

# wwPDB EM Validation Summary Report (i)

Dec 10, 2022 – 10:32 am GMT

PDB ID : 5G06

EMDB ID : EMD-3366

Title : Cryo-EM structure of yeast cytoplasmic exosome

Authors: Liu, J.J.; Niu, C.Y.; Wu, Y.; Tan, D.; Wang, Y.; Ye, M.D.; Liu, Y.; Zhao,

W.W.; Zhou, K.; Liu, Q.S.; Dai, J.B.; Yang, X.R.; Dong, M.Q.; Huang, N.;

Wang, H.W.

Deposited on : 2016-03-17

Resolution : 4.20 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
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 (1)) 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)

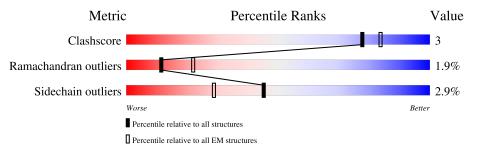
Validation Pipeline (wwPDB-VP) : 2.31.3

### 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 4.20 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 $(\#  ext{Entries})$	${ m EM~structures} \ (\#{ m Entries})$
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	305	7% 79%		1	.9%	
2	В	246	73%		21%	•	<del>-</del>
3	С	394	7%	17%	• •	15%	
4	D	223	81%			18%	-
5	Е	265	79%		1	.9%	•
6	F	250	61%	20%	••	16%	_
7	G	240	75%		20%	-	
8	Н	359		21%	•	18%	_

Continued on next page...



 $Continued\ from\ previous\ page...$ 

Mol	Chain	Length		Quality of chain						
9	I	292	11%	57%	18%		22%	_		
10	J	1001	6%	74%			22%			
11	Р	747	8% • •		87%			_		



## 2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 26794 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 EXOSOME COMPLEX COMPONENT RRP45.

Mol	Chain	Residues		At	oms			AltConf	Trace
1	A	302	Total 2350	C 1474	N 402	O 457	S 17	0	0

• Molecule 2 is a protein called EXOSOME COMPLEX COMPONENT SKI6.

Mol	Chain	Residues		Ato	oms			AltConf	Trace
2	В	240	Total 1890	C 1182	N 338	O 362	S 8	0	0

• Molecule 3 is a protein called EXOSOME COMPLEX COMPONENT RRP43.

M	lol	Chain	Residues		At	oms			AltConf	Trace
;	3	С	335	Total 2622	C 1653	N 455	O 503	S 11	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
С	C 363 MET		VAL	conflict	UNP P25359

• Molecule 4 is a protein called EXOSOME COMPLEX COMPONENT RRP46.

Mol	Chain	Residues		At	oms			AltConf	Trace
4	D	223	Total 1707	C 1073	N 289	O 335	S 10	0	0

• Molecule 5 is a protein called EXOSOME COMPLEX COMPONENT RRP42.

Mol	Chain	Residues		Ato	oms			AltConf	Trace
5	F	265	Total	С	N	О	S	0	0
]	יבו	200	2048	1307	336	400	5		U

There is a discrepancy between the modelled and reference sequences:



Chain	Residue	Modelled	Actual	Comment	Reference
${ m E}$	138	ILE	VAL	conflict	UNP Q12277

• Molecule 6 is a protein called EXOSOME COMPLEX COMPONENT MTR3.

Me	ol	Chain	Residues		At	oms			AltConf	Trace
6		F	211	Total 1627	C 1018	N 276	O 323	S 10	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	75	SER	THR	conflict	UNP P48240

• Molecule 7 is a protein called EXOSOME COMPLEX COMPONENT RRP40.

Mol	Chain	Residues	Atoms				AltConf	Trace	
7	G	236	Total 1831	C 1169	N 301	O 350	S 11	0	0

• Molecule 8 is a protein called EXOSOME COMPLEX COMPONENT RRP4.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	Н	293	Total 2277	C 1419	N 411	O 435	S 12	0	0

• Molecule 9 is a protein called EXOSOME COMPLEX COMPONENT CSL4.

Mol	Chain	Residues	Atoms				AltConf	Trace	
9	I	227	Total 1750	C 1093	N 310	O 338	S 9	0	0

• Molecule 10 is a protein called EXOSOME COMPLEX EXONUCLEASE DIS3.

Mol	Chain	Residues	Atoms				AltConf	Trace	
10	J	993	Total 7942	C 5002	N 1395	O 1509	S 36	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	171	ASN	ASP	conflict	UNP Q08162
J	551	ASN	ASP	conflict	UNP Q08162



 $\bullet\,$  Molecule 11 is a protein called SUPERKILLER PROTEIN 7.

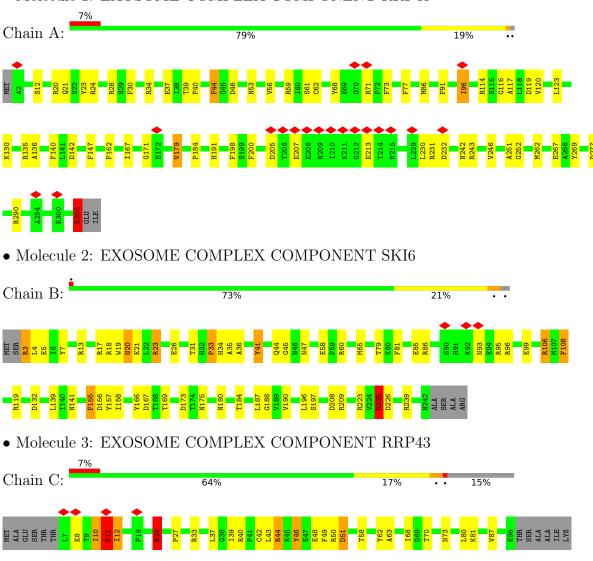
Mol	Chain	Residues	Atoms				AltConf	Trace	
11	Р	95	Total 750	C 475	N 128	O 146	S 1	0	0



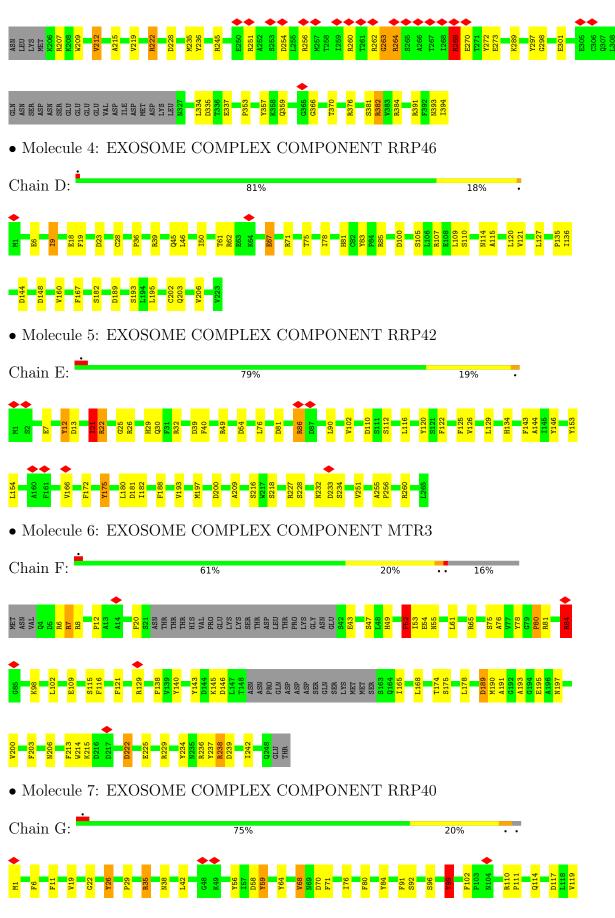
### 3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and atom inclusion in map density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red diamond above a residue indicates a poor fit to the EM map for this residue (all-atom inclusion < 40%). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

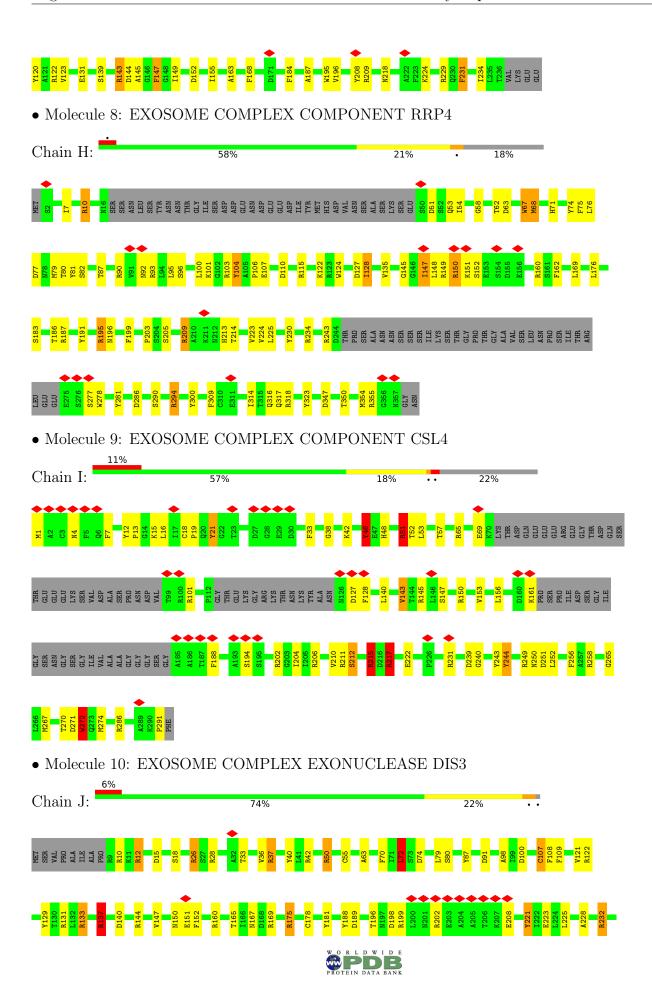
• Molecule 1: EXOSOME COMPLEX COMPONENT RRP45

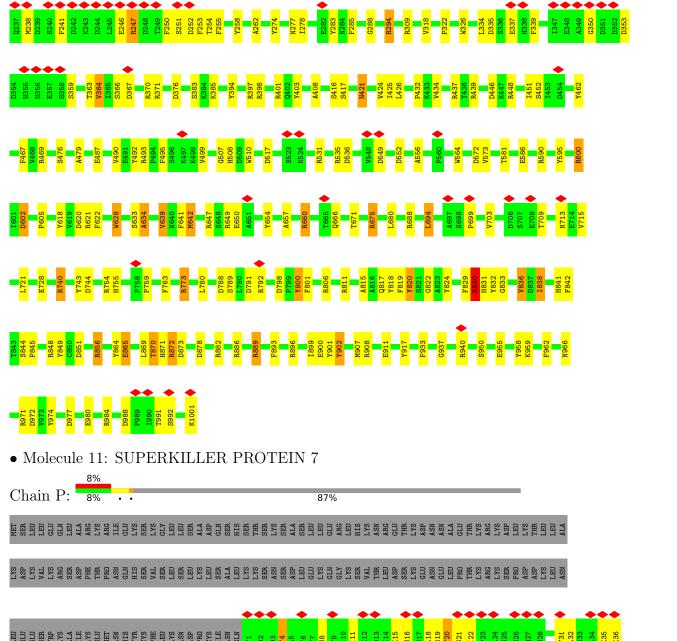
















| 1116 | SER | LEU | TTR | TTR | LEU | LEU | SER | LEU



# 4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	25000	Depositor
Resolution determination method	Not provided	
CTF correction method	MICROGRAPHS	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose $(e^-/\text{Å}^2)$	21	Depositor
Minimum defocus (nm)	1500	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	22500	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.112	Depositor
Minimum map value	-0.050	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.006	Depositor
Recommended contour level	0.02	Depositor
Map size (Å)	235.1772, 235.1772, 235.1772	wwPDB
Map dimensions	180, 180, 180	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.30654, 1.30654, 1.30654	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	В	ond lengths	Е	Bond angles
IVIOI	Chain	RMSZ	# Z  > 5	RMSZ	# Z >5
1	A	1.70	14/2386~(0.6%)	1.91	46/3218 (1.4%)
2	В	1.75	25/1914 (1.3%)	1.91	33/2577 (1.3%)
3	С	1.69	23/2659~(0.9%)	2.01	69/3596 (1.9%)
4	D	1.67	9/1725 (0.5%)	1.86	27/2339 (1.2%)
5	Е	1.68	11/2087~(0.5%)	1.90	42/2836 (1.5%)
6	F	1.79	16/1649 (1.0%)	2.04	$48/2222 \ (2.2\%)$
7	G	1.76	17/1868 (0.9%)	1.99	53/2531 (2.1%)
8	Н	1.80	30/2311 (1.3%)	2.02	61/3118 (2.0%)
9	I	1.78	26/1774 (1.5%)	1.98	42/2398 (1.8%)
10	J	1.75	90/8097 (1.1%)	1.98	$214/10972 \ (2.0\%)$
11	Р	1.69	6/763~(0.8%)	1.87	15/1031 (1.5%)
All	All	1.74	267/27233 (1.0%)	1.96	650/36838 (1.8%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a maintain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	3
2	В	0	7
3	С	0	10
4	D	0	2
5	Е	0	4
6	F	0	5
7	G	0	7
8	Н	0	5
9	I	0	6
10	J	0	26
11	Р	0	1
All	All	0	76

The worst 5 of 267 bond length outliers are listed below:



Mol	Chain	Res	Type	Atoms	Z	$\operatorname{Observed}(\text{\AA})$	$Ideal(\AA)$
7	G	110	ARG	CZ-NH2	9.83	1.45	1.33
10	J	886	ARG	CZ-NH2	9.40	1.45	1.33
8	Н	74	TYR	CZ-OH	8.53	1.52	1.37
10	J	590	ARG	NE-CZ	8.02	1.43	1.33
1	A	34	ARG	CZ-NH2	7.95	1.43	1.33

The worst 5 of 650 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$Observed(^o)$	$\mathrm{Ideal}(^{o})$
2	В	7	TYR	CB-CG-CD2	20.64	133.39	121.00
8	Н	281	TYR	CB-CG-CD2	-16.23	111.26	121.00
3	С	44	ARG	NE-CZ-NH2	-15.84	112.38	120.30
7	G	110	ARG	NE-CZ-NH1	15.73	128.16	120.30
3	С	262	ARG	NE-CZ-NH1	15.64	128.12	120.30

There are no chirality outliers.

5 of 76 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	269	TYR	Sidechain
1	A	303	ARG	Sidechain
1	A	44	PHE	Sidechain
2	В	23	ARG	Sidechain
2	В	3	ARG	Sidechain

#### 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	2350	0	2345	4	0
2	В	1890	0	1943	5	0
3	С	2622	0	2688	25	0
4	D	1707	0	1766	4	0
5	Е	2048	0	2079	4	0
6	F	1627	0	1607	50	0
7	G	1831	0	1829	2	0
8	Н	2277	0	2283	9	0
9	I	1750	0	1755	45	0

Continued on next page...



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
10	J	7942	0	7947	12	0
11	Р	750	0	759	112	0
All	All	26794	0	27001	157	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 3.

The worst 5 of 157 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} \ ({\rm \AA}) \end{array}$	$egin{aligned}  ext{Clash} \  ext{overlap } ( ext{Å}) \end{aligned}$
6:F:49:HIS:NE2	11:P:36:LYS:HD2	1.28	1.43
3:C:51:ASP:CB	11:P:47:PRO:HD3	1.49	1.41
9:I:250:ASN:OD1	11:P:86:LEU:CD1	1.71	1.38
3:C:81:LYS:HE2	11:P:35:LYS:CB	1.45	1.38
6:F:52:PHE:CD2	11:P:32:LEU:HD22	1.61	1.34

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	300/305~(98%)	278 (93%)	18 (6%)	4 (1%)	12 48
2	В	238/246~(97%)	223 (94%)	11 (5%)	4 (2%)	9 44
3	C	327/394~(83%)	294 (90%)	22 (7%)	11 (3%)	3 30
4	D	$221/223\ (99\%)$	211 (96%)	9 (4%)	1 (0%)	29 68
5	E	263/265~(99%)	248 (94%)	10 (4%)	5 (2%)	8 41
6	F	$205/250\ (82\%)$	194 (95%)	8 (4%)	3 (2%)	10 46
7	G	234/240 (98%)	219 (94%)	14 (6%)	1 (0%)	34 72
8	Н	$287/359\ (80\%)$	263 (92%)	15 (5%)	9 (3%)	4 31

Continued on next page...



Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Perce	entiles
9	I	219/292 (75%)	200 (91%)	14 (6%)	5 (2%)	6	37
10	J	991/1001 (99%)	905 (91%)	68 (7%)	18 (2%)	8	42
11	Р	93/747 (12%)	82 (88%)	9 (10%)	2 (2%)	6	38
All	All	3378/4322 (78%)	3117 (92%)	198 (6%)	63 (2%)	11	41

5 of 63 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	251	ALA
3	С	10	ILE
3	С	11	GLU
3	С	12	ILE
3	С	143	ALA

#### 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	263/266~(99%)	259 (98%)	4 (2%)	65 80
2	В	$214/218\ (98\%)$	207 (97%)	7 (3%)	38 61
3	С	294/349~(84%)	286 (97%)	8 (3%)	44 66
4	D	$197/197\ (100\%)$	193 (98%)	4 (2%)	55 73
5	E	240/240 (100%)	233 (97%)	7 (3%)	42 64
6	F	181/219~(83%)	174 (96%)	7 (4%)	32 57
7	G	$205/209\ (98\%)$	197 (96%)	8 (4%)	32 57
8	Н	$252/311\ (81\%)$	245 (97%)	7 (3%)	43 65
9	Ι	190/240~(79%)	184 (97%)	6 (3%)	39 62
10	J	895/901~(99%)	865 (97%)	30 (3%)	37 61
11	Р	89/702 (13%)	88 (99%)	1 (1%)	73 84
All	All	3020/3852~(78%)	2931 (97%)	89 (3%)	45 64



5 of 89 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
9	I	272	TRP
10	J	552	ASP
10	J	74	ASP
10	J	251	SER
10	J	671	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
10	J	871	HIS
10	J	888	HIS
11	Р	16	ASN
6	F	171	HIS
6	F	99	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.



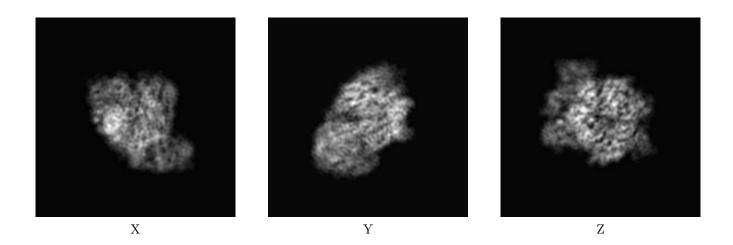
## 6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-3366. 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)

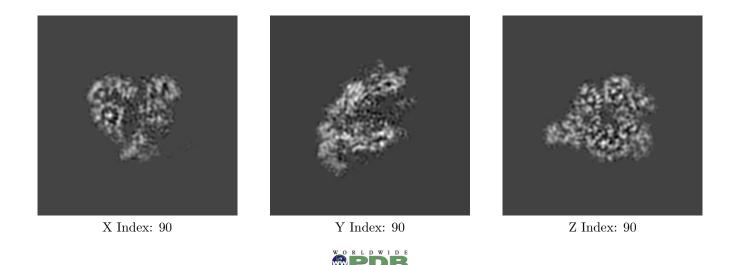
#### 6.1.1 Primary map



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

#### 6.2 Central slices (i)

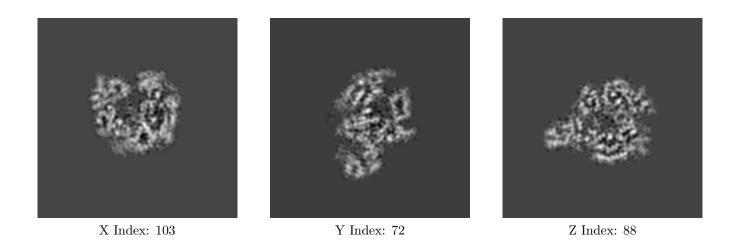
#### 6.2.1 Primary map



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

#### 6.3 Largest variance slices (i)

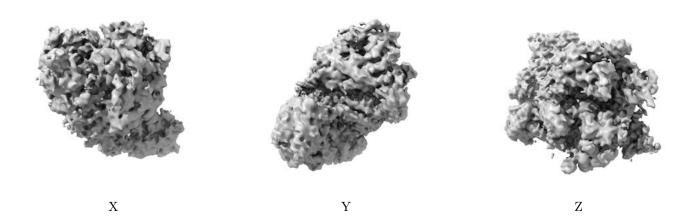
#### 6.3.1 Primary map



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 0.02. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.



## 6.5 Mask visualisation (i)

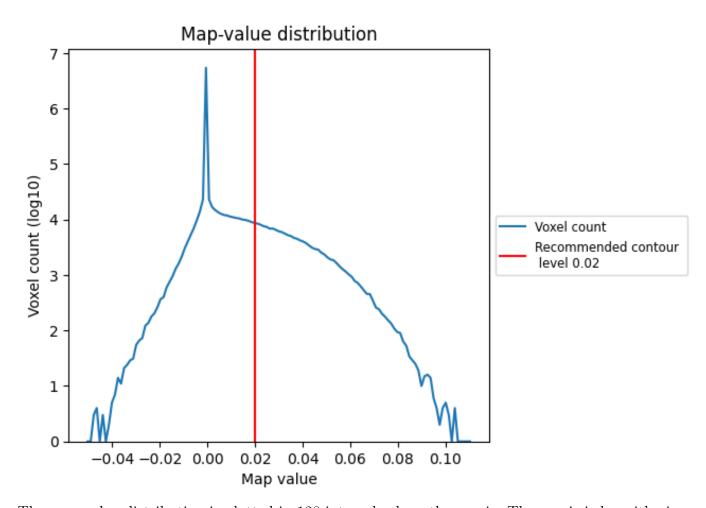
This section was not generated. No masks/segmentation were deposited.



## 7 Map analysis (i)

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

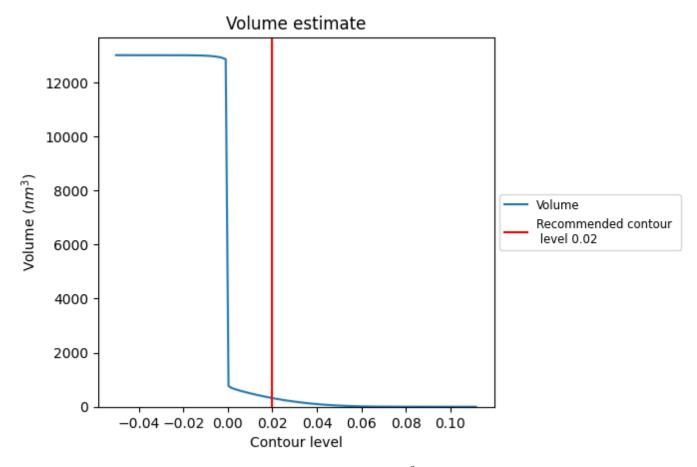
#### 7.1 Map-value distribution (i)



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)

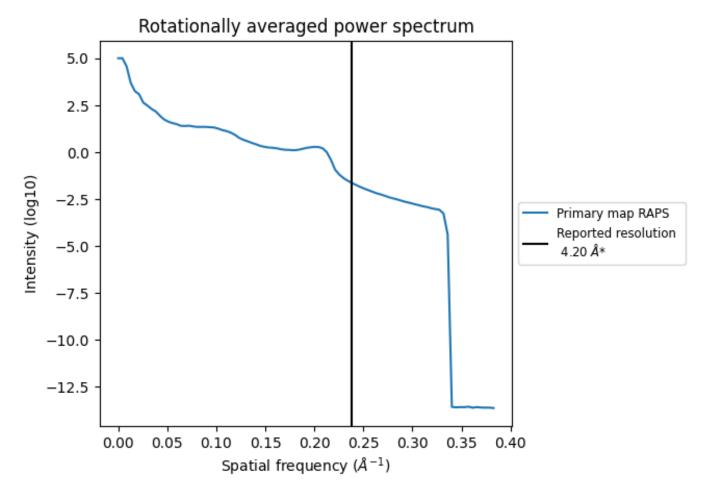


The volume at the recommended contour level is  $321~\mathrm{nm}^3$ ; this corresponds to an approximate mass of  $290~\mathrm{kDa}$ .

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)



<sup>\*</sup>Reported resolution corresponds to spatial frequency of 0.238  $\rm \mathring{A}^{-1}$ 



# 8 Fourier-Shell correlation (i)

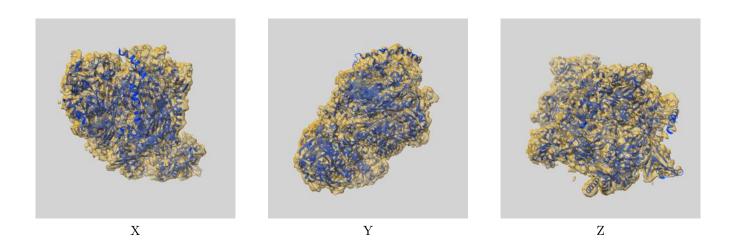
This section was not generated. No FSC curve or half-maps provided.



## 9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-3366 and PDB model 5G06. Per-residue inclusion information can be found in section 3 on page 7.

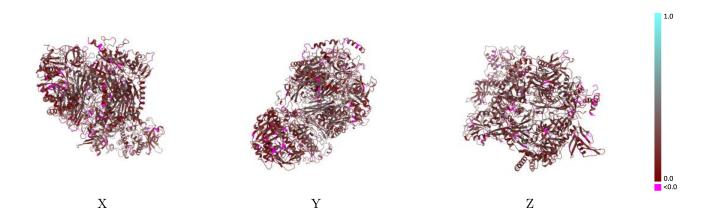
### 9.1 Map-model overlay (i)



The images above show the 3D surface view of the map at the recommended contour level 0.02 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.

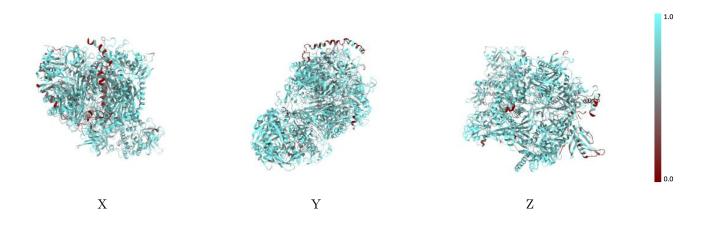


### 9.2 Q-score mapped to coordinate model (i)



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.

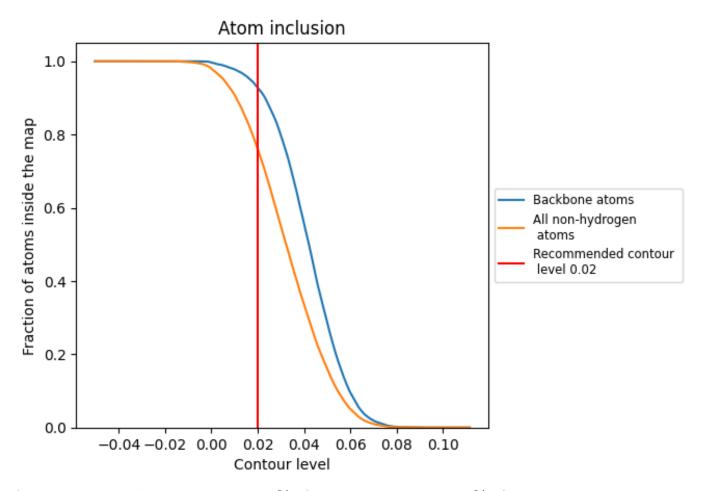
#### 9.3 Atom inclusion mapped to coordinate model (i)



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 (0.02).



### 9.4 Atom inclusion (i)



At the recommended contour level, 93% 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 (0.02) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.7609	0.2210
A	0.7728	0.2500
В	0.7973	0.2640
С	0.7608	0.2210
D	0.7871	0.2500
E	0.7833	0.2560
F	0.7886	0.2430
G	0.8117	0.2240
Н	0.7978	0.2180
I	0.6842	0.1960
J	0.7695	0.1990
P	0.3070	0.1120



