



wwPDB EM Validation Summary Report ⓘ

Feb 23, 2023 – 12:14 PM EST

PDB ID : 8FJP
EMDB ID : EMD-29245
Title : Cryo-EM structure of native mosquito salivary gland surface protein 1 (SGS1)
Authors : Liu, S.; Xia, X.; Calvo, E.; Zhou, Z.H.
Deposited on : 2022-12-20
Resolution : 3.30 Å (reported)

This is a wwPDB EM Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

EMDB validation analysis : 0.0.1.dev43
Mogul : 1.8.5 (274361), CSD as541be (2020)
MolProbity : 4.02b-467
buster-report : 1.1.7 (2018)
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ : 1.9.9
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.32.1

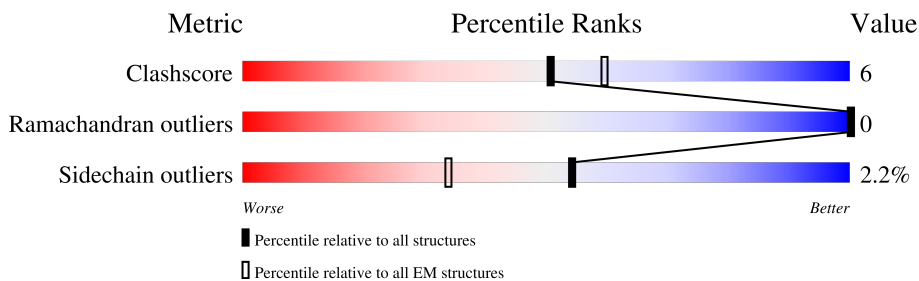
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.30 Å.

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



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

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 $\leq 5\%$. 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	3364	
2	B	2	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 23390 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 salivary gland surface protein 1.

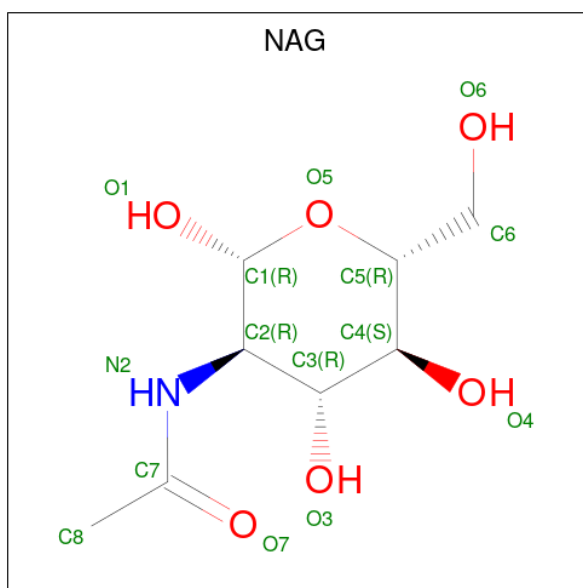
Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	2932	23348	14826	3974	4477	71	0	0

- Molecule 2 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
2	B	2	28	16	2	10	0	0

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C₈H₁₅NO₆) (labeled as "Ligand of Interest" by depositor).

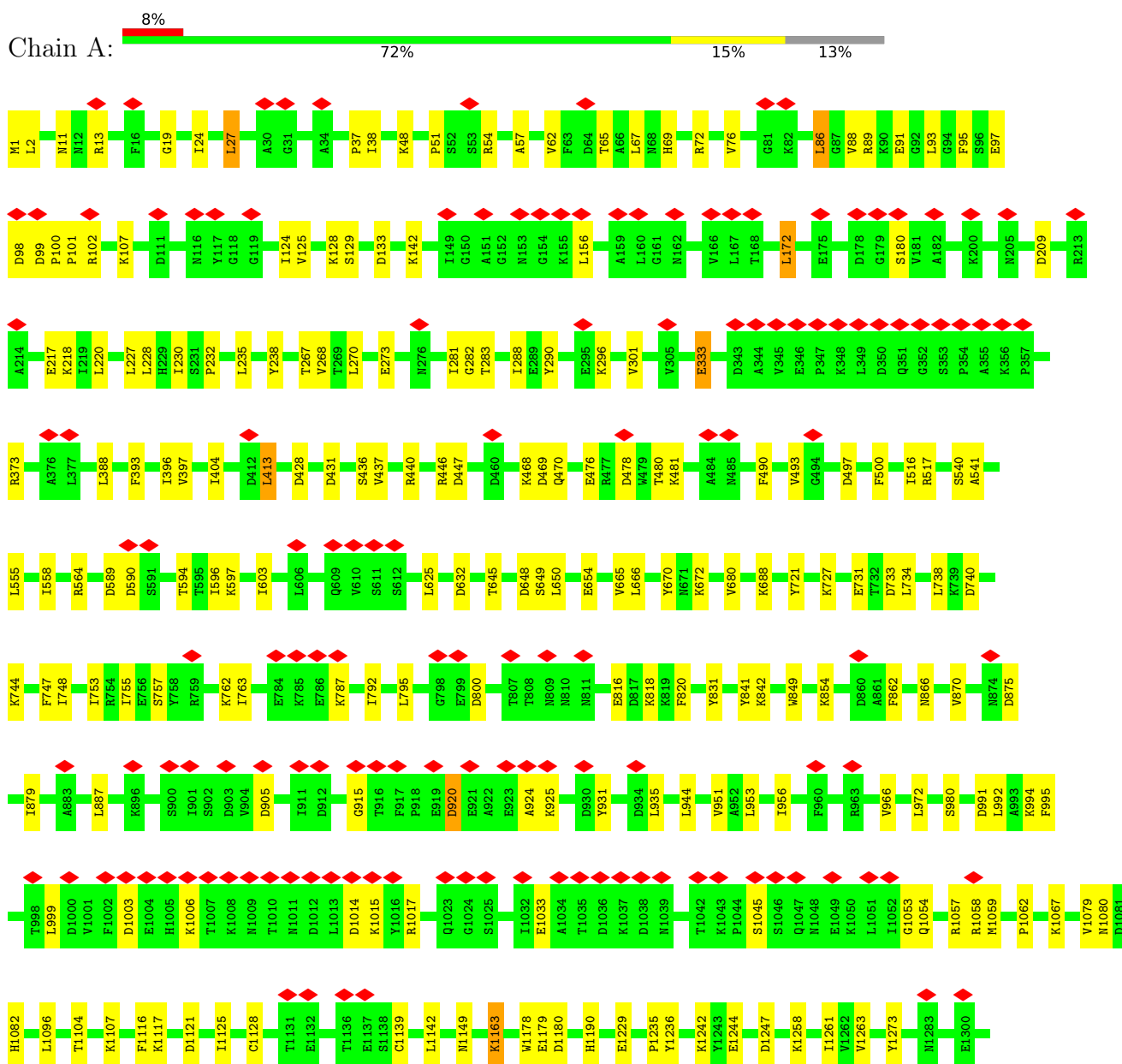


Mol	Chain	Residues	Atoms				AltConf
			Total	C	N	O	
3	A	1	14	8	1	5	0

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: salivary gland surface protein 1



ALA	PHE	THR	W2967	R2968	K2971	D2972	T2973	L2974	L2976	K2980	S2986	K2987	Q2988	L2989	R2915	M2920	I2921	GLU	LYS	ALA	ASP	LEU	ALA	LYS	THR	ASP	PHE	ILE	ILE	ILE	HIS	ARG	GLY	PHE	PHE	MET	PRO	TRP	GLN	ARG	ILE	GLY	TVR	ALA	ALA	ILE	HIS	VAL	MET	PRO	ASN	PRO	ILE	PRO	SER	LYS	F2868
L2889	L2890	F2891	Y2892	L2893	L2894	G2895	G2896	A2899	Q2905	I2906	D2910	W2911	K2912	S2913	R2914	R2915	M2920	I2921	GLU	LYS	ALA	ASP	LEU	ALA	LYS	THR	ASP	PHE	ILE	ILE	HIS	ARG	GLY	PHE	PHE	MET	PRO	TRP	GLN	ARG	ILE	GLY	TVR	ALA	ALA	ILE	HIS	VAL	MET	PRO	ASN	PRO	ILE	PRO	SER	LYS	F2868
GLY	ALA	ALA	THR	PHE	SER	THR	PHE	ALA	MET	PHE	GLY	GLY	SER	MET	ALA	GLY	A2809	L2810	A2811	V2814	A2820	F2821	G2822	G2823	A2824	A2825	A2826	S2828	M2829	Q2830	W2831	M2832	P2833	A2834	K2835	W2836	D2837	M2845	I2853	A2854	V2855	S2856	F2857	P2858	S2873	L2874	M2875	M2884	F2888								
G2399	K2400	G2401	G2402	D2403	S2404	S2405	I2406	W2410	D2414	R2415	T2416	E2417	E2418	C2421	K2422	R2423	Q2426	Q2427	A2343	N2344	A2345	G2346	M2347	K2348	S2349	H2350	E2351	R2354	F2355	R2361	E2367	A2371	E2374	G2375	T2376	Q2377	Q2378	L2379	T2380	T2381	V2384	P2391	K2392	L2393	L2396	L2397	E2398										
D2104	K2105	V2106	S2107	L2122	R2129	R2137	W2141	K2141	T2142	S2143	G2144	Q2145	T2149	P2150	L2151	D2155	K2164	N2168	V2169	G2170	D2171	S2172	K2173	K2179	V2180	L2181	K2182	F2183	V2187	E2188	L2198	T2199	Y2200	T2201	E2202	P2209	I2215	T2218	P2234	T2236	A2237	R2238	W2241														
Y1983	Q1984	I1985	I1996	Y1997	F1998	F1999	L2000	N2002	F2003	F2004	G2005	M1841	F1859	R1862	A2015	N2016	P2017	T2018	Q2021	I2033	P2034	N2035	R2036	D2037	Q2038	D2043	R2059	K2063	S2070	E2071	E2077	E2078	K2084	I2087	T2091	E2092	E2093	T2094	L2095	S2096	Q2101	E2102	N2103														
S1819	M1826	L1827	L1828	R1831	H1832	H1836	P1837	G1838	V1839	M1840	M1841	F1859	R1862	K1866	S1877	P1878	K1881	R1890	Y1897	T1898	Q1899	I1900	K1901	D1902	I1906	H1915	M1921	S1931	R1932	S1933	G1934	V1949	F1959	I1960	T1961	S1962	K1963	R1970	I1974																		
V1578	D1679	L1682	L1683	I1684	E1685	M1694	A1697	V1698	I1712	V1713	D1716	K1559	K1560	R1561	Q1570	M1571	L1572	A1573	K1574	V1578	M1586	Q1591	I1592	T1595	A1596	V1597	E1599	K1612	E1613	K1614	F1615	I1622	R1623	I1638	K1653	P1660	A1663	E1664	L1665	V1666	S1675	I1676	W1677														
E1311	D1317	G1318	A1319	Q1320	L1321	D1322	A1323	Q1325	S1343	A1344	N1366	K1367	R1368	M1374	I1375	R1376	K1392	L1396	I1397	A1398	E1399	F1400	T1403	A1404	V1405	F1409	R1458	E1459	I1460	V1461	V1462	S1463	E1464	S1465	V1471	L1472	F1473	K1474	I1475	I1476	A1482	P1494	F1499	Q1516													



- Molecule 2: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain B: 50% 50%

MAG1
MAG2

4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	161092	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 ($e^-/\text{\AA}^2$)	30	Depositor
Minimum defocus (nm)	1500	Depositor
Maximum defocus (nm)	3000	Depositor
Magnification	105000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.189	Depositor
Minimum map value	-0.100	Depositor
Average map value	-0.000	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.03	Depositor
Map size (Å)	408.0, 408.0, 408.0	wwPDB
Map dimensions	300, 300, 300	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.36, 1.36, 1.36	Depositor

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NAG

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	0.30	0/23854	0.59	0/32262

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	23348	0	23006	273	0
2	B	28	0	25	3	0
3	A	14	0	13	0	0
All	All	23390	0	23044	274	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 6.

The worst 5 of 274 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:B:1:NAG:H62	2:B:2:NAG:HN2	1.41	0.85

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:97:GLU:HB2	1:A:102:ARG:HA	1.62	0.80
1:A:99:ASP:HB2	1:A:100:PRO:HD3	1.62	0.80
1:A:97:GLU:CB	1:A:102:ARG:HA	2.18	0.73
1:A:1586:MET:HG2	1:A:1592:ILE:HD11	1.77	0.67

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	2922/3364 (87%)	2756 (94%)	166 (6%)	0	100 100

There are no Ramachandran outliers to report.

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	2546/2935 (87%)	2490 (98%)	56 (2%)	52 74

5 of 56 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1516	GLN
1	A	3041	PHE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	1832	HIS
1	A	2943	TYR
1	A	2757	ASN

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

Mol	Chain	Res	Type
1	A	1570	GLN
1	A	1720	GLN
1	A	1937	GLN
1	A	1763	GLN
1	A	1334	HIS

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](#)

2 monosaccharides are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	NAG	B	1	1,2	14,14,15	0.36	0	17,19,21	0.86	1 (5%)
2	NAG	B	2	2	14,14,15	0.28	0	17,19,21	0.68	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the

Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	NAG	B	1	1,2	-	2/6/23/26	0/1/1/1
2	NAG	B	2	2	-	3/6/23/26	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1	NAG	C1-O5-C5	2.45	115.51	112.19

There are no chirality outliers.

All (5) torsion outliers are listed below:

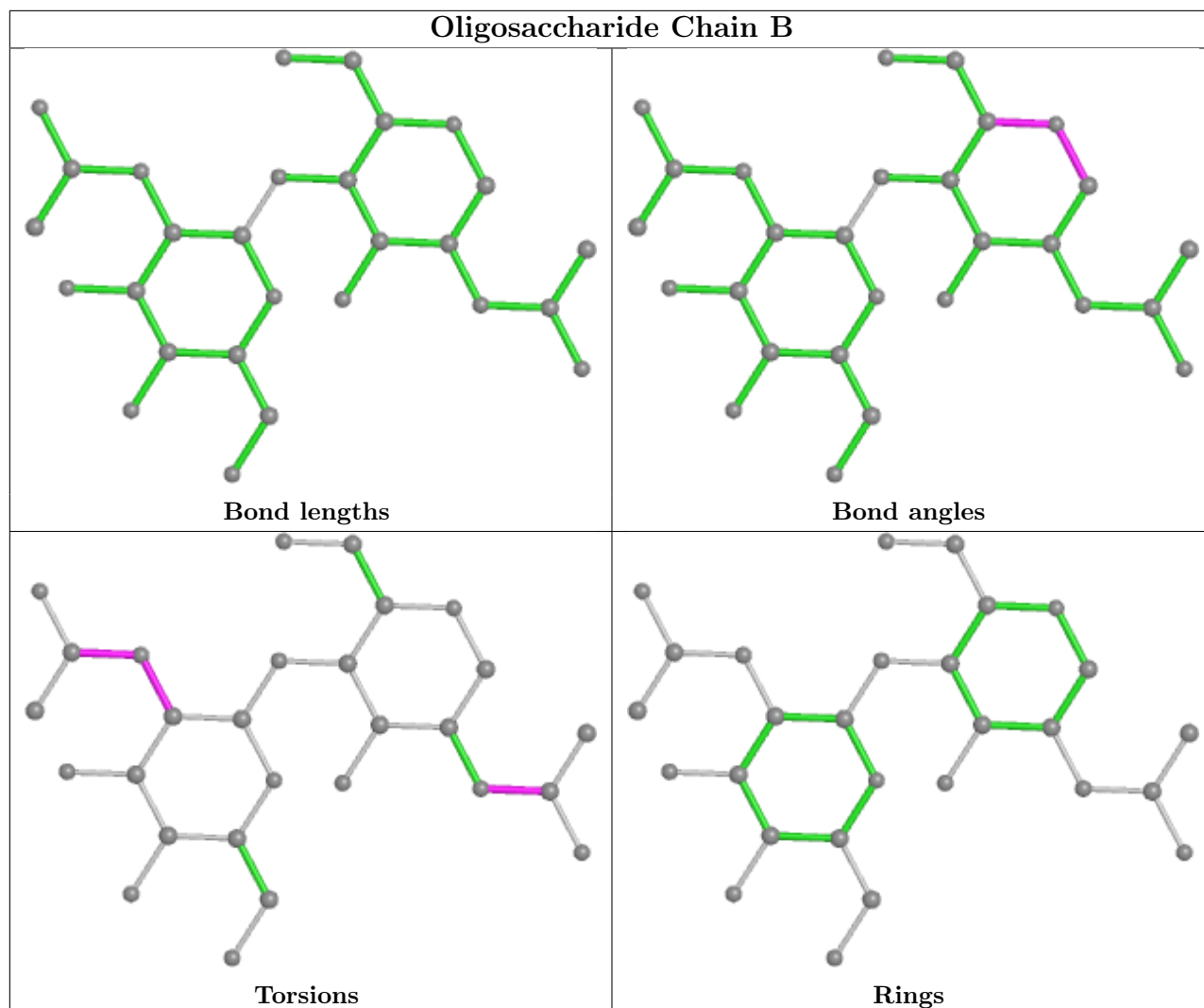
Mol	Chain	Res	Type	Atoms
2	B	2	NAG	C3-C2-N2-C7
2	B	2	NAG	C8-C7-N2-C2
2	B	2	NAG	O7-C7-N2-C2
2	B	1	NAG	C8-C7-N2-C2
2	B	1	NAG	O7-C7-N2-C2

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	2	NAG	2	0
2	B	1	NAG	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.



5.6 Ligand geometry [i](#)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
3	NAG	A	4001	1	14,14,15	0.28	0	17,19,21	0.76	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	A	4001	1	-	1/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

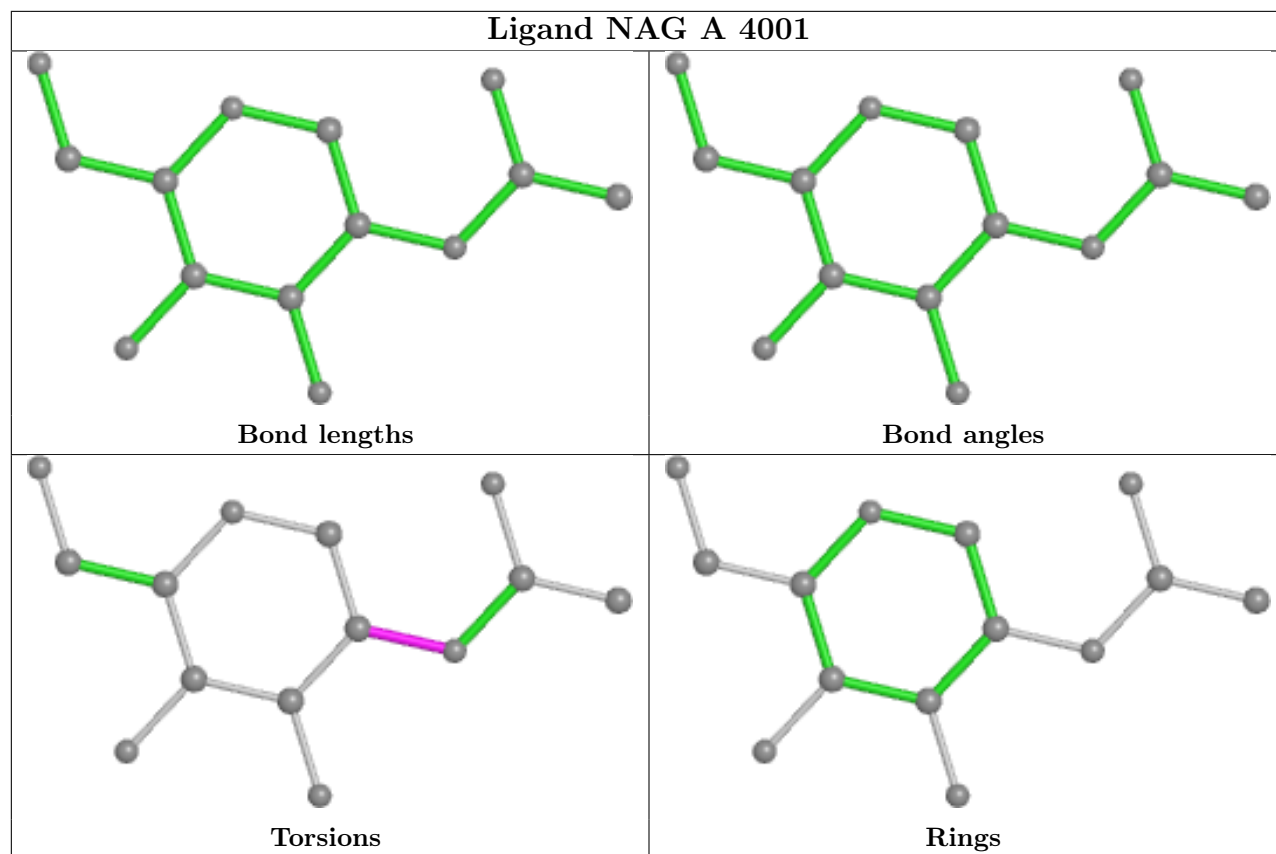
All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	4001	NAG	C1-C2-N2-C7

There are no ring outliers.

No monomer is involved in short contacts.

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



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.

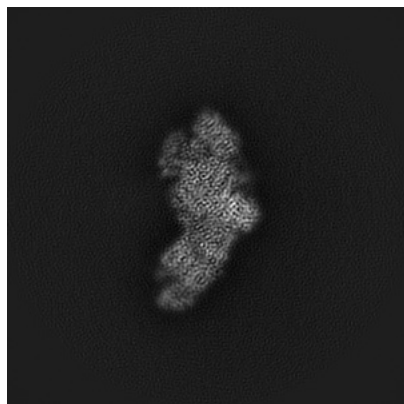
6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-29245. These allow visual inspection of the internal detail of the map and identification of artifacts.

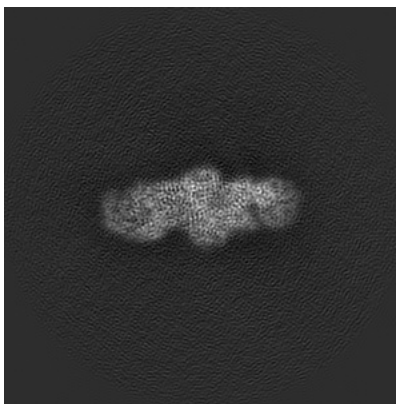
Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

6.1 Orthogonal projections [i](#)

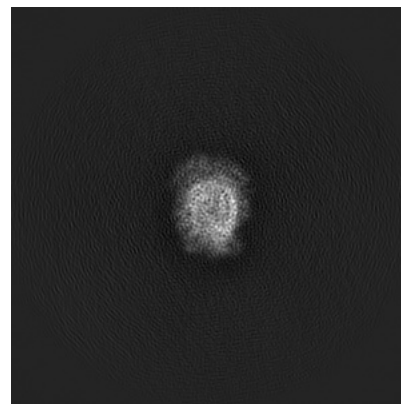
6.1.1 Primary map



X

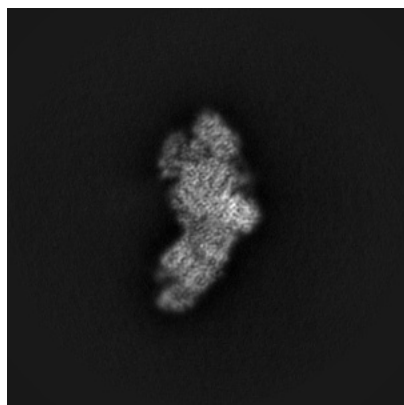


Y

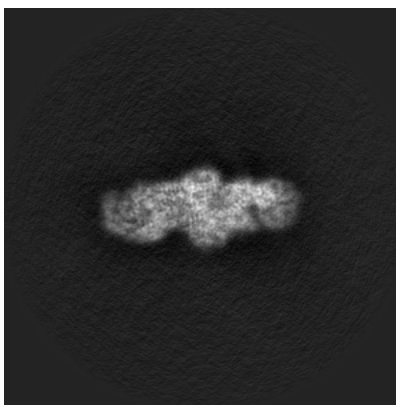


Z

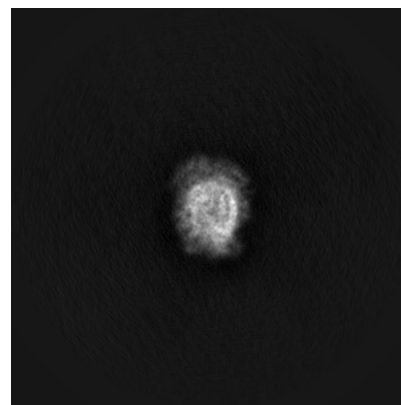
6.1.2 Raw map



X



Y

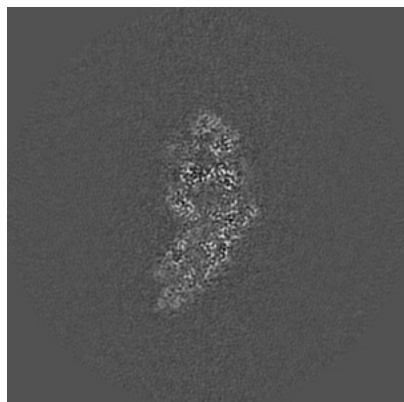


Z

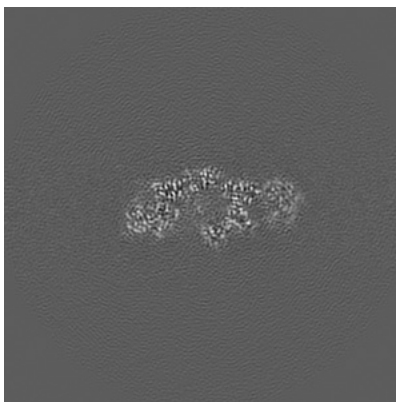
The images above show the map projected in three orthogonal directions.

6.2 Central slices [i](#)

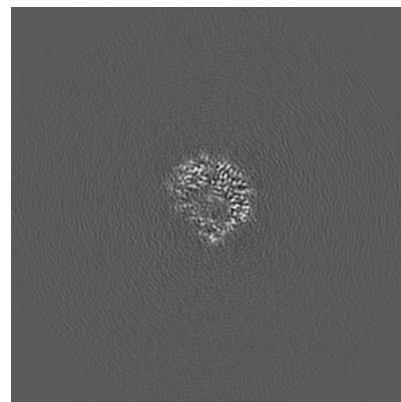
6.2.1 Primary map



X Index: 150

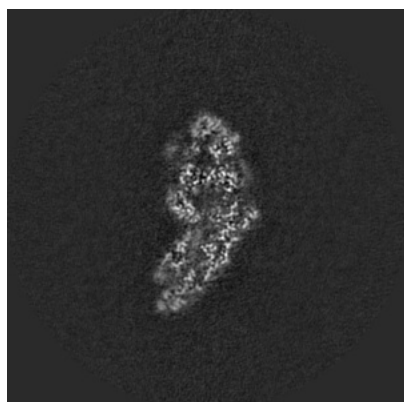


Y Index: 150

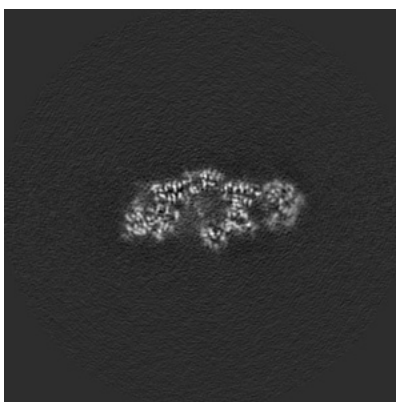


Z Index: 150

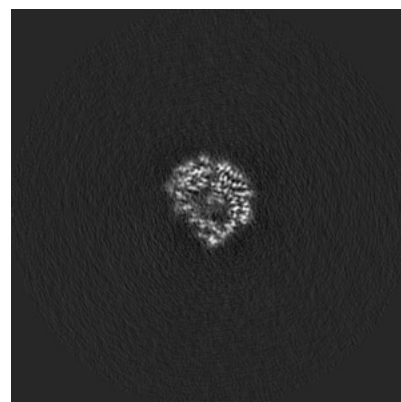
6.2.2 Raw map



X Index: 150



Y Index: 150

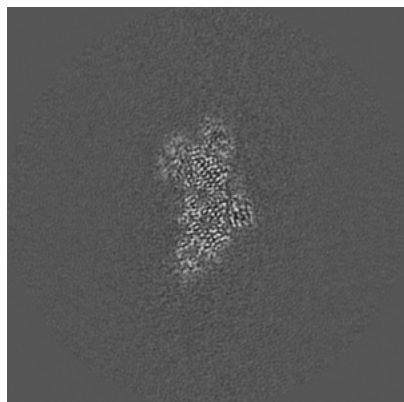


Z Index: 150

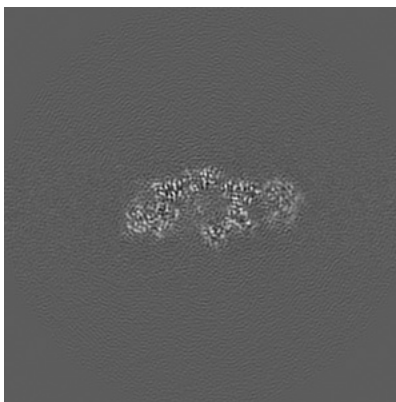
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices [i](#)

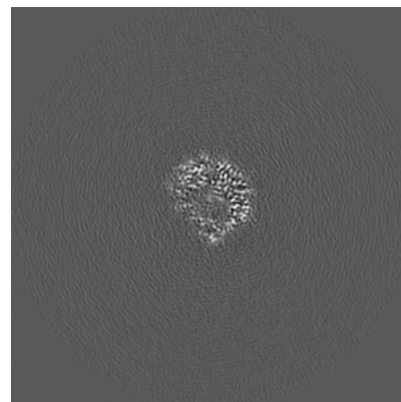
6.3.1 Primary map



X Index: 164

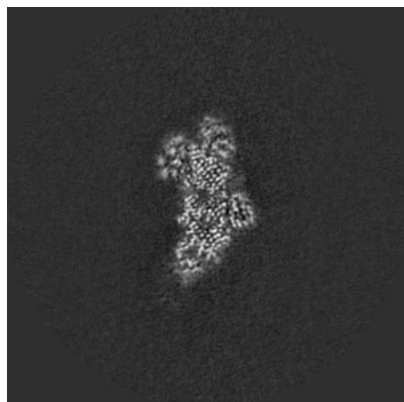


Y Index: 150

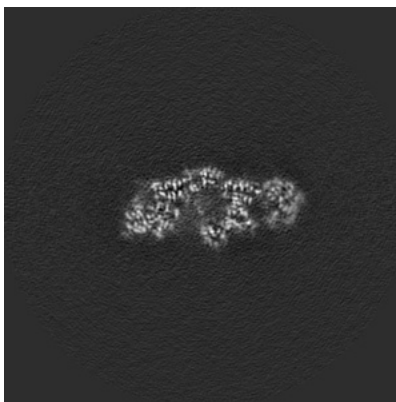


Z Index: 150

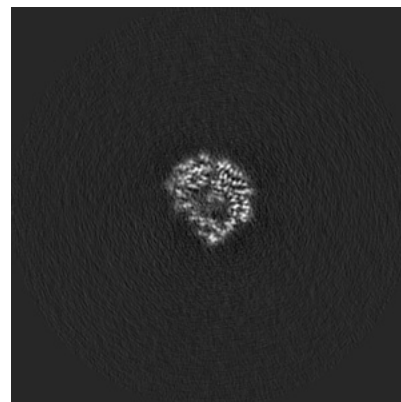
6.3.2 Raw map



X Index: 164



Y Index: 150

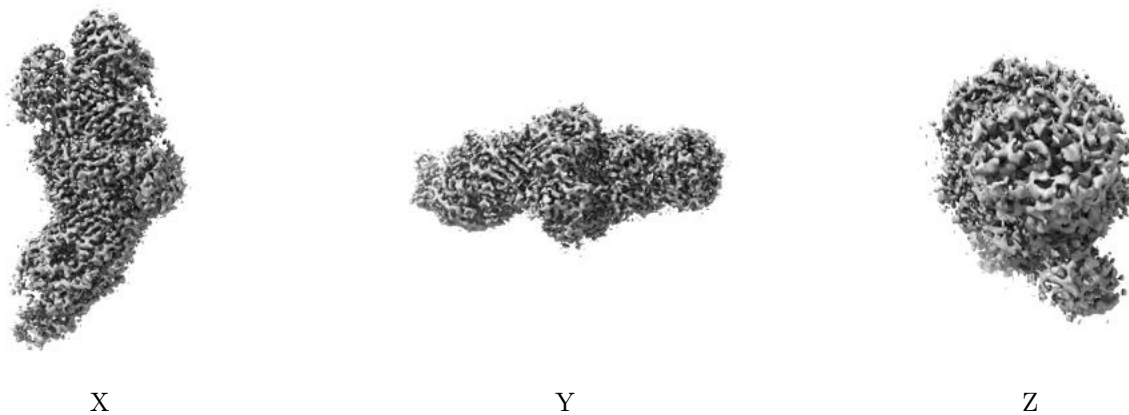


Z Index: 150

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

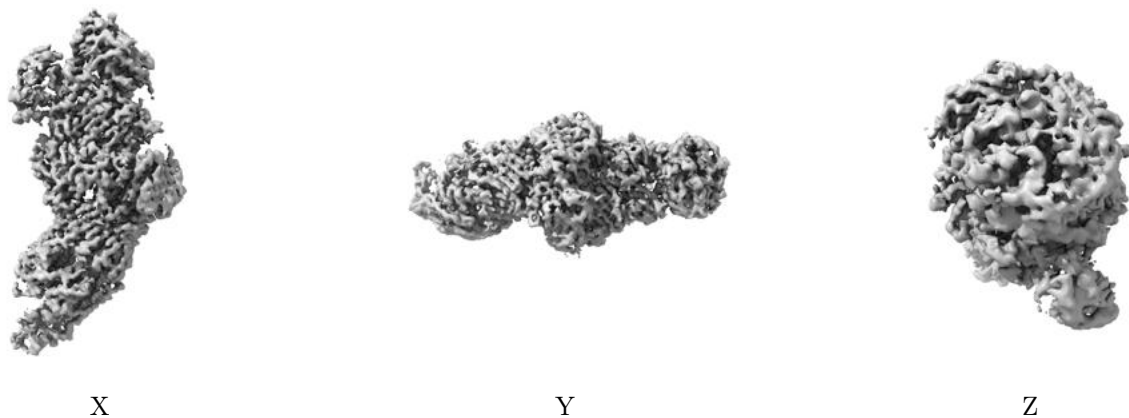
6.4 Orthogonal surface views [i](#)

6.4.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.03. 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



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.

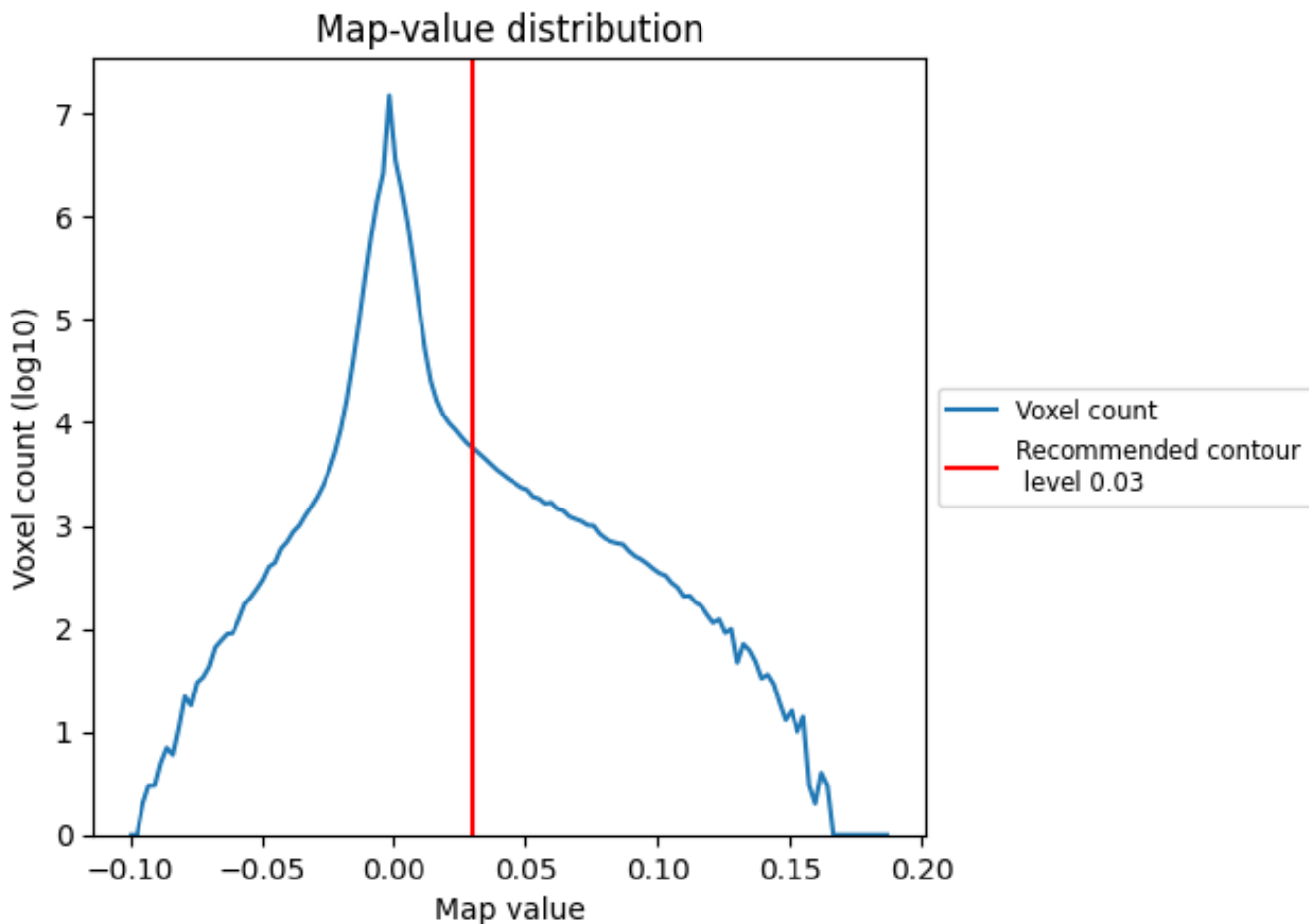
6.5 Mask visualisation [i](#)

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

7 Map analysis [i](#)

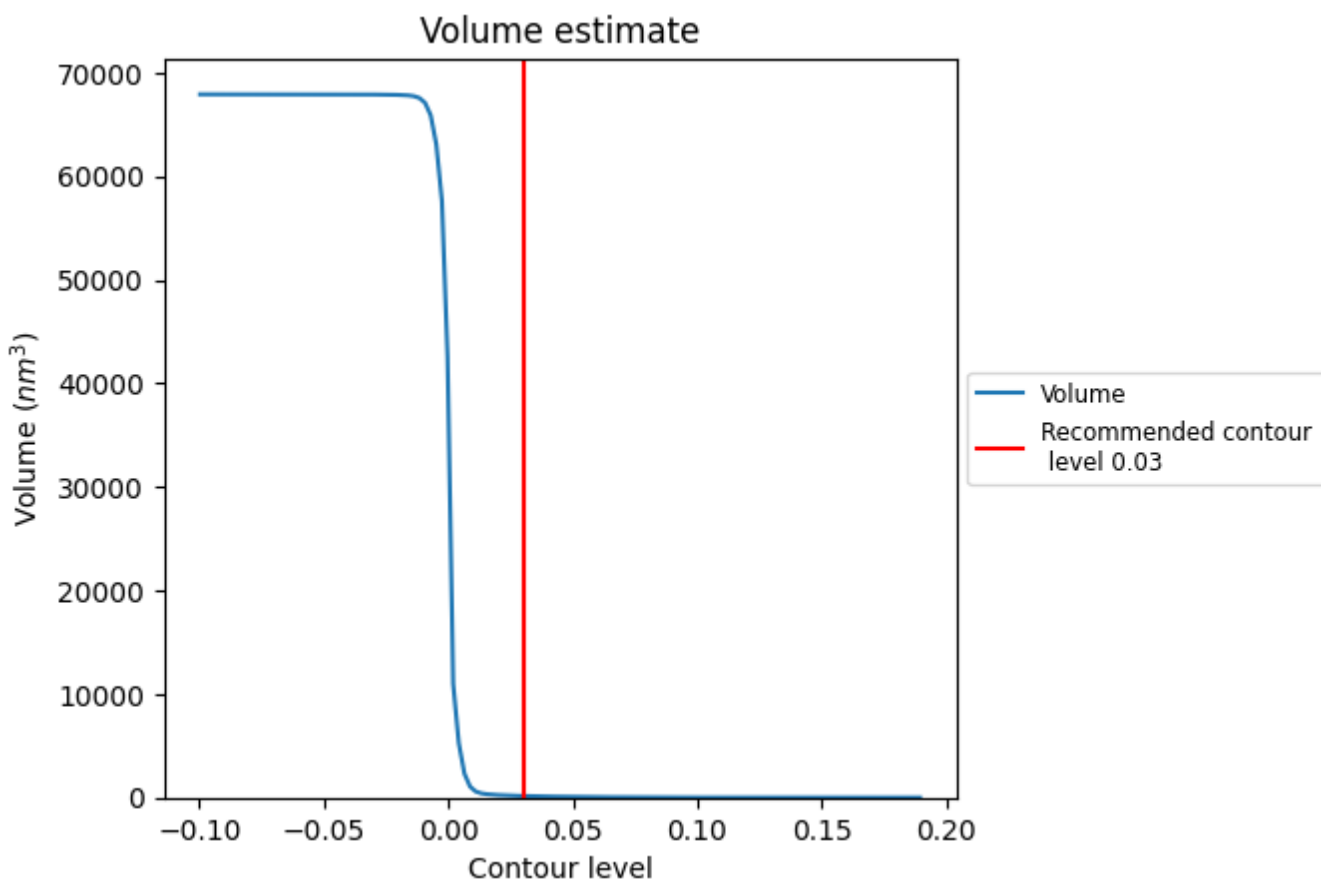
This section contains the results of statistical analysis of the map.

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.

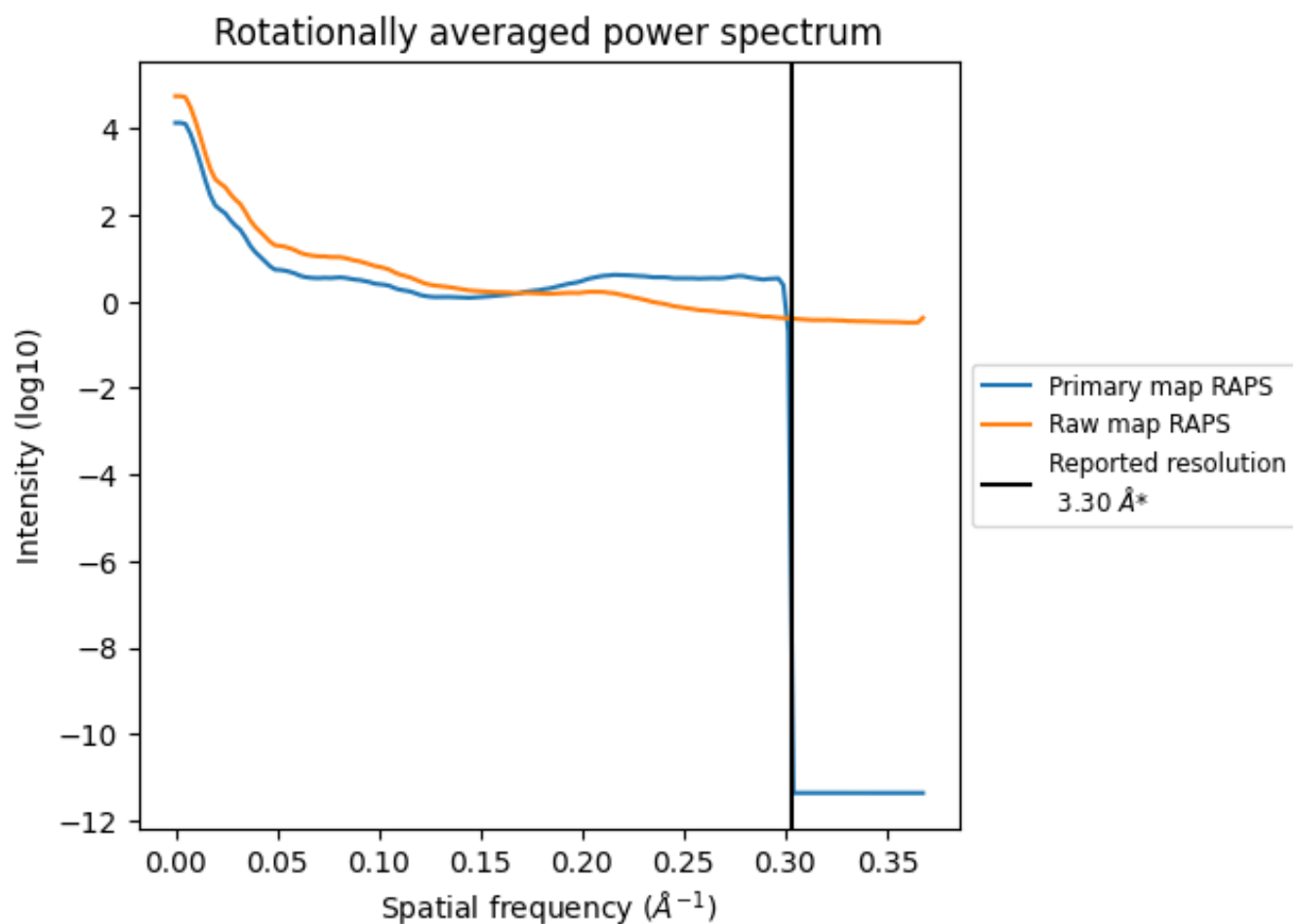
7.2 Volume estimate [i](#)



The volume at the recommended contour level is 151 nm³; this corresponds to an approximate mass of 137 kDa.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

7.3 Rotationally averaged power spectrum [i](#)

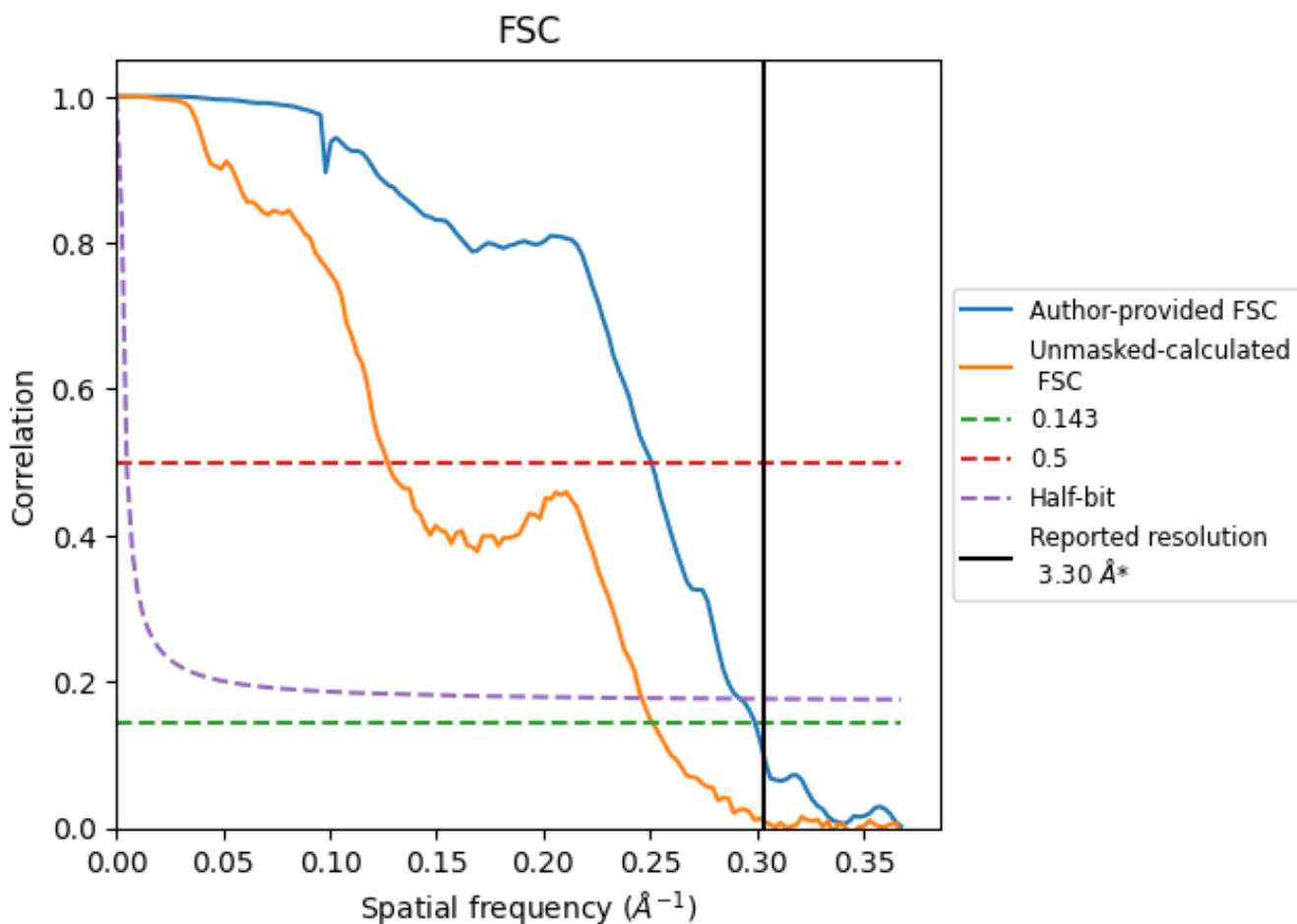


*Reported resolution corresponds to spatial frequency of 0.303 Å⁻¹

8 Fourier-Shell correlation [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 0.303 Å⁻¹

8.2 Resolution estimates [i](#)

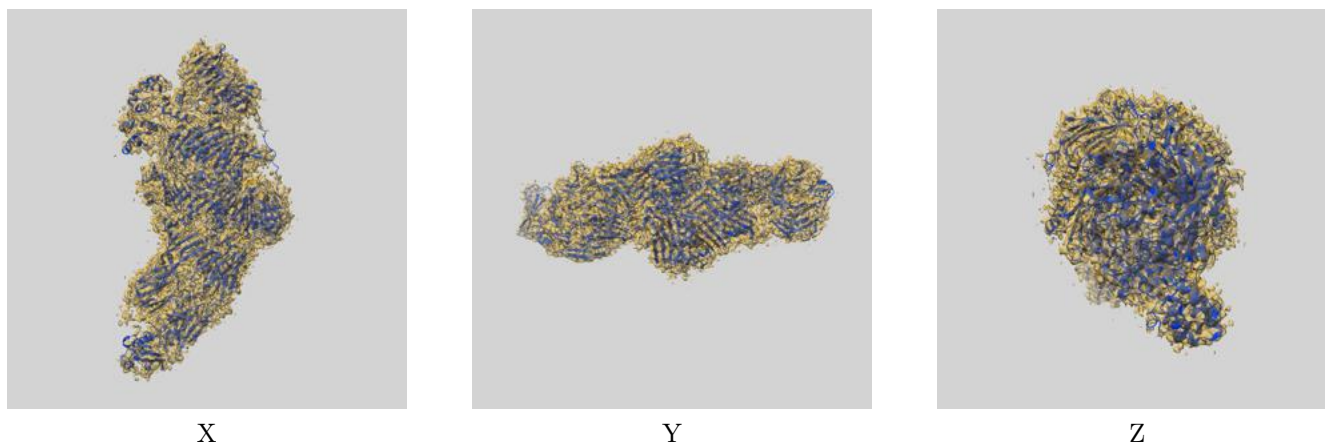
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.30	-	-
Author-provided FSC curve	3.34	3.99	3.42
Unmasked-calculated*	3.98	7.88	4.06

*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 3.98 differs from the reported value 3.3 by more than 10 %

9 Map-model fit [i](#)

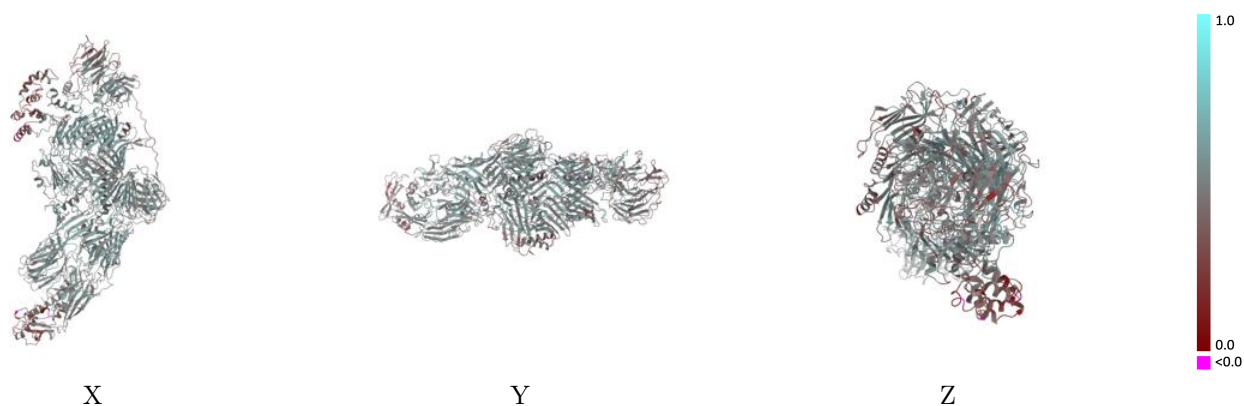
This section contains information regarding the fit between EMDB map EMD-29245 and PDB model 8FJP. Per-residue inclusion information can be found in section 3 on page 5.

9.1 Map-model overlay [i](#)



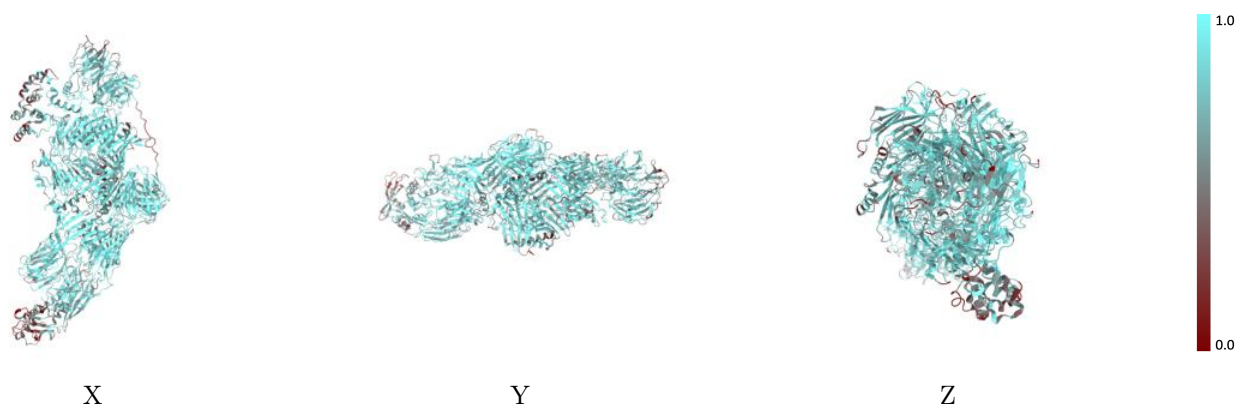
The images above show the 3D surface view of the map at the recommended contour level 0.03 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2 Q-score mapped to coordinate model [i](#)



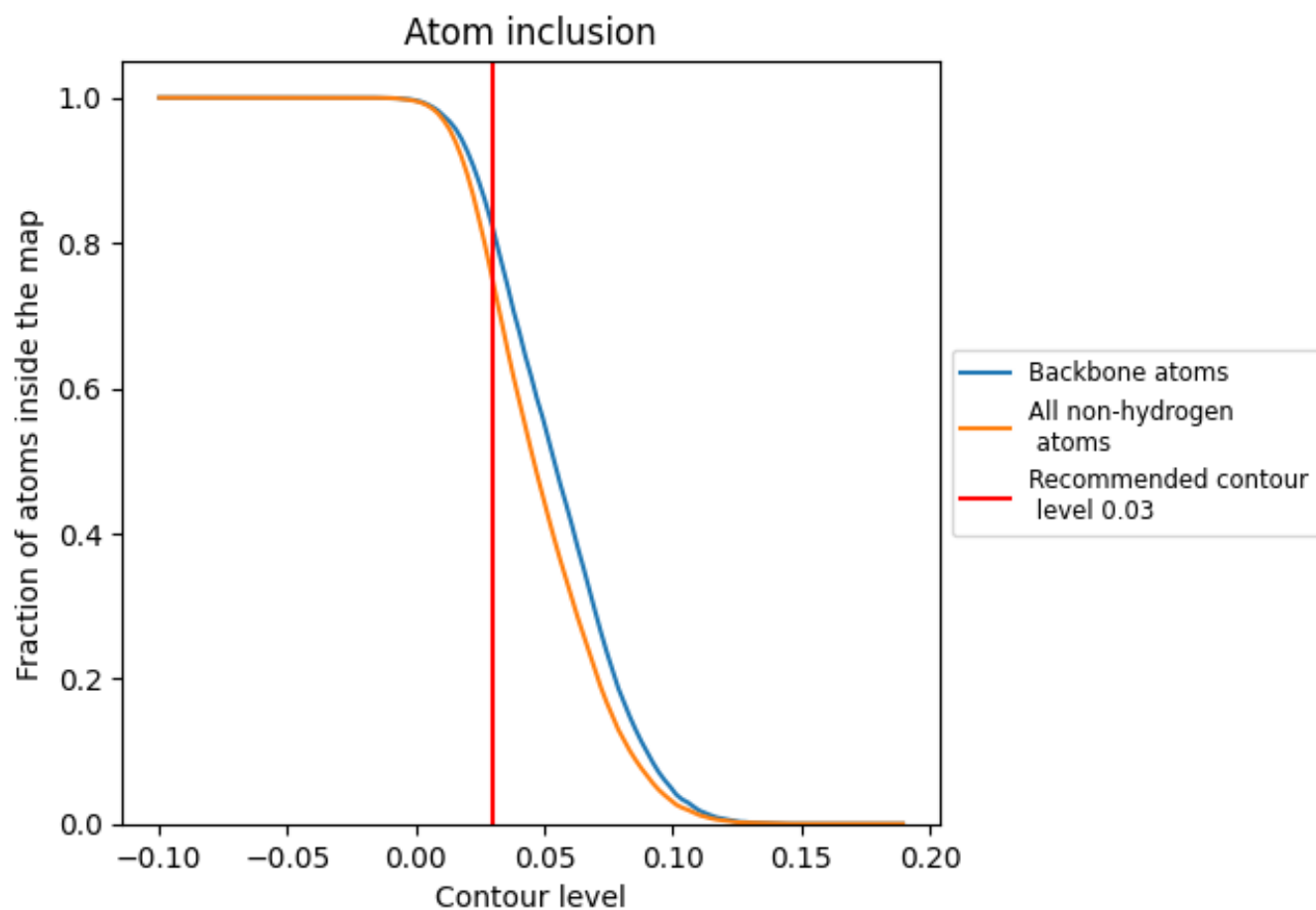
The images above show the model with each residue coloured according to its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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](#)



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 (0.03).







9.4 Atom inclusion [i](#)



At the recommended contour level, 82% of all backbone atoms, 75% 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 (0.03) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	 0.7477	 0.4870
A	 0.7478	 0.4870
B	 0.7143	 0.4780

