



## wwPDB EM Validation Summary Report ⓘ

Nov 20, 2022 – 03:38 AM EST

PDB ID : 3J7H  
EMDB ID : EMD-5995  
Title : Structure of beta-galactosidase at 3.2-A resolution obtained by cryo-electron microscopy  
Authors : Bartesaghi, A.; Matthies, D.; Banerjee, S.; Merk, A.; Subramaniam, S.  
Deposited on : 2014-06-30  
Resolution : 3.20 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](mailto: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  
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

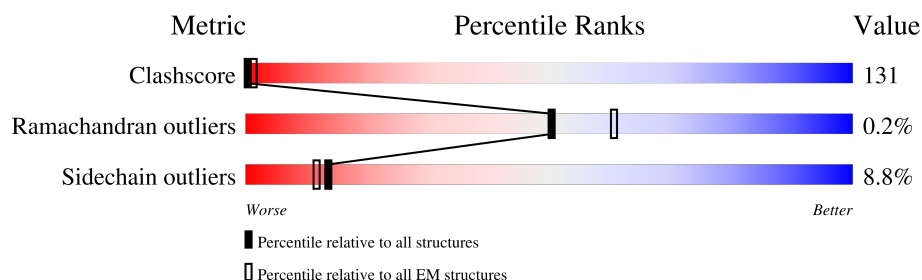
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*


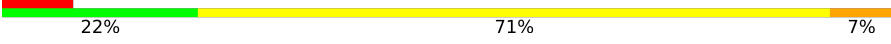

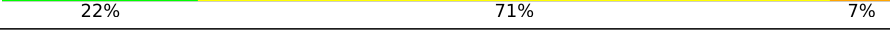
The reported resolution of this entry is 3.20 Å.

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  $\geq 3$ , 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	1024	
1	B	1024	
1	C	1024	
1	D	1024	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 32828 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 Beta-galactosidase.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1022	Total	C	N	O	S	0	0
			8206	5190	1452	1525	39		
1	B	1022	Total	C	N	O	S	0	0
			8206	5190	1452	1525	39		
1	C	1022	Total	C	N	O	S	0	0
			8206	5190	1452	1525	39		
1	D	1022	Total	C	N	O	S	0	0
			8206	5190	1452	1525	39		

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

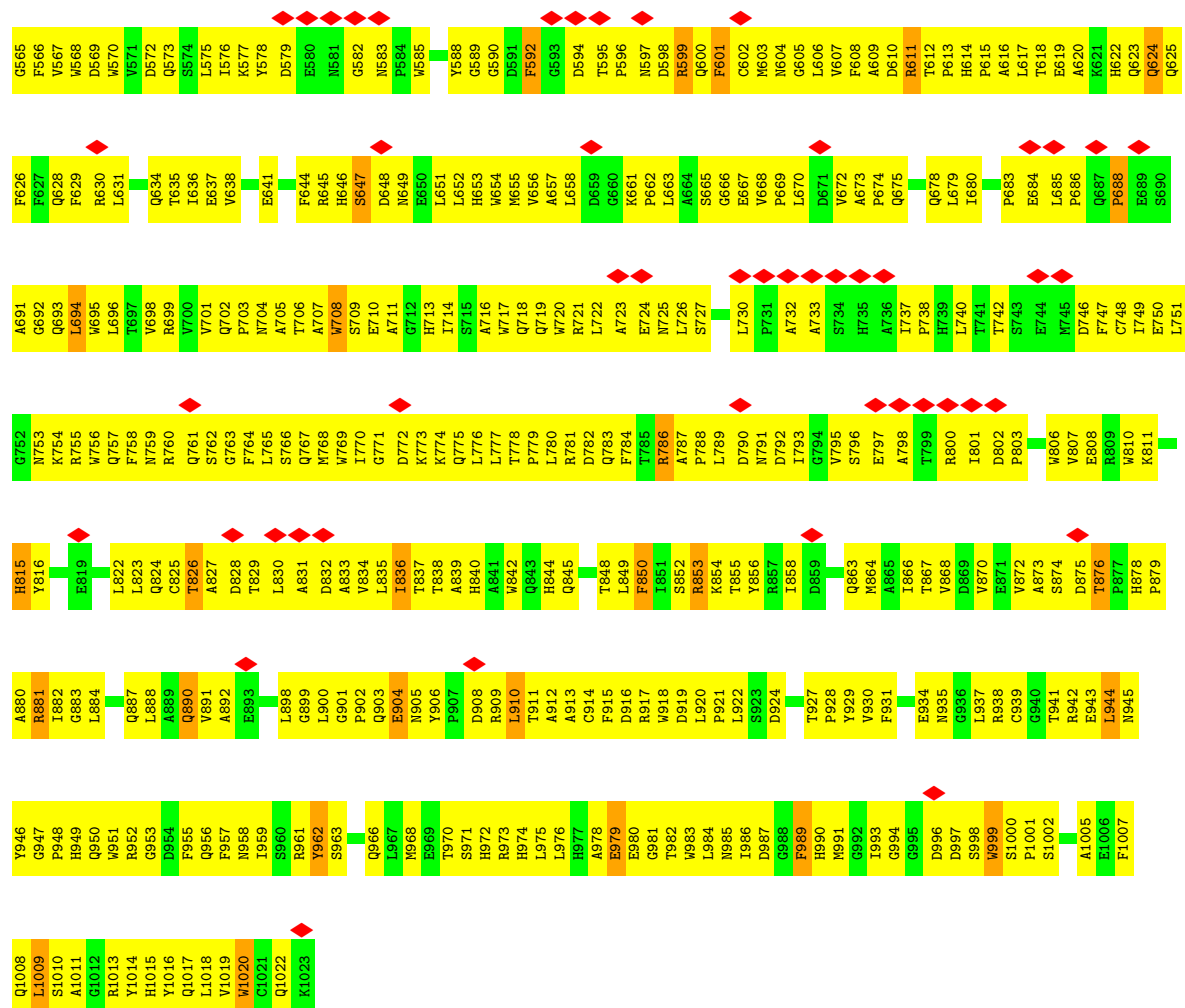
Mol	Chain	Residues	Atoms		AltConf
2	A	1	Total	Mg	0
			1	1	
2	B	1	Total	Mg	0
			1	1	
2	C	1	Total	Mg	0
			1	1	
2	D	1	Total	Mg	0
			1	1	

### 3 Residue-property plots

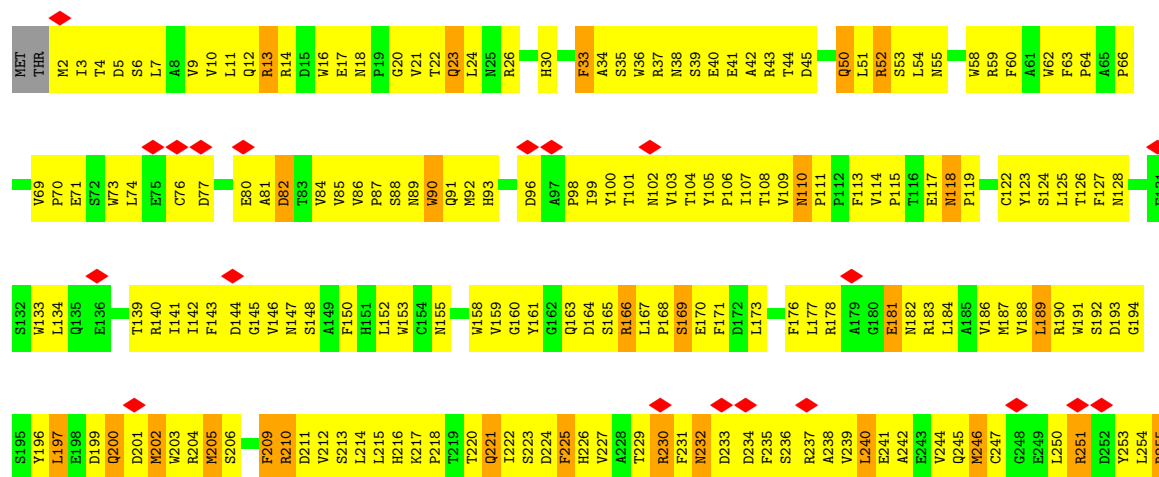
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion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





### • Molecule 1: Beta-galactosidase

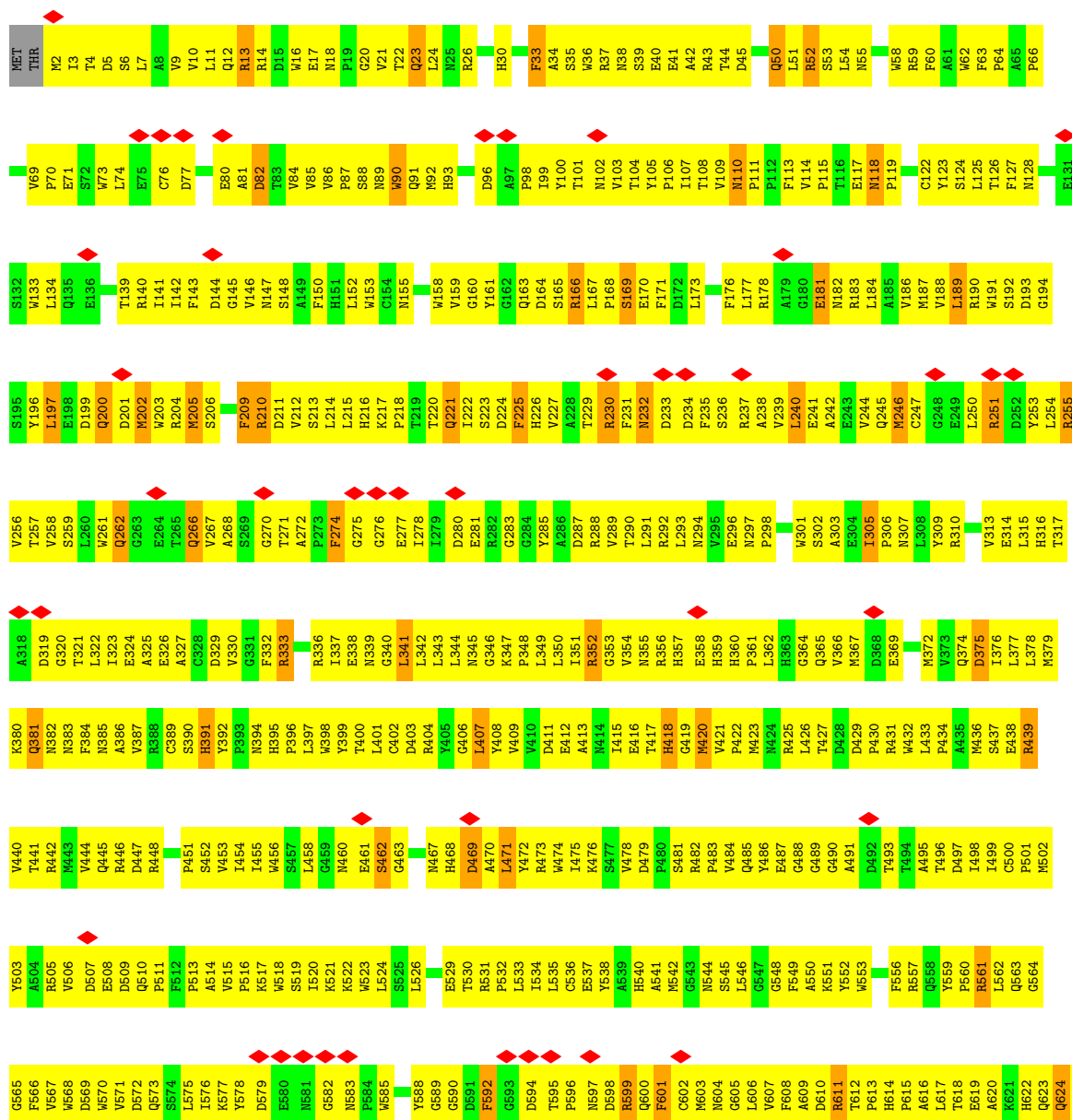




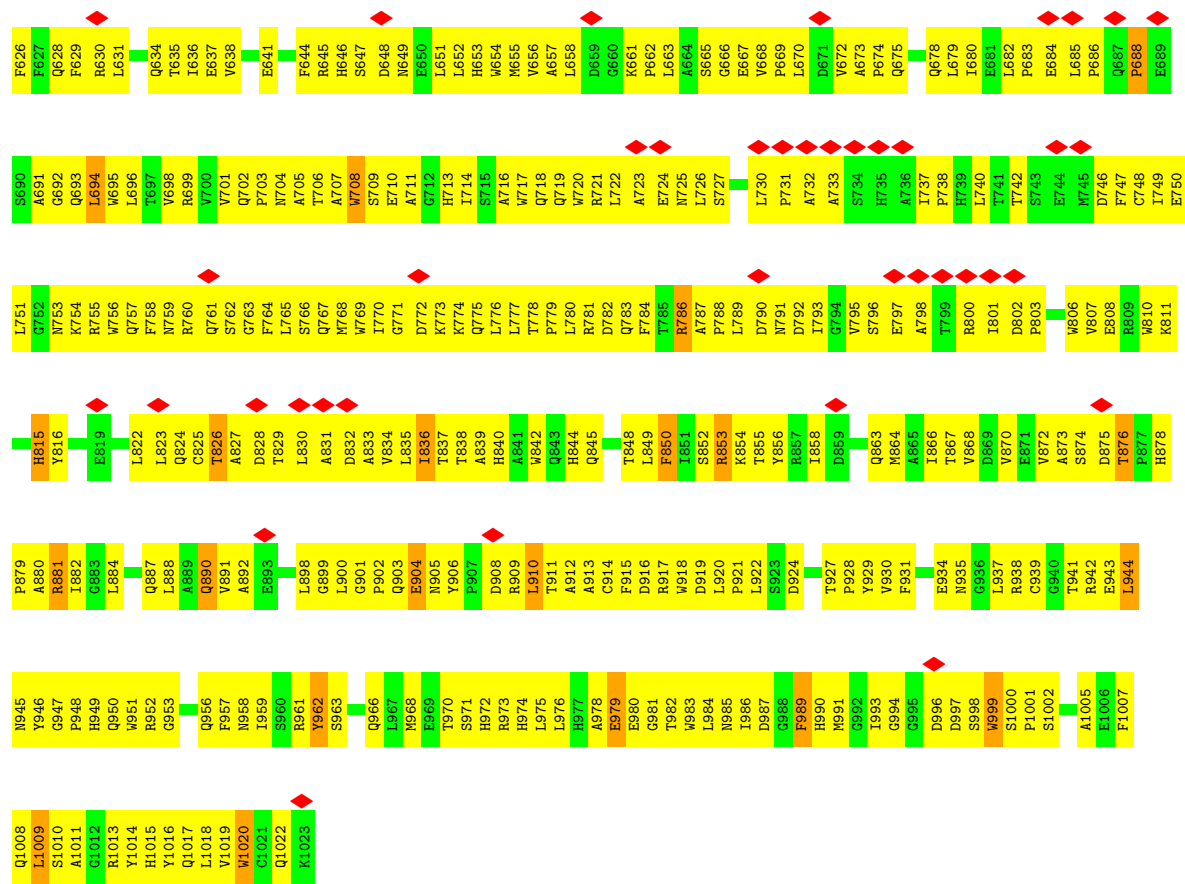
L751	S690	F626	G565	Y503	V440	K380	A318	V256	S195	S132	V69	MET
G752	A691	F627	F566	A504	T441	Q381	D319	T257	Y196	W133	P70	THR
K753	G692	Q628	N753	R505	R442	N382	G320	V258	L197	L134	E71	
R754	Q693	F629	W568	V506	R443	N383	S259	E198	E198	Q135	S72	
W755	L694	B630	D569	D507	V444	F384	L321	L260	D199	E136	L74	
W756	W695	L631	W570	E508	Q445	N385	L322	W261	Q200		W73	
L696	L696		V571	D509	Q446	A386	L323	Q262	D201		L74	
T697	T697		D572	Q510	D447	R387	E324	E263	M202		C76	
V698	V698		D573	P511	R448	V388	E325	E264	W203		D77	
W699	W699		Q573	P512		C389	A327	T265	R204			
G700	G700		S574	R512		S390	C328	Q266	M205			
V701	V701		L575	P513	P451	S390	C328	Q266	M205			
S637	S637		L576	P513	S452	H391	D329	A268	S206			
V638	V638		K577	P515	S453	Y392	V330	A268				
E641	E641		Y578	V515	T454	F393	G331	S269	F209			
			Y578	P516	T454	F393	G331	G146	R13			
			D579	P517	T455	H395	F332	N147	R14			
			W580	K517	T456	P396	R333	S148	D15			
			N581	S519	S457	L397		A149	W16			
			G582	S520	L458	P273	F274	F150	W17			
			K521	S520	G459	V398	P274	H151	N18			
			N583	K522	M460	Y399	G275	L152	P19			
			P584	W523	E461	T400	G276	W153	G20			
			W585	L524	S462	L401	E277	C154	W21			
				S525	G463	C402	L341	N155	W90			
				L526		D403	L342		Q91			
			Y588	E529	M467	R404	L343	T220	L24			
			G589	T530	H468	Y405	L344	Q221	L26			
			G590	W531	G406	G406	N345	E281	N25			
			D591	P532	D469	L407	G346	Q282	R22			
			F592	L533	A470	Y408	K347	G283	D224			
			G593	L534	L471	V409	P348	G284	F225			
			D594	L535	Y472	D411	L349	Y285	H226			
			T595	C536	R473	E412	L350	A286	H226			
			P596	E537	T475	E413	R351	D287	Y100			
			N597	D598	S477	L415	R352	R288	S165			
			R599	H540	V478	E416	V354	R289	R166			
			Q600	A541	D479	T417	N355	F231	L167			
			F601	M542	P480	Y418	R356	R292	P168			
			C602	G543	S481	G419	H357	L291	S169			
			N603	N544	R482	V421	E358	N294	E170			
			M604	S545	P483	Q485	H360	E296	F171			
			G605	L546	Q485	M423	P361	N297	D233			
			D671	G547	Y486	D428	L362	P298	D234			
			V607	G548	E487	M424	G363		F235			
			F608	F549	G488	R425	Q364	W301	S236			
			A609	A550	G489	L426	Q365	S302	R237			
			P674	D610	K551	T427	V366	A242	L177			
			Q675	R611	A491	D429	M367	E243	R178			
			T612	W553	D492	P430	D368	V244	P111			
			P613	F556	T493	R431	E369	Q245	A179			
			H614	R557	T494	L432	M372	M246	L240			
			P615	Q568	T495	L433	V373	C247	G190			
			A616	V559	T496	R309	L308	G248	E181			
			L617	P560	D497	Y310	Y307	E249	N182			
			T618	V560	T497	D435	D375	R189	R183			
			E684	R561	T498	M436	Q374	L250	E117			
			E688	A620	T499	S437	D376	R251	N118			
			L685	L562	C500	S437	L377	D252	P119			
			P686	G621	P501	E438	L378	Y253	C122			
			Q687	H622	M502	R439	L378	L254	G194			
			F688	Q623			K379	T317				
			E689	Q625								
					</							



● Molecule 1: Beta-galactosidase







## 4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, D2	Depositor
Number of particles used	11726	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Individual frames of each movie were aligned by cross-correlation using the cumulative average of previously aligned frames as a reference to align the remaining frames. Parameters of the contrast transfer function for each micrograph were estimated from power spectra obtained using periodogram averaging with tiles of size 512x512 pixels extracted from all frames of each movie. These power spectra were then radially averaged and used to estimate the defocus for each image using frequencies in the 15.0-3.0 Angstrom range. CTF correction was done for each particle as implemented in FREALIGN's reconstruction protocol.	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	45	Depositor
Minimum defocus (nm)	1000	Depositor
Maximum defocus (nm)	2500	Depositor
Magnification	105000	Depositor
Image detector	GATAN K2 (4k x 4k)	Depositor
Maximum map value	0.088	Depositor
Minimum map value	-0.051	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.006	Depositor
Recommended contour level	0.0224	Depositor
Map size (Å)	216.75, 216.75, 216.75	wwPDB
Map dimensions	340, 340, 340	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.6375, 0.6375, 0.6375	Depositor

## 5 Model quality

### 5.1 Standard geometry

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

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	0.26	0/8448	0.47	2/11526 (0.0%)
1	B	0.26	1/8448 (0.0%)	0.47	2/11526 (0.0%)
1	C	0.26	0/8448	0.47	2/11526 (0.0%)
1	D	0.26	0/8448	0.47	2/11526 (0.0%)
All	All	0.26	1/33792 (0.0%)	0.47	8/46104 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	306	PRO	N-CD	5.03	1.54	1.47

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	305	ILE	C-N-CD	5.72	140.41	128.40
1	C	305	ILE	C-N-CD	5.72	140.41	128.40
1	D	305	ILE	C-N-CD	5.71	140.39	128.40
1	B	305	ILE	C-N-CD	5.71	140.39	128.40
1	B	110	ASN	C-N-CD	5.11	139.14	128.40

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	8206	0	7791	2164	0
1	B	8206	0	7791	2165	0
1	C	8206	0	7791	2156	0
1	D	8206	0	7791	2161	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
All	All	32828	0	31164	8379	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 131.

The worst 5 of 8379 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:C:427:THR:HA	1:C:436:MET:CE	1.21	1.67
1:D:427:THR:HA	1:D:436:MET:CE	1.21	1.67
1:C:159:VAL:HG22	1:C:176:PHE:CE1	1.25	1.64
1:D:159:VAL:HG22	1:D:176:PHE:CE1	1.25	1.64
1:A:159:VAL:HG22	1:A:176:PHE:CE1	1.25	1.63

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	1020/1024 (100%)	993 (97%)	25 (2%)	2 (0%)	47 79
1	B	1020/1024 (100%)	993 (97%)	25 (2%)	2 (0%)	47 79
1	C	1020/1024 (100%)	993 (97%)	25 (2%)	2 (0%)	47 79
1	D	1020/1024 (100%)	993 (97%)	25 (2%)	2 (0%)	47 79

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	4080/4096 (100%)	3972 (97%)	100 (2%)	8 (0%)	50	79

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	511	PRO
1	B	511	PRO
1	C	511	PRO
1	D	511	PRO
1	A	688	PRO

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	874/876 (100%)	797 (91%)	77 (9%)	10	36
1	B	874/876 (100%)	797 (91%)	77 (9%)	10	36
1	C	874/876 (100%)	797 (91%)	77 (9%)	10	36
1	D	874/876 (100%)	797 (91%)	77 (9%)	10	36
All	All	3496/3504 (100%)	3188 (91%)	308 (9%)	13	36

5 of 308 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	13	ARG
1	D	667	GLU
1	D	118	ASN
1	D	274	PHE
1	D	910	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 160 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	905	ASN
1	D	646	HIS
1	C	1022	GLN
1	D	266	GLN
1	D	815	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](#)

There are no monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

### 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.

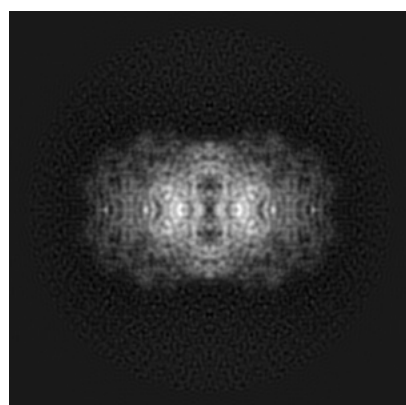
## 6 Map visualisation [i](#)

This section contains visualisations of the EMDB entry EMD-5995. These allow visual inspection of the internal detail of the map and identification of artifacts.

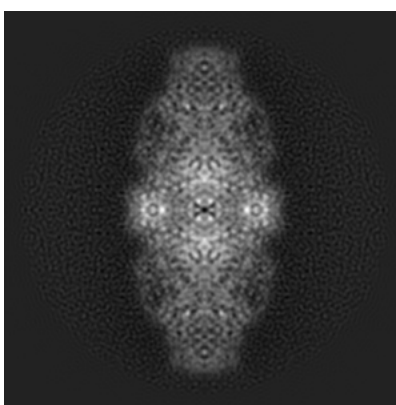
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

### 6.1 Orthogonal projections [i](#)

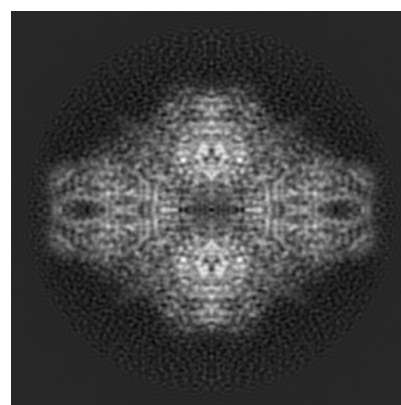
#### 6.1.1 Primary map



X



Y

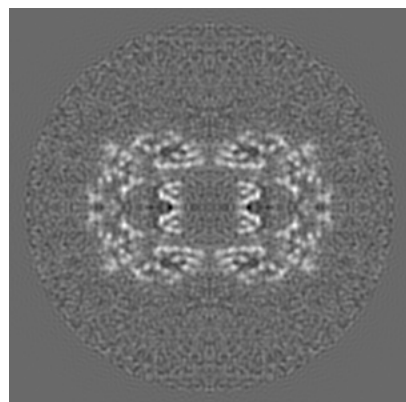


Z

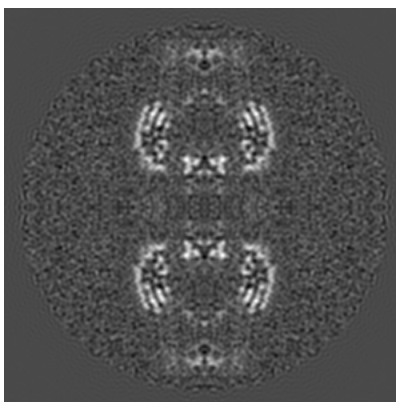
The images above show the map projected in three orthogonal directions.

### 6.2 Central slices [i](#)

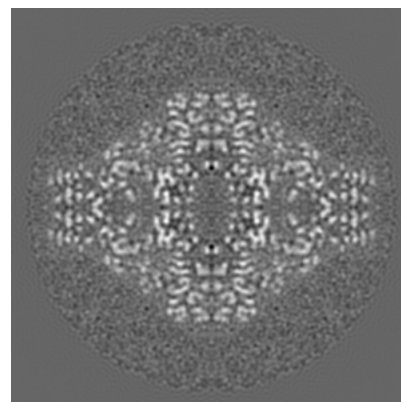
#### 6.2.1 Primary map



X Index: 170



Y Index: 170

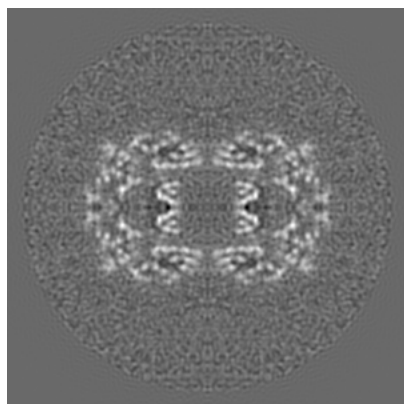


Z Index: 170

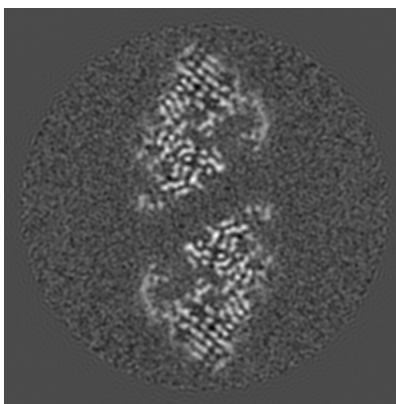
The images above show central slices of the map in three orthogonal directions.

## 6.3 Largest variance slices [i](#)

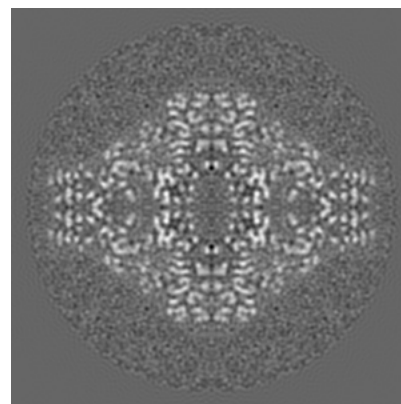
### 6.3.1 Primary map



X Index: 170



Y Index: 183

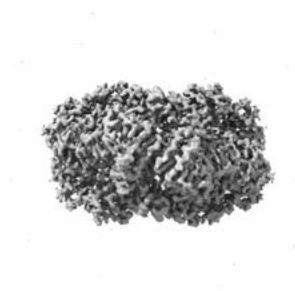


Z Index: 170

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

## 6.4 Orthogonal surface views [i](#)

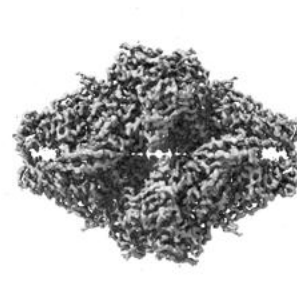
### 6.4.1 Primary map



X



Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.0224. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.



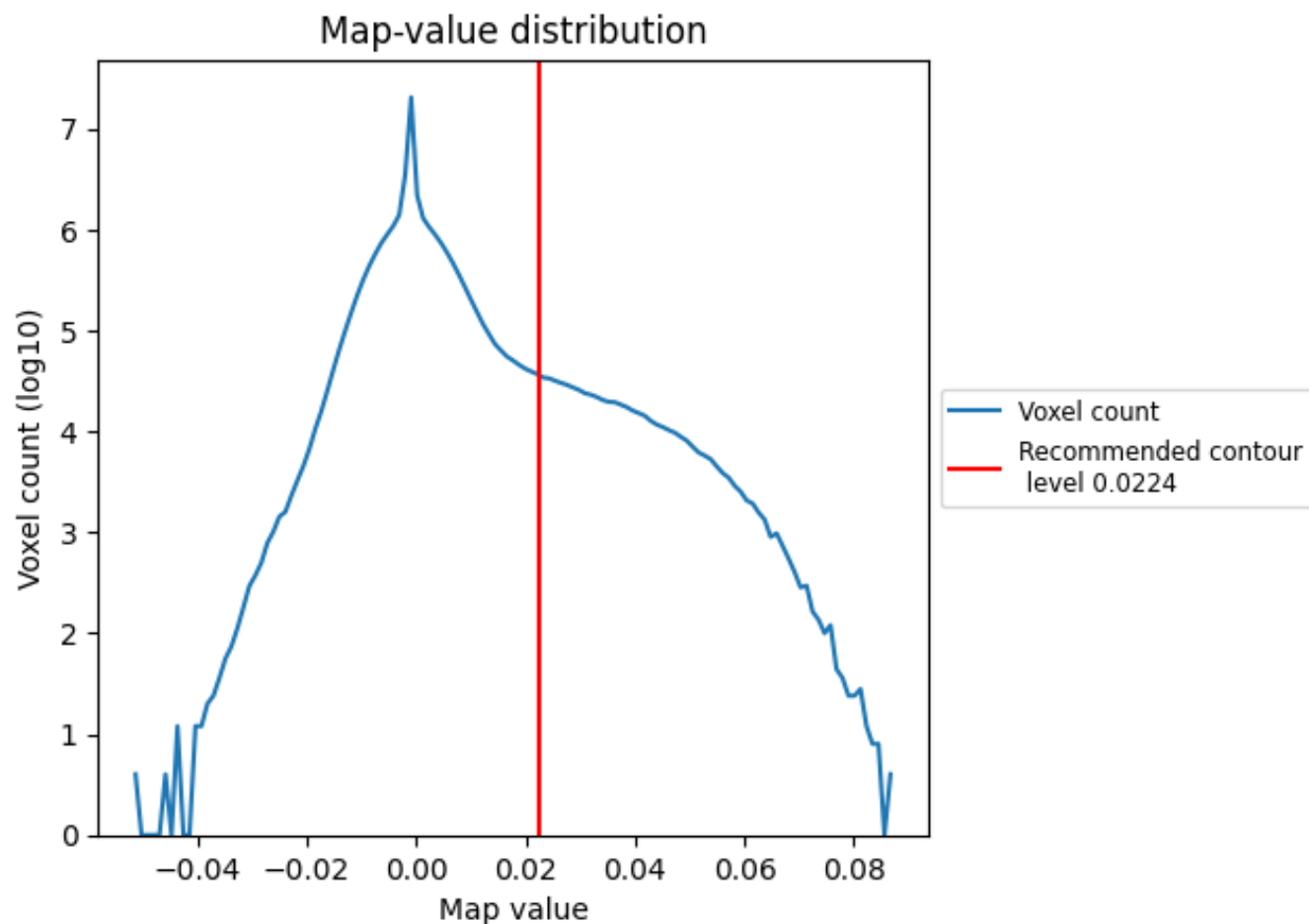
## 6.5 Mask visualisation

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

## 7 Map analysis [i](#)

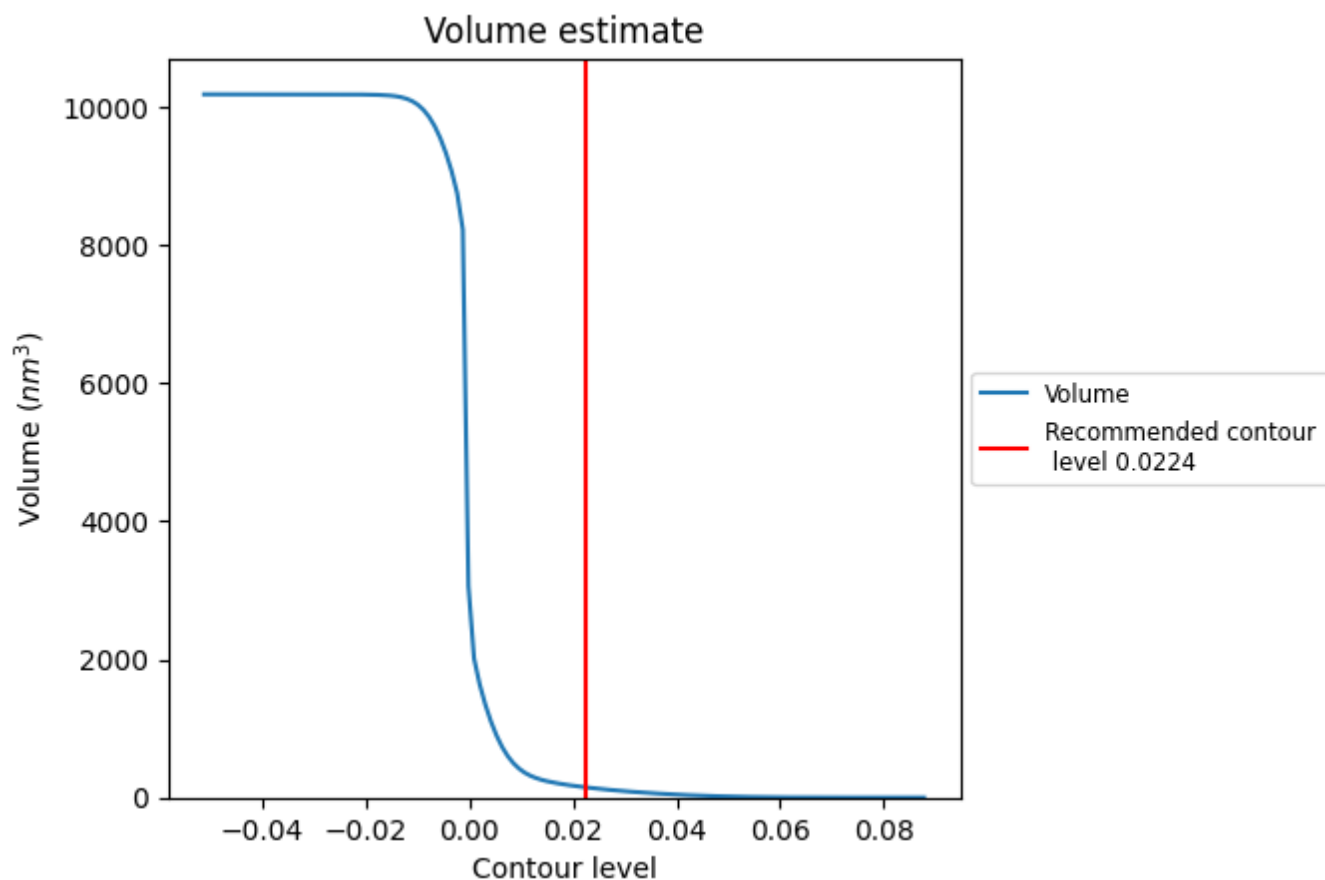
This section contains the results of statistical analysis of the map.

### 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.

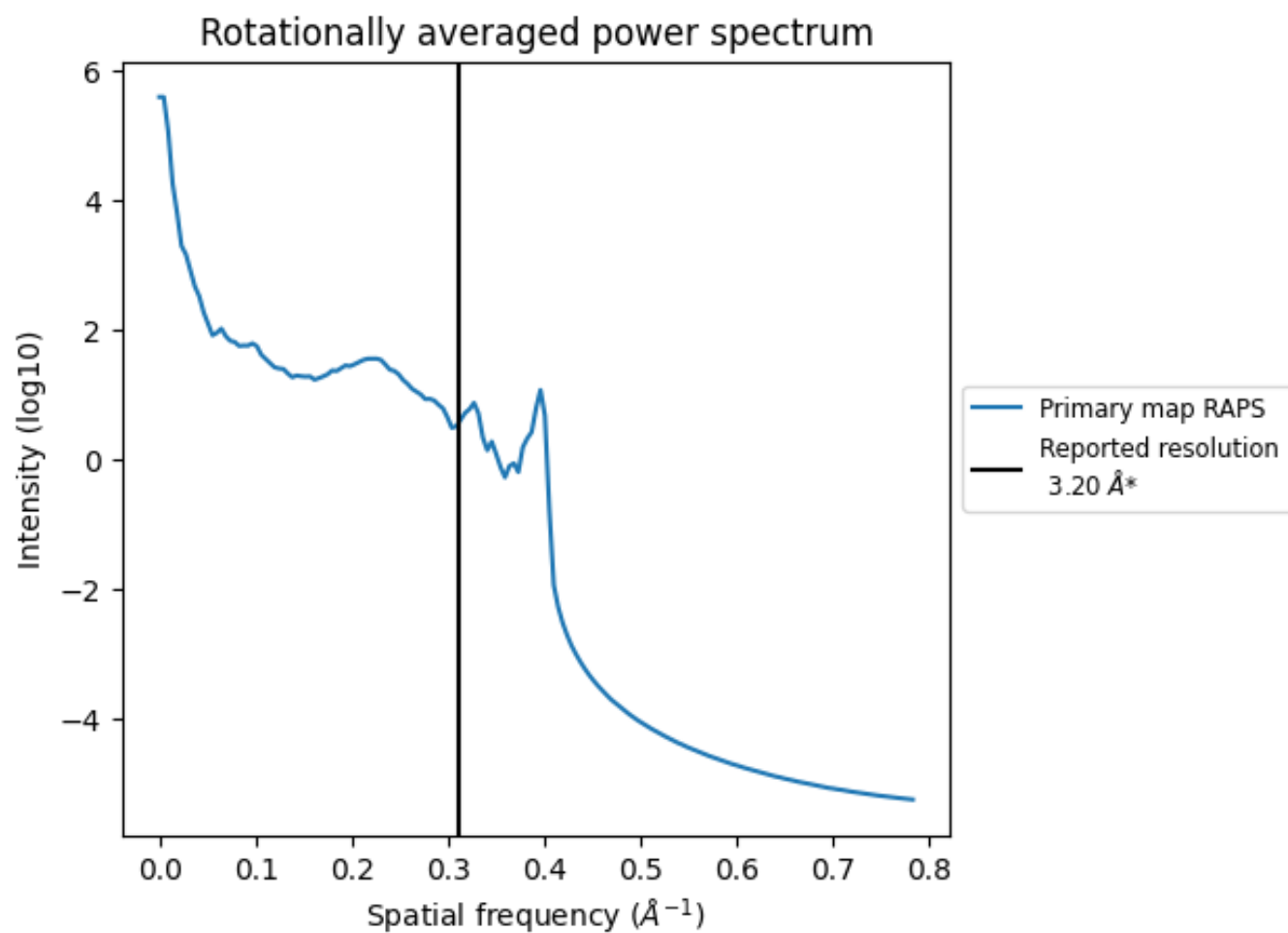
## 7.2 Volume estimate [i](#)



The volume at the recommended contour level is 148 nm<sup>3</sup>; this corresponds to an approximate mass of 134 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 ⓘ

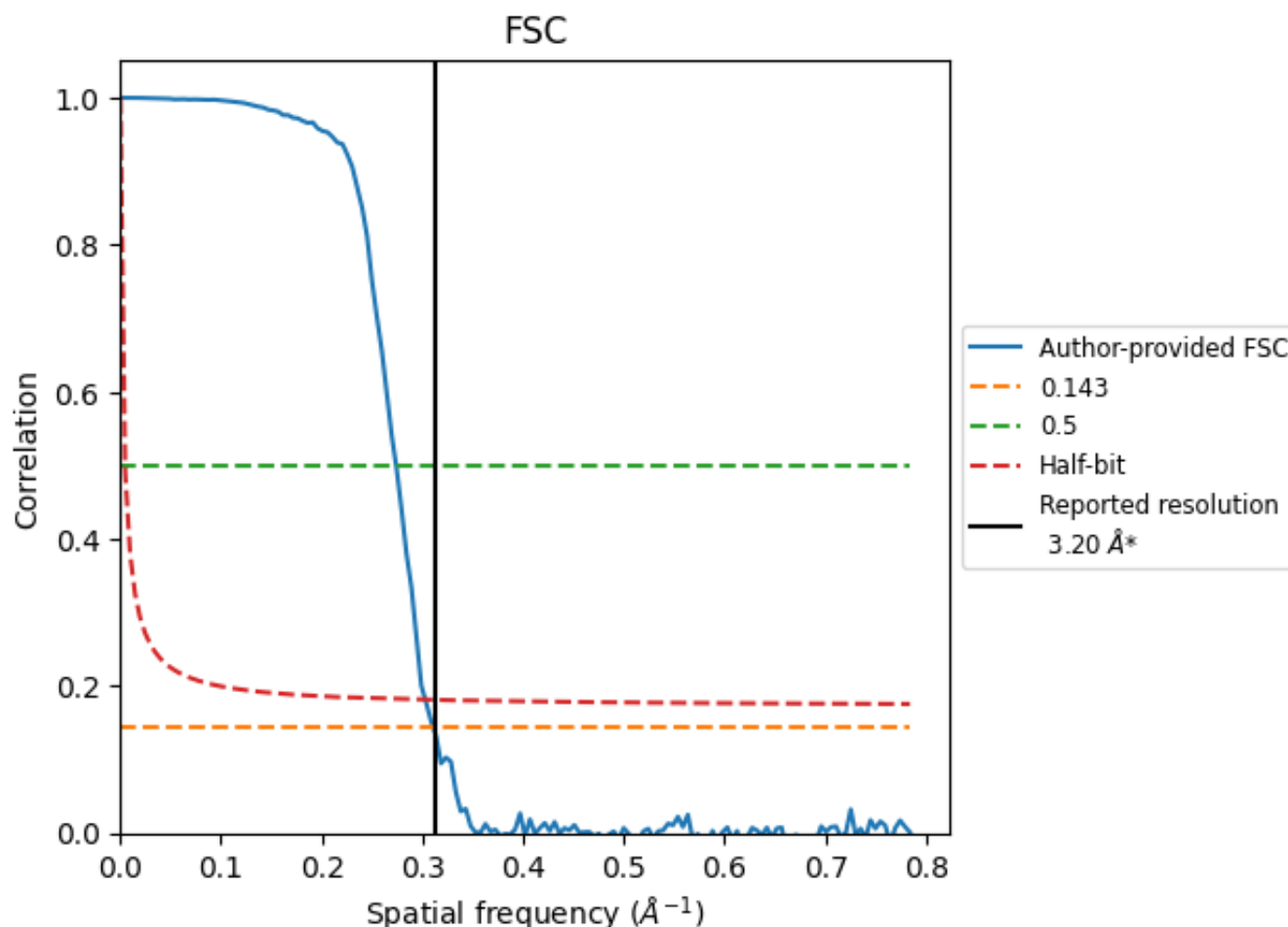


\*Reported resolution corresponds to spatial frequency of 0.312 Å<sup>-1</sup>

## 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 0.312 Å<sup>-1</sup>

## 8.2 Resolution estimates [i](#)

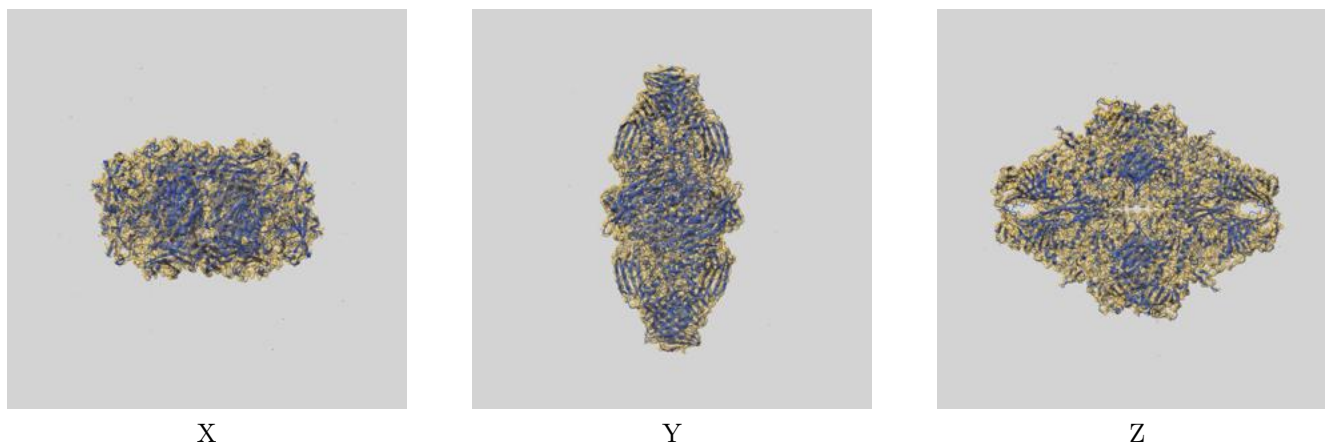
Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.20	-	-
Author-provided FSC curve	3.22	3.65	3.30
Unmasked-calculated*	-	-	-

\*Resolution estimate based on FSC curve calculated by comparison of deposited half-maps.

## 9 Map-model fit [i](#)

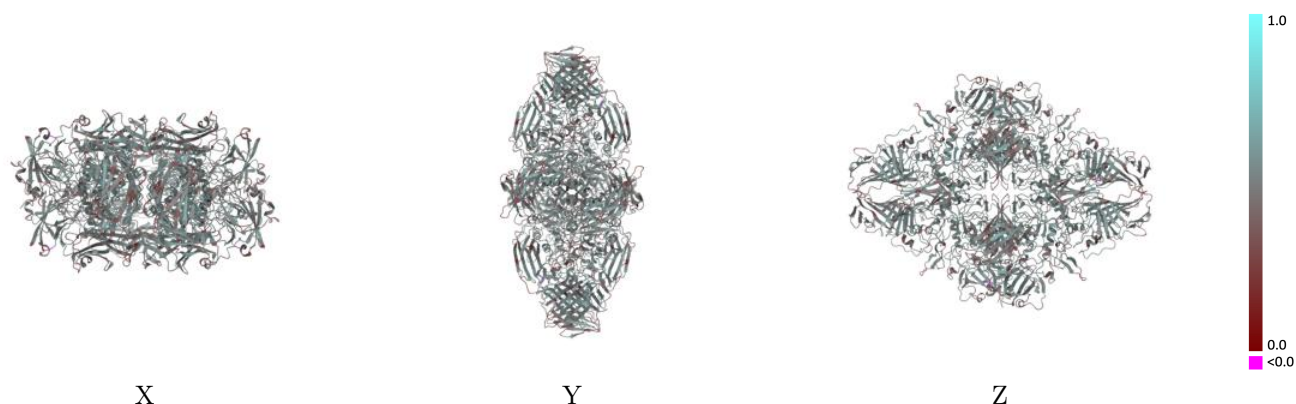
This section contains information regarding the fit between EMDB map EMD-5995 and PDB model 3J7H. Per-residue inclusion information can be found in [section 3](#) on [page 4](#).

### 9.1 Map-model overlay [i](#)



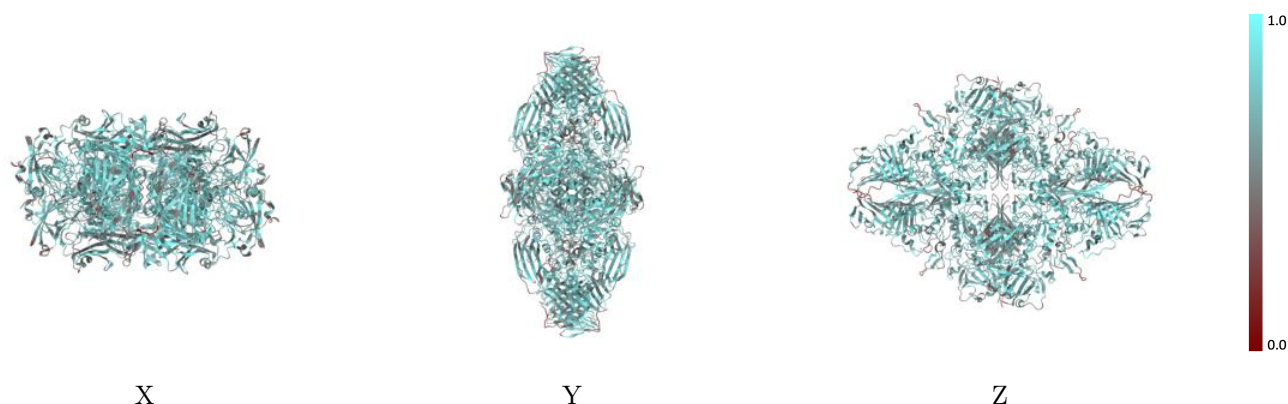
The images above show the 3D surface view of the map at the recommended contour level 0.0224 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

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



The images above show the model with each residue coloured according 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.

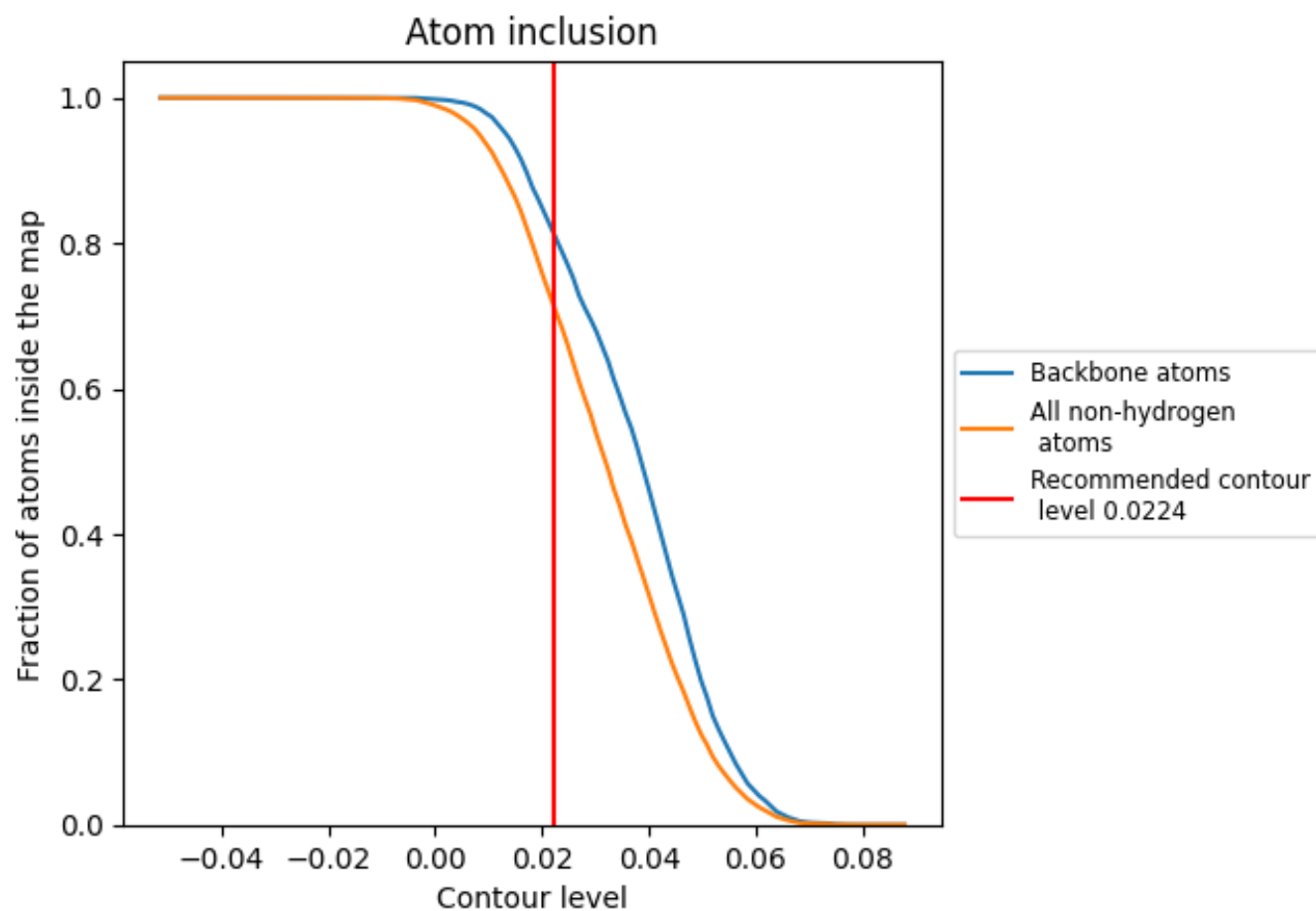
## 9.3 Atom inclusion mapped to coordinate model [i](#)



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 (0.0224).



## 9.4 Atom inclusion [i](#)



At the recommended contour level, 81% of all backbone atoms, 71% 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 (0.0224) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	<div></div> 0.7097	<div></div> 0.4920
A	<div></div> 0.7098	<div></div> 0.4930
B	<div></div> 0.7097	<div></div> 0.4920
C	<div></div> 0.7096	<div></div> 0.4920
D	<div></div> 0.7098	<div></div> 0.4920

