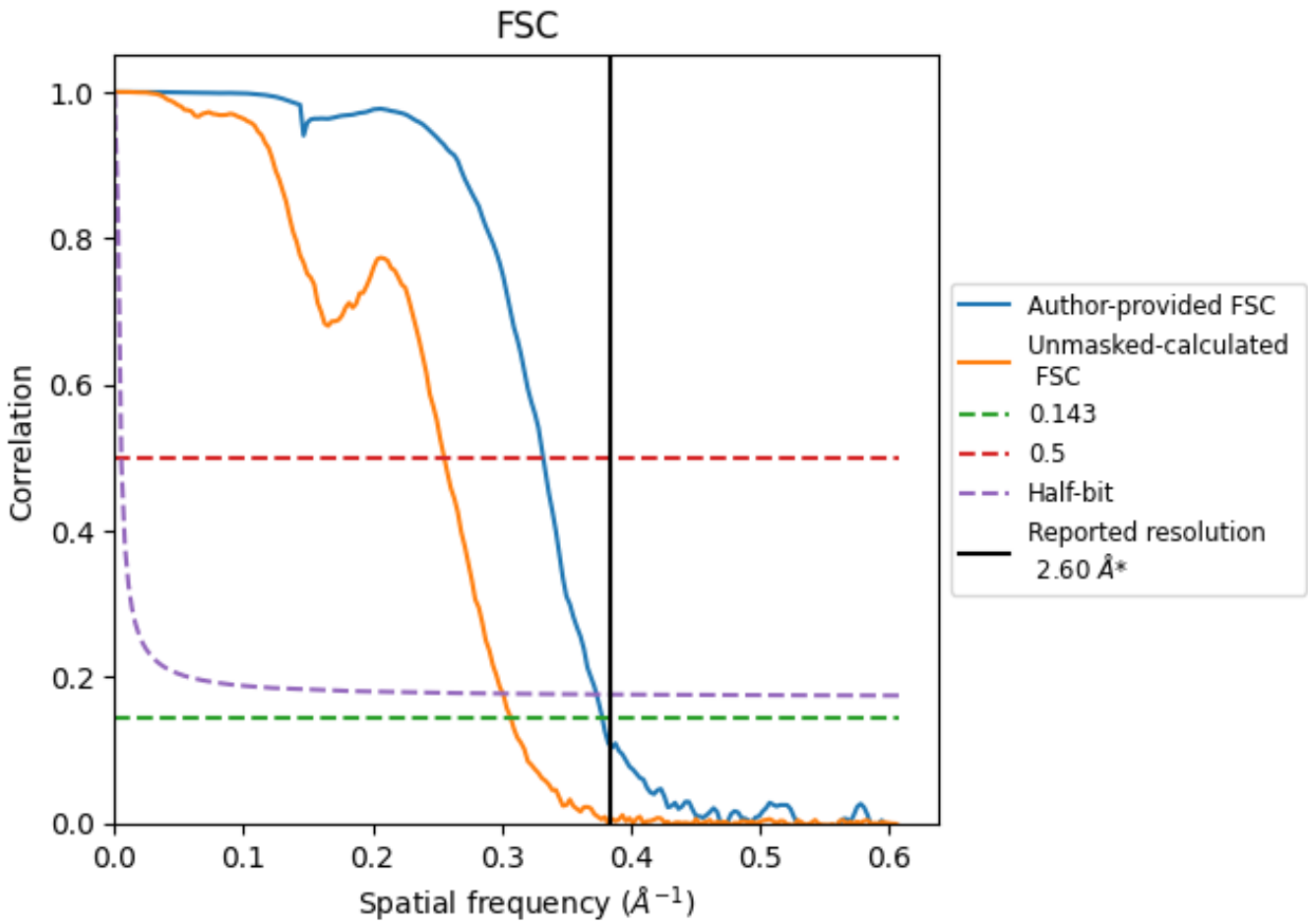


*Reported resolution corresponds to spatial frequency of 0.385 Å^{-1}

8 Fourier-Shell correlation (i)

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 (i)



*Reported resolution corresponds to spatial frequency of 0.385 Å^{-1}

8.2 Resolution estimates (i)

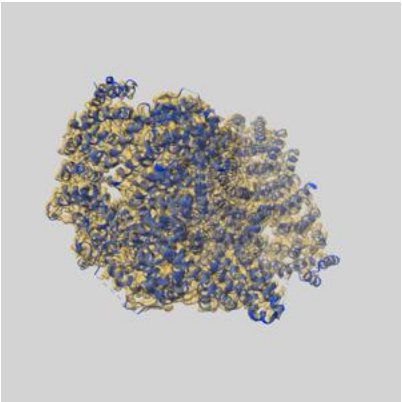
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	2.60	-	-
Author-provided FSC curve	2.64	3.01	2.68
Unmasked-calculated*	3.26	3.91	3.32

*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 3.26 differs from the reported value 2.6 by more than 10 %

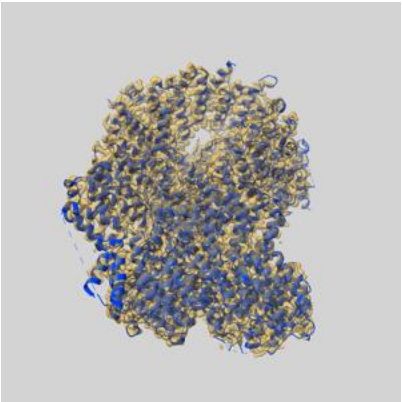
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-22106 and PDB model 6X9O. Per-residue inclusion information can be found in section 3 on page 5.

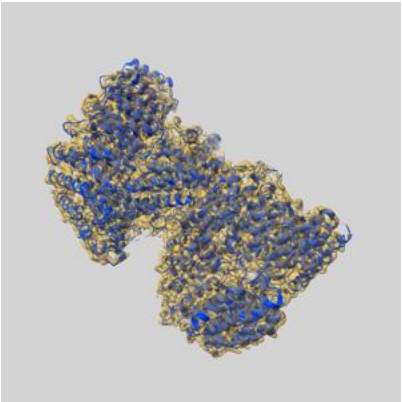
9.1 Map-model overlay (i)



X



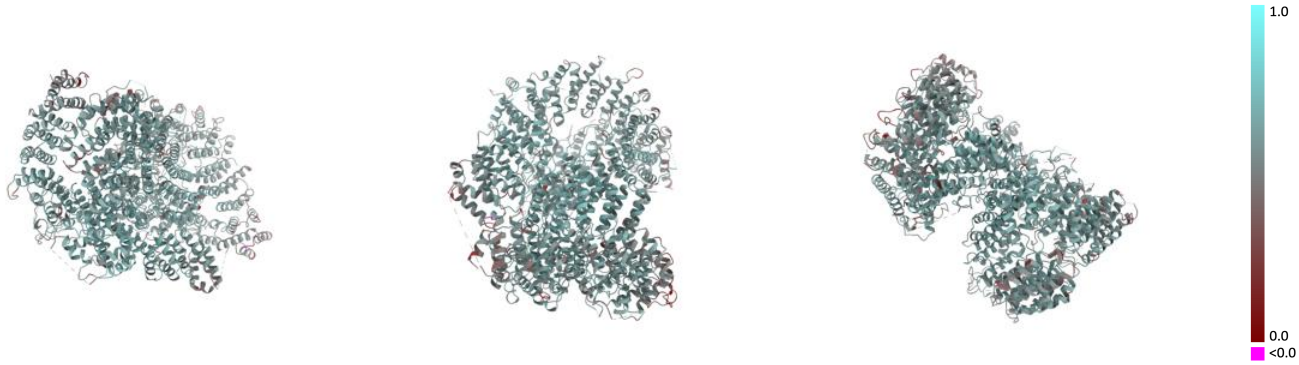
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.035 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.

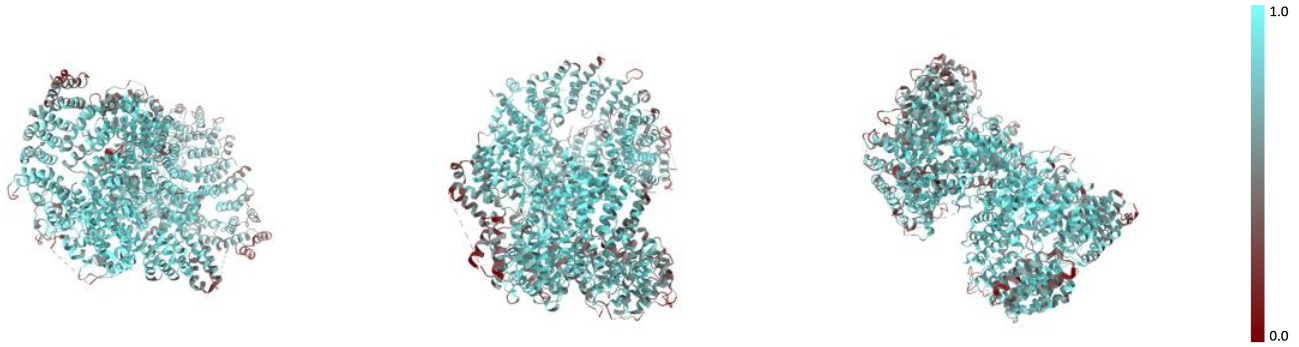
9.2 Q-score mapped to coordinate model (i)



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.

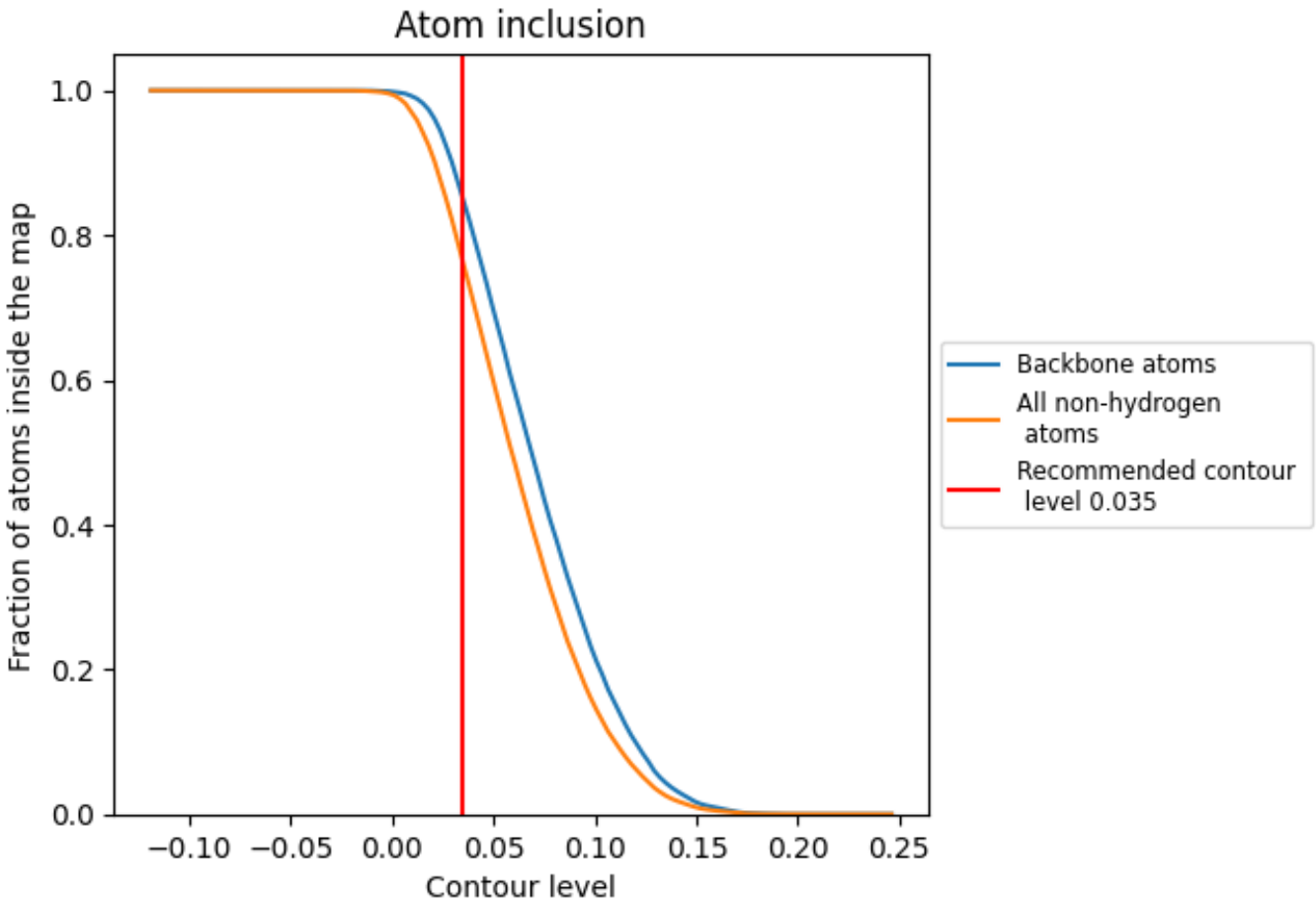
9.3 Atom inclusion mapped to coordinate model (i)



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 (0.035).

9.4 Atom inclusion (i)



At the recommended contour level, 85% of all backbone atoms, 76% 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 (0.035) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7611	0.5790
A	0.7560	0.5760
B	0.8136	0.6140