

# wwPDB EM Validation Summary Report (i)

#### Nov 14, 2022 – 09:55 PM EST

PDB ID	:	7KJV				
EMDB ID	:	EMD-22899				
Title	:	tructure of HIV-1 reverse transcriptase initiation complex core				
Authors	:	Ia, B.; Larsen, K.P.; Zhang, J.; Fu, Z.; Montabana, E.; Jackson, L.N.; Chen,				
		D.H.; Puglisi, E.V.				
Deposited on	:	2020-10-26				
Resolution	:	2.80  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:

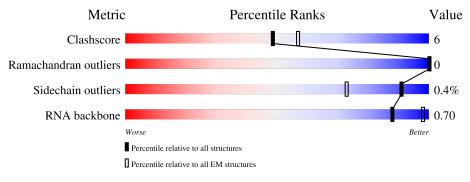
EMDB validation analysis Mogul		0.0.1.dev43 1.8.5 (274361), CSD as541be (2020)
MolProbity	:	4.02b-467
buster-report	:	1.1.7(2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.9
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

# 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 2.80 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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f EM} {f structures} \ (\#{f Entries})$
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826
RNA backbone	4643	859

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  $\leq 40\%$ ). The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain						
1	А	562		82%			13% •		
2	В	442			7% 9%				
3	С	26	5(	0%	27%	8%	15%		
4	D	39	18%	36%	•	44%			



# 2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 8361 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 Reverse transcriptase/ribonuclease H.

Mol	Chain	Residues	Atoms			AltConf	Trace		
1	А	538	Total 4186	С 2717	N 708	0 755	S 6	0	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Actual Comment	
А	-1	MET	-	expression tag	UNP P03366
А	0	VAL	-	expression tag	UNP P03366
А	258	CYS	GLN	engineered mutation	UNP P03366
А	280	SER	CYS	conflict	UNP P03366
А	478	GLN	GLU	engineered mutation	UNP P03366

• Molecule 2 is a protein called reverse transcriptase p51 subunit.

Mol	Chain	Residues	Atoms			AltConf	Trace		
2	В	403	Total 3214	C 2099	N 534	O 576	${ m S}{ m 5}$	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
В	-1	MET	-	expression tag	UNP P03366
В	0	VAL	-	expression tag	UNP P03366
В	280	SER	CYS	engineered mutation	UNP P03366

• Molecule 3 is a RNA chain called HIV-1 viral RNA fragment.

Mol	Chain	Residues	Atoms			AltConf	Trace		
3	C	22	Total 474	C 211	N 89	0 152	Р 22	0	0

• Molecule 4 is DNA/RNA hybrid called tRNA lysine 3 fragment with GNRA tetraloop.

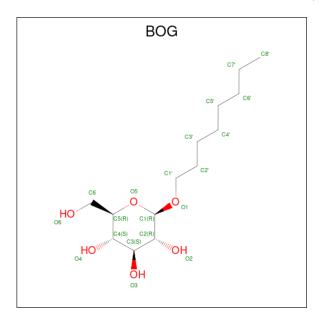


Mol	Chain	Residues	Atoms				AltConf	Trace		
4	D	22	Total 466	C 209	N 79	0 155	Р 22	S 1	0	0

• Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	AltConf
5	А	1	Total Mg 1 1	0

• Molecule 6 is octyl beta-D-glucopyranoside (three-letter code: BOG) (formula:  $C_{14}H_{28}O_6$ ).



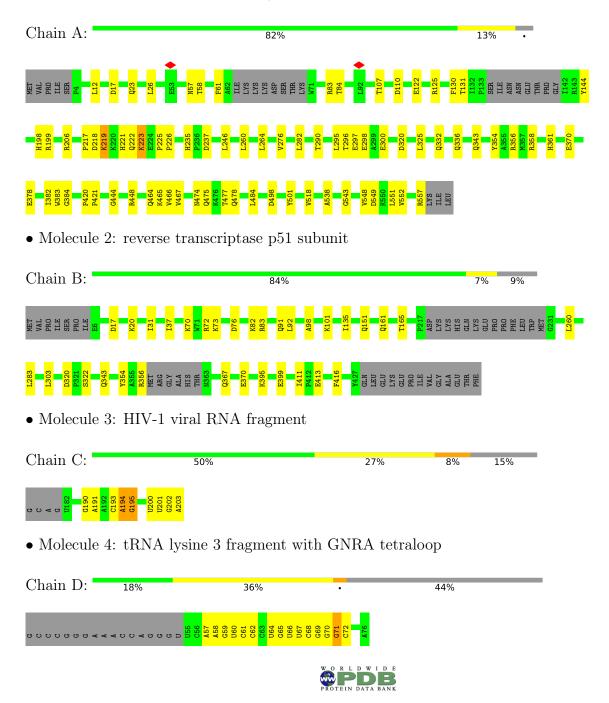
Mol	Chain	Residues	Atoms	AltConf
6	А	1	Total         C         O           20         14         6	0



# 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: Reverse transcriptase/ribonuclease H



# 4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	1344402	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 $(e^-/\text{\AA}^2)$	100	Depositor
Minimum defocus (nm)	Not provided	
Maximum defocus (nm)	Not provided	
Magnification	Not provided	
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	3.890	Depositor
Minimum map value	-2.589	Depositor
Average map value	0.001	Depositor
Map value standard deviation	0.086	Depositor
Recommended contour level	0.3	Depositor
Map size (Å)	314.88, 314.88, 314.88	wwPDB
Map dimensions	384, 384, 384	wwPDB
Map angles $(^{\circ})$	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	0.82, 0.82, 0.82	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: MG, BOG, G47  $\,$ 

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		
IVIOI	Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.24	0/4298	0.41	0/5864	
2	В	0.24	0/3305	0.41	0/4508	
3	С	0.14	0/530	0.70	0/825	
4	D	0.35	0/489	0.81	0/756	
All	All	0.24	0/8622	0.47	0/11953	

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	А	4186	0	4074	61	0
2	В	3214	0	3164	19	0
3	С	474	0	240	10	0
4	D	466	0	241	16	0
5	А	1	0	0	0	0
6	А	20	0	28	1	0
All	All	8361	0	7747	99	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.



Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:223:LYS:HE3	6:A:702:BOG:H1	1.57	0.85
1:A:475:GLN:HB3	1:A:501:TYR:CE2	2.17	0.79
1:A:122:GLU:OE1	1:A:125:ARG:NH1	2.23	0.72
1:A:557:ARG:NE	3:C:201:U:OP1	2.23	0.72
1:A:444:GLY:O	1:A:478:GLN:NE2	2.23	0.72

The worst 5 of 99 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

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	Perce	ntiles
1	А	532/562~(95%)	511 (96%)	21 (4%)	0	100	100
2	В	397/442~(90%)	387~(98%)	10 (2%)	0	100	100
All	All	929/1004~(92%)	898~(97%)	31 (3%)	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	Analysed Rotameric		Percentiles		
1	А	425/502~(85%)	423 (100%)	2 (0%)	88 96		

Continued on next page...



Continued from previous page...

Mol	Chain	Analysed Rotameric		Outliers	Percentiles		
2	В	336/402~(84%)	335~(100%)	1 (0%)	92 98		
All	All	761/904~(84%)	758 (100%)	3~(0%)	91 97		

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	219	LYS
1	А	223	LYS
2	В	70	LYS

Sometimes side chains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such side chains are listed below:

Mol	Chain	Res	Type
1	А	242	GLN
1	А	478	GLN
2	В	91	GLN

#### 5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	С	21/26~(80%)	2 (9%)	0
4	D	20/39~(51%)	1 (5%)	0
All	All	41/65~(63%)	3~(7%)	0

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	С	194	А
3	С	195	G
4	D	70	G

There are no RNA pucker outliers to report.

### 5.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue 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 7	Type Chain		Dog	Link	Bo	ond leng	$\operatorname{ths}$	B	ond ang	les
		туре	e Cham Res	nes		Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
	4	G47	D	71	1,4,3	19,27,28	1.11	3 (15%)	17,38,41	1.42	2 (11%)

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
4	G47	D	71	1,4,3	-	3/7/25/26	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	71	G47	C5-C6	-2.56	1.42	1.47
4	D	71	G47	C8-N7	-2.45	1.30	1.35
4	D	71	G47	C5-C4	-2.33	1.37	1.43

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
4	D	71	G47	C6A-N2-C2	-4.05	115.82	123.41
4	D	71	G47	O6-C6-C5	2.70	129.65	124.37

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	D	71	G47	N1-C2-N2-C6A
4	D	71	G47	N3-C2-N2-C6A
4	D	71	G47	N2-C6A-C7A-SG

There are no ring outliers.

1 monomer is involved in 2 short contacts:



Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	71	G47	2	0

#### 5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

### 5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

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).

Γ	Mol	Type	Chain	Res Link		Bond lengths			В	ond ang	les
	101	туре	Ullalli	nes		Counts   RMSZ		# Z >2	Counts	RMSZ	# Z  > 2
	6	BOG	А	702	-	20,20,20	0.47	0	$25,\!25,\!25$	0.59	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
6	BOG	А	702	-	-	6/11/31/31	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	А	702	BOG	O5-C1-O1-C1'
6	А	702	BOG	C2-C1-O1-C1'
6	А	702	BOG	C2'-C1'-O1-C1
6	А	702	BOG	C5'-C6'-C7'-C8'
6	А	702	BOG	C1'-C2'-C3'-C4'

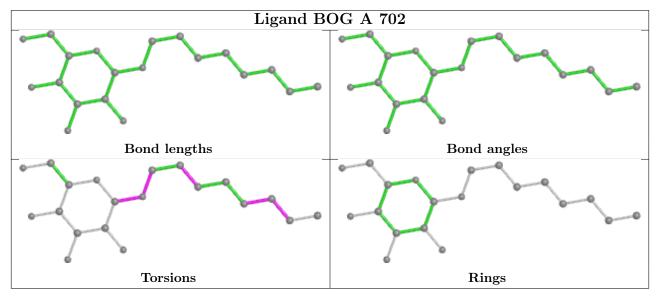


There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	А	702	BOG	1	0

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 then 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 sufficient 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.



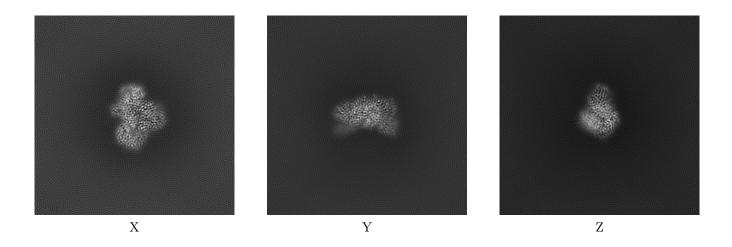
# 6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-22899. 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)

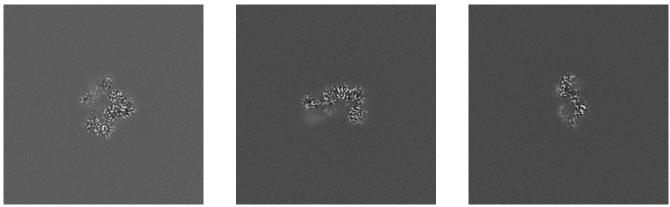
#### 6.1.1 Primary map



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

### 6.2 Central slices (i)

#### 6.2.1 Primary map



X Index: 192

Y Index: 192



Z Index: 192

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: 203

Y Index: 182

Z Index: 208

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 0.3. 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)

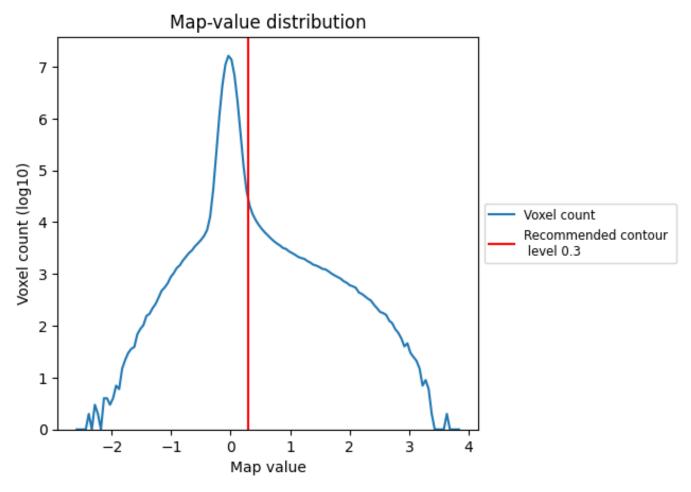
This section was not generated. No masks/segmentation were deposited.



## 7 Map analysis (i)

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

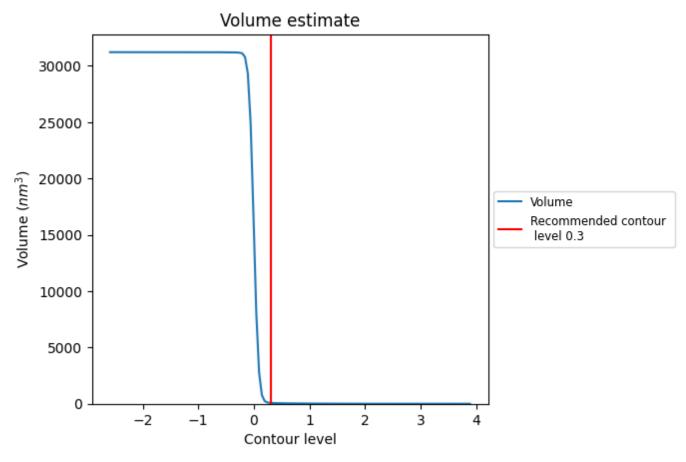
## 7.1 Map-value distribution (i)



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)

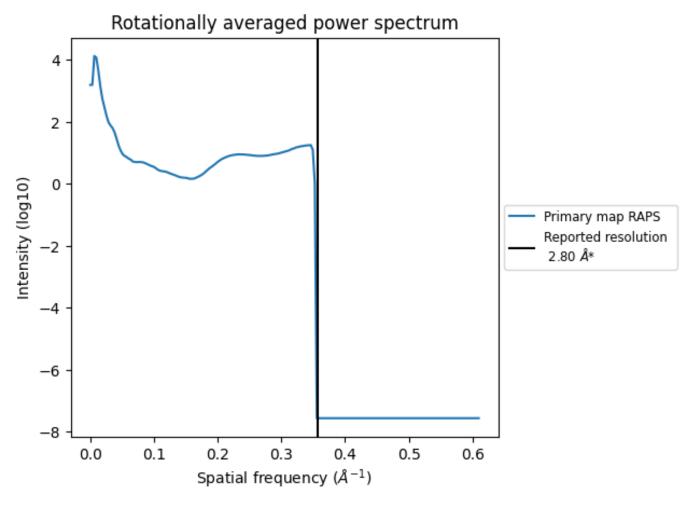


The volume at the recommended contour level is 80  $\rm nm^3;$  this corresponds to an approximate mass of 72 kDa.

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 0.357  ${\rm \AA^{-1}}$ 



# 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-22899 and PDB model 7KJV. Per-residue inclusion information can be found in section 3 on page 5.

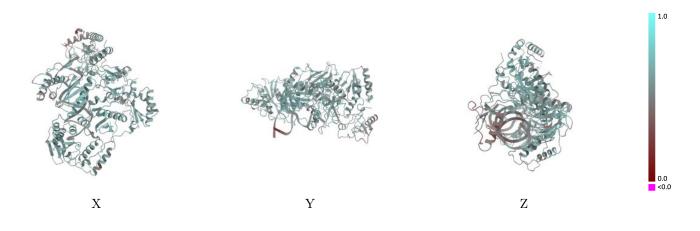
## 9.1 Map-model overlay (i)



The images above show the 3D surface view of the map at the recommended contour level 0.3 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

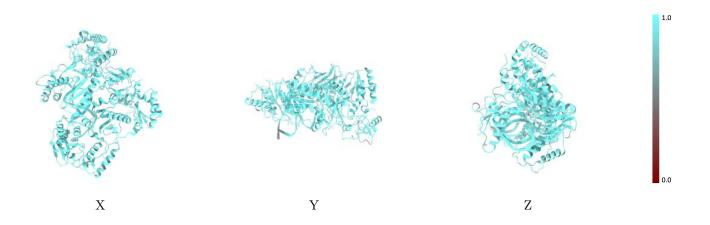


### 9.2 Q-score mapped to coordinate model (i)



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.

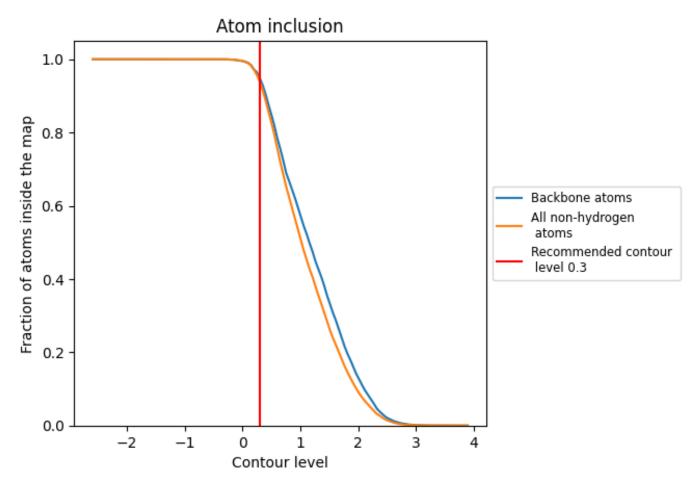
#### 9.3 Atom inclusion mapped to coordinate model (i)



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 (0.3).



### 9.4 Atom inclusion (i)



At the recommended contour level, 95% of all backbone atoms, 94% of all non-hydrogen atoms, are inside the map.



1.0

0.0 <0.0

### 9.5 Map-model fit summary (i)

The table lists the average atom inclusion at the recommended contour level (0.3) and Q-score for the entire model and for each chain.

		-
$\mathbf{Chain}$	Atom inclusion	Q-score
All	0.9373	0.5640
А	0.9302	0.5600
В	0.9588	0.5960
С	0.8650	0.4610
D	0.9292	0.4800

