wwPDB EM Validation Summary Report (i)

Feb 25, 2024 – 09:58 PM EST

PDB ID	:	6WB2
EMDB ID	:	EMD-21585
Title	:	+3 extended HIV-1 reverse transcriptase initiation complex core (displaced state)
Authors	:	Larsen, K.P.; Jackson, L.N.; Kappel, K.; Zhang, J.; Chen, D.H.; Puglisi, E.V.
Deposited on	:	2020-03-26
Resolution	:	4.50 Å(reported)

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

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

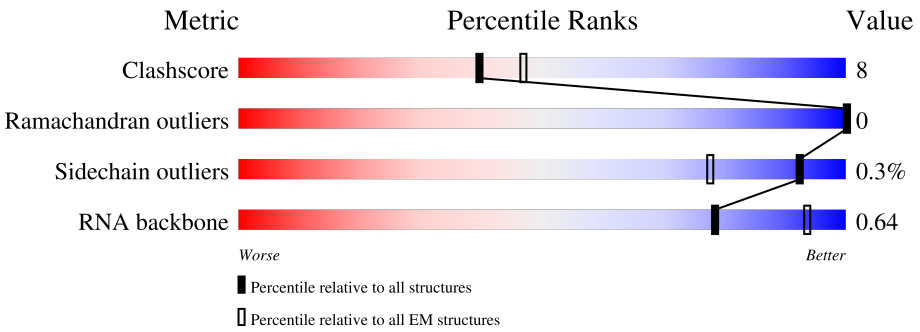
EMDB validation analysis	:	0.0.1.dev70
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	1.9.13
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 4.50 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
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 >=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	C	101	14% 17% • 67%
2	D	79	16% 37% 47%
3	A	570	81% 13% 5%
4	B	442	77% 15% 8%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 8407 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 RNA chain called HIV-1 viral RNA genome fragment.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
1	C	33	705	316	132	225	32	0	0

- Molecule 2 is DNA/RNA hybrid called tRNA lysine 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	D	42	880	398	155	287	40	0	0

- Molecule 3 is a protein called Reverse transcriptase/ribonuclease H.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	A	540	3854	2494	667	688	5	0	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	expression tag	UNP P03366
A	0	VAL	-	expression tag	UNP P03366
A	258	CYS	GLN	engineered mutation	UNP P03366
A	280	SER	CYS	engineered mutation	UNP P03366
A	478	GLN	GLU	engineered mutation	UNP P03366
A	561	ASP	-	expression tag	UNP P03366
A	562	LEU	-	expression tag	UNP P03366
A	563	GLY	-	expression tag	UNP P03366
A	564	THR	-	expression tag	UNP P03366
A	565	LEU	-	expression tag	UNP P03366
A	566	VAL	-	expression tag	UNP P03366
A	567	PRO	-	expression tag	UNP P03366
A	568	ARG	-	expression tag	UNP P03366

- Molecule 4 is a protein called reverse transcriptase p51 subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	B	406	2968	1942	511	511	4	0	0

There are 3 discrepancies between the modelled and reference sequences:

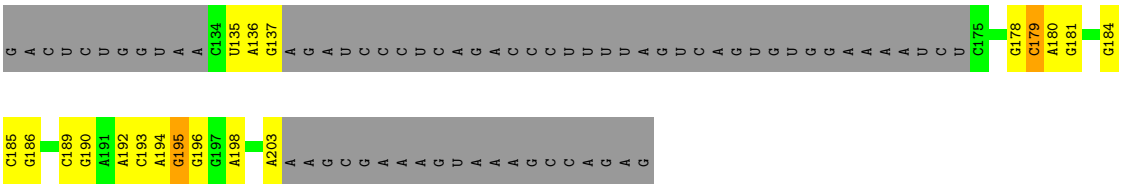
Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	MET	-	expression tag	UNP P03366
B	0	VAL	-	expression tag	UNP P03366
B	280	SER	CYS	engineered mutation	UNP P03366

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.

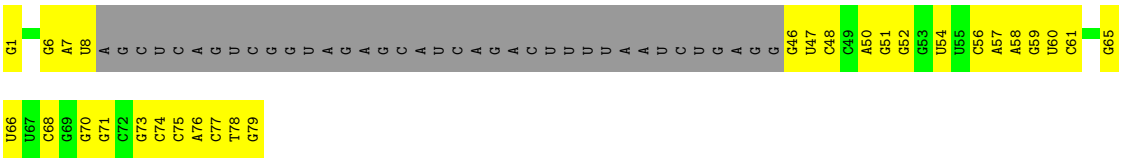
• Molecule 1: HIV-1 viral RNA genome fragment

Chain C: 14% 17% • 67%



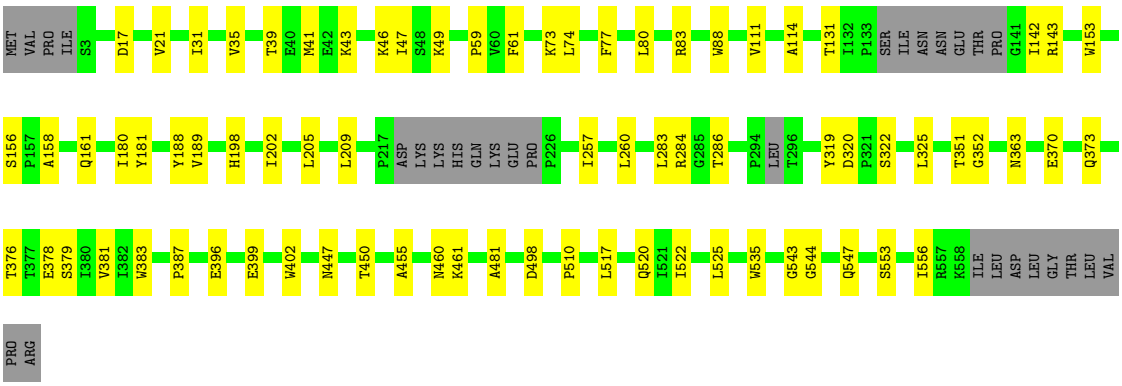
• Molecule 2: tRNA lysine 3

Chain D: 16% 37% 47%



• Molecule 3: Reverse transcriptase/ribonuclease H

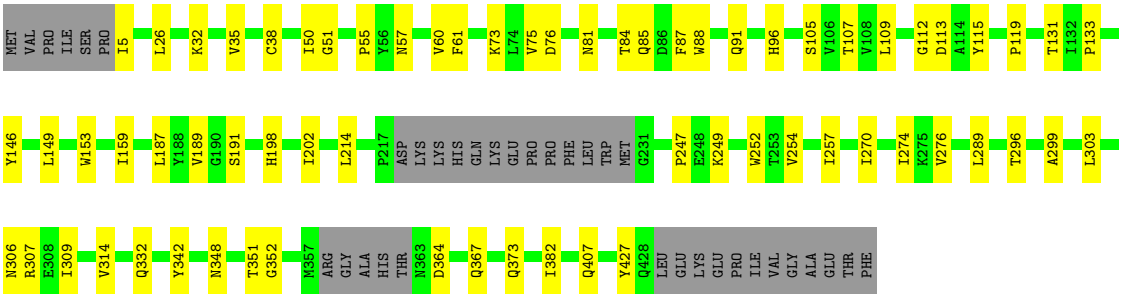
Chain A: 81% 13% 5%



• Molecule 4: reverse transcriptase p51 subunit

Chain B:

77%
15%
8%


MET VAL PRO ILE SER PRO I5 L26 K32 V35 C38 I50 G51 P55 Y56 N57 V60 F61 K73 L74 V75 D76 N81 T84 Q85 D86 F87 W88 Q91 H96 S105 V106 T107 V108 L109 G112 D113 A114 Y115 P119 T131 I132 P133
Y146 L149 W153 I159 L187 Y188 V189 G190 S191 H198 I202 L214 P217 ASP LYS LYS HIS GLN LYS GLU PRO PRO PHE LEU TRP MET G231 P247 E248 K249 W252 T253 V254 I257 I270 I274 K275 V276 L289 T296 A299 L303
N306 R307 E308 I309 V314 Q332 Y342 N348 T351 G352 M357 ARG GLY ALA HIS THR N363 D364 Q367 Q373 I382 Q407 Y427 Q428 LEU GLU LYS GLU PRO ILE VAL GLY ALA GLU THR PHE

4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	291904	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	77.6	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	0.102	Depositor
Minimum map value	-0.038	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.003	Depositor
Recommended contour level	0.013	Depositor
Map size (Å)	271.36, 271.36, 271.36	wwPDB
Map dimensions	256, 256, 256	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.06, 1.06, 1.06	Depositor

5 Model quality (i)

5.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\|>5$	RMSZ	$\#\|Z\|>5$
1	C	0.18	0/788	0.76	0/1226
2	D	0.23	0/981	0.79	0/1525
3	A	0.23	0/3960	0.42	0/5441
4	B	0.23	0/3060	0.42	0/4215
All	All	0.23	0/8789	0.52	0/12407

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	C	705	0	358	16	0
2	D	880	0	451	22	0
3	A	3854	0	3407	44	0
4	B	2968	0	2669	39	0
All	All	8407	0	6885	117	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 8.

The worst 5 of 117 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:D:1:G:H1	2:D:54:U:H3	1.42	0.65
3:A:351:THR:HG22	3:A:352:GLY:H	1.60	0.65
3:A:46:LYS:HG3	3:A:47:ILE:HD12	1.79	0.64
3:A:399:GLU:HA	3:A:402:TRP:CD1	2.31	0.64
3:A:181:TYR:HB2	3:A:188:TYR:HB3	1.78	0.64

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	
3	A	532/570 (93%)	514 (97%)	18 (3%)	0	100	100
4	B	400/442 (90%)	389 (97%)	11 (3%)	0	100	100
All	All	932/1012 (92%)	903 (97%)	29 (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	Rotameric	Outliers	Percentiles	
3	A	326/509 (64%)	324 (99%)	2 (1%)	86	92
4	B	262/402 (65%)	262 (100%)	0	100	100
All	All	588/911 (64%)	586 (100%)	2 (0%)	92	95

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	43	LYS
3	A	73	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	269	GLN
4	B	348	ASN
4	B	373	GLN
4	B	367	GLN
3	A	547	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	C	31/101 (30%)	6 (19%)	0
2	D	36/79 (45%)	6 (16%)	0
All	All	67/180 (37%)	12 (17%)	0

5 of 12 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	C	178	G
1	C	179	C
1	C	192	A
1	C	195	G
1	C	196	G

There are no RNA pucker outliers to report.

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.

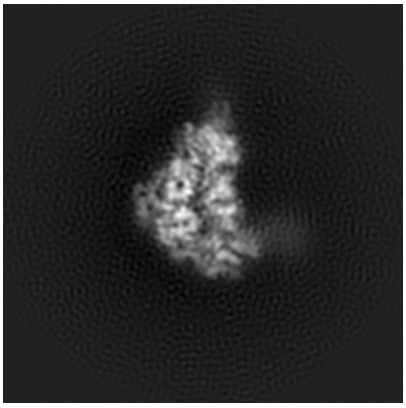
6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-21585. These allow visual inspection of the internal detail of the map and identification of artifacts.

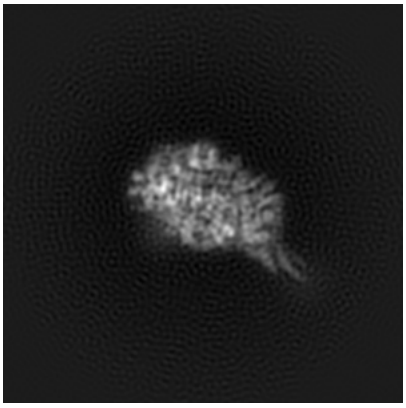
No raw map or half-maps were deposited for this entry and therefore no images, graphs, etc. pertaining to the raw map can be shown.

6.1 Orthogonal projections (i)

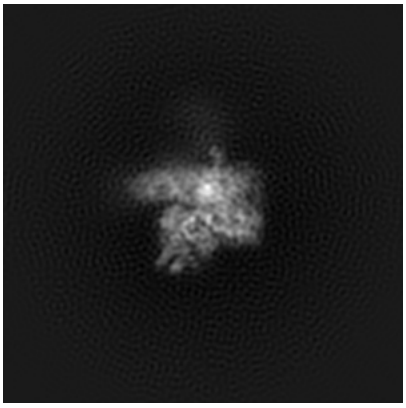
6.1.1 Primary map



X



Y

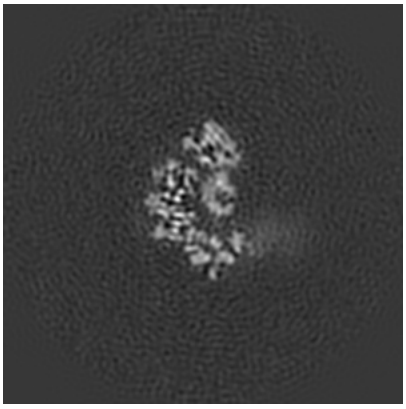


Z

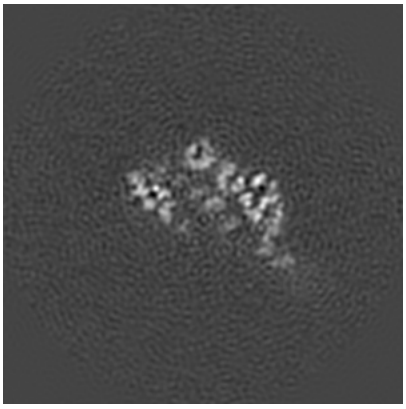
The images above show the map projected in three orthogonal directions.

6.2 Central slices (i)

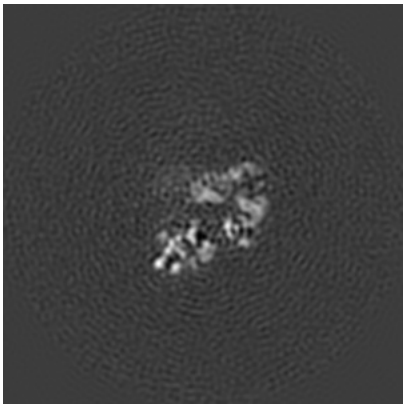
6.2.1 Primary map



X Index: 128



Y Index: 128

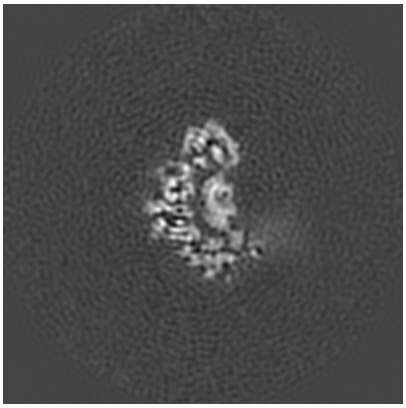


Z Index: 128

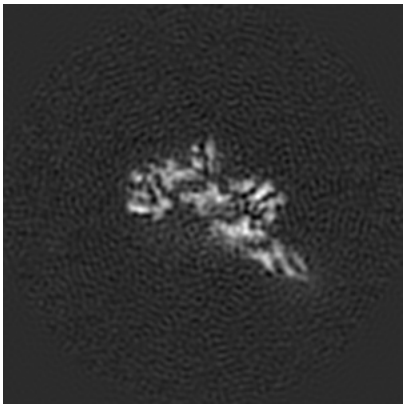
The images above show central slices of the map in three orthogonal directions.

6.3 Largest variance slices (i)

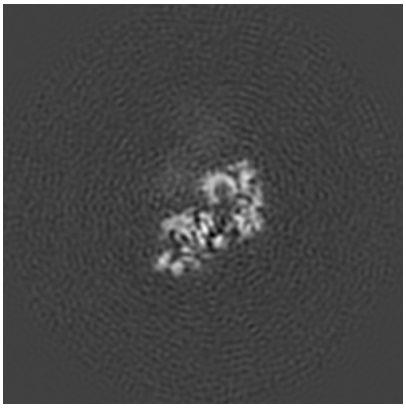
6.3.1 Primary map



X Index: 131



Y Index: 135

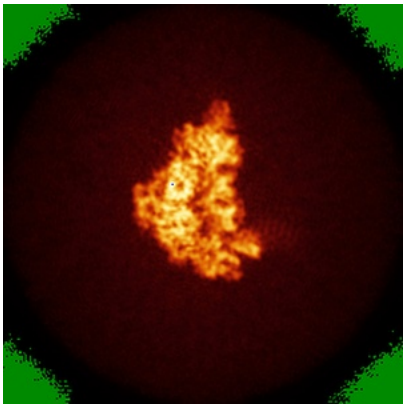


Z Index: 121

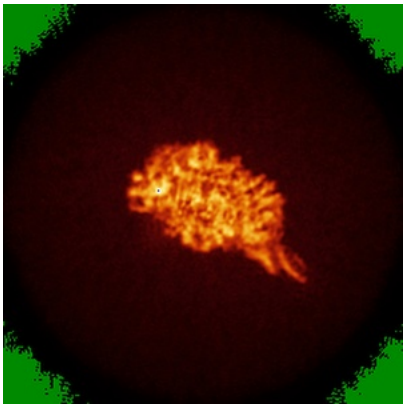
The images above show the largest variance slices of the map in three orthogonal directions.

6.4 Orthogonal standard-deviation projections (False-color) (i)

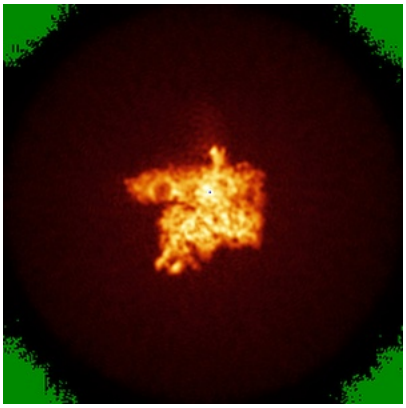
6.4.1 Primary map



X



Y

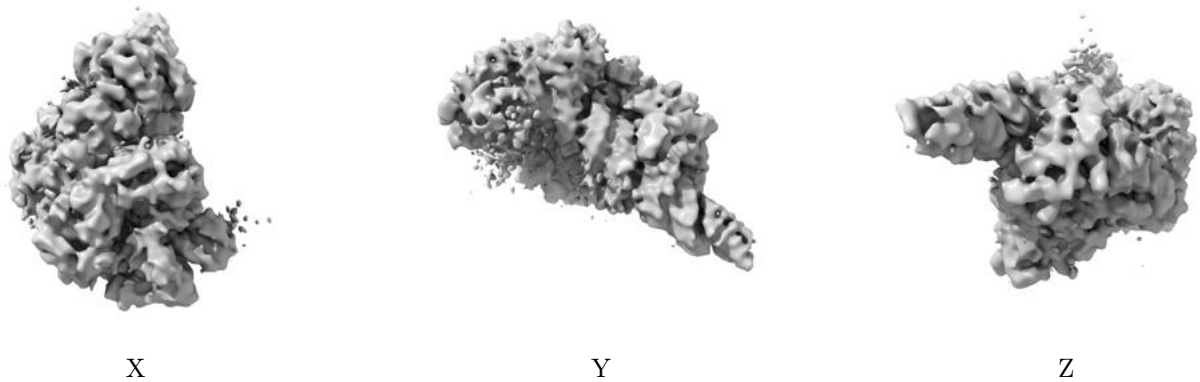


Z

The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views (i)

6.5.1 Primary map



The images above show the 3D surface view of the map at the recommended contour level 0.013. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

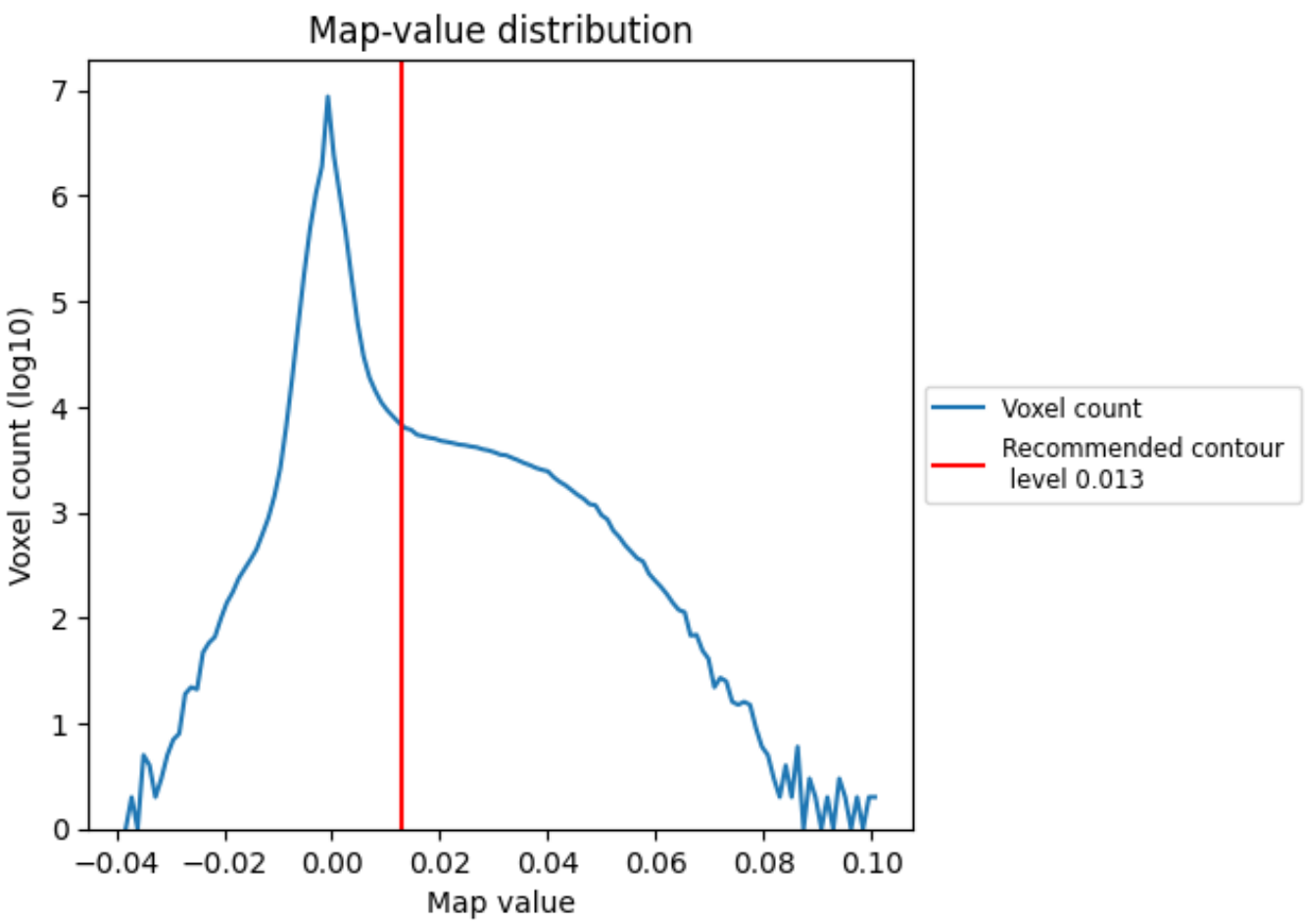
6.6 Mask visualisation (i)

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

7 Map analysis (i)

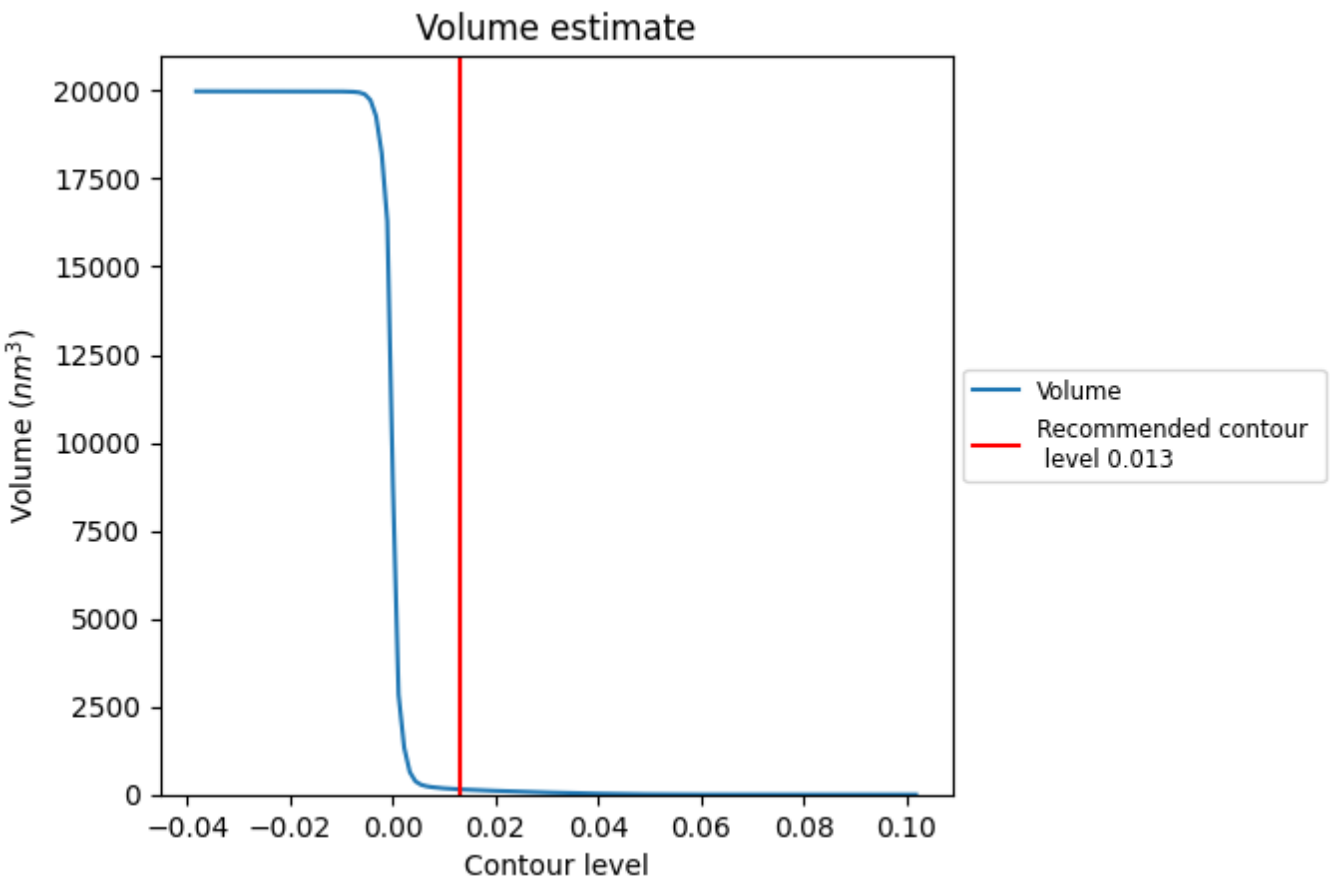
This section contains the results of statistical analysis of the map.

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.

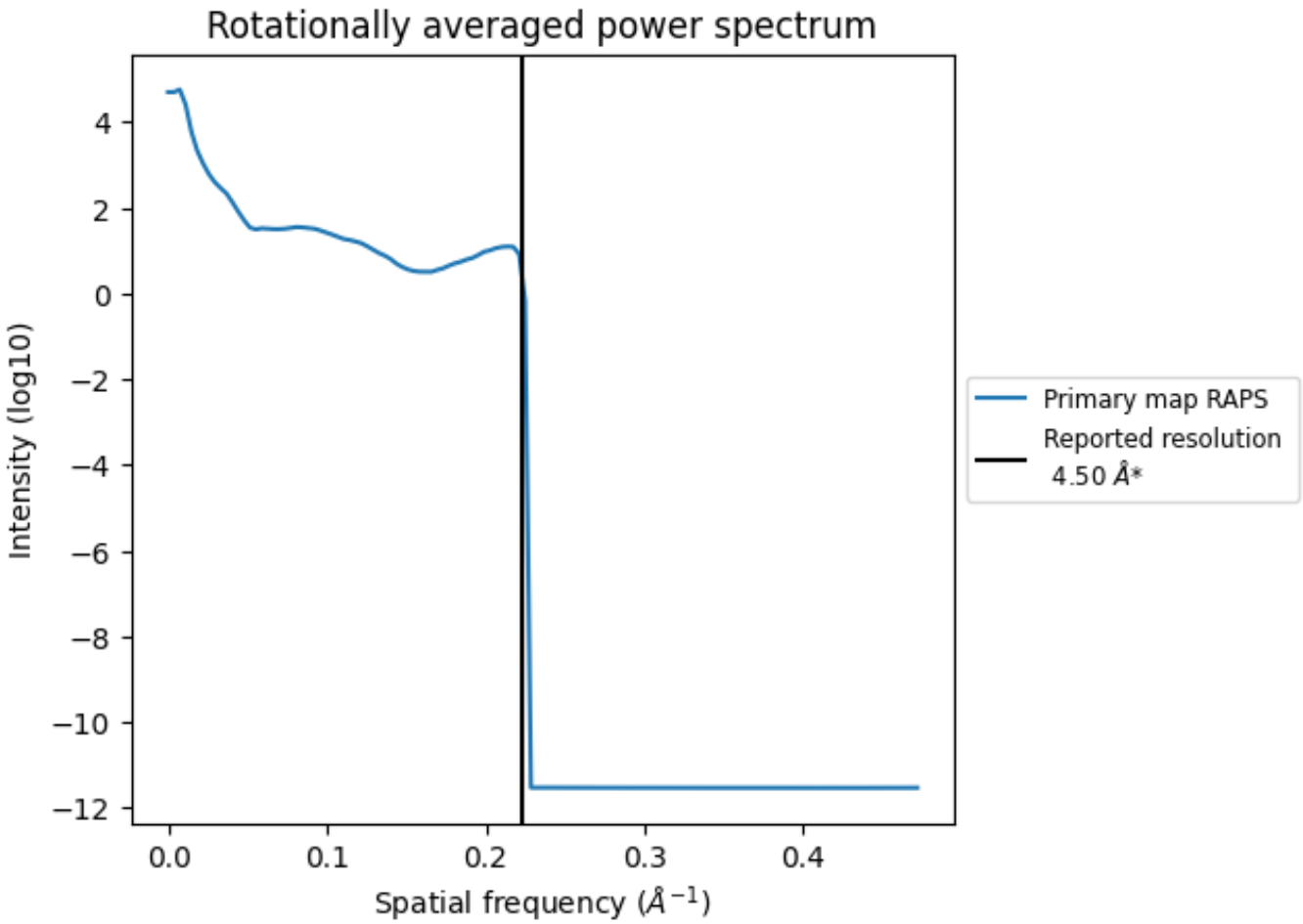
7.2 Volume estimate (i)



The volume at the recommended contour level is 149 nm^3; this corresponds to an approximate mass of 135 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 0.222 Å^{-1}

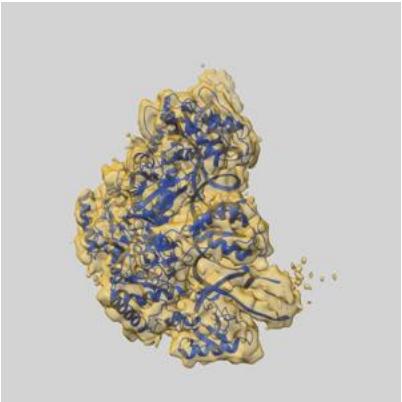
8 Fourier-Shell correlation (i)

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

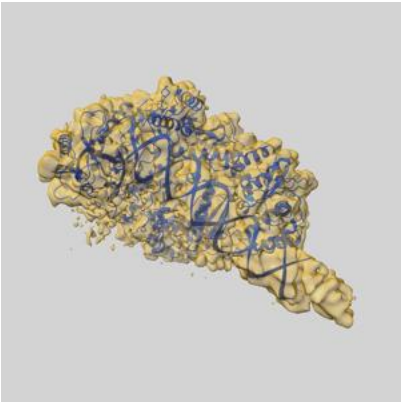
9 Map-model fit (i)

This section contains information regarding the fit between EMDB map EMD-21585 and PDB model 6WB2. Per-residue inclusion information can be found in section 3 on page 5.

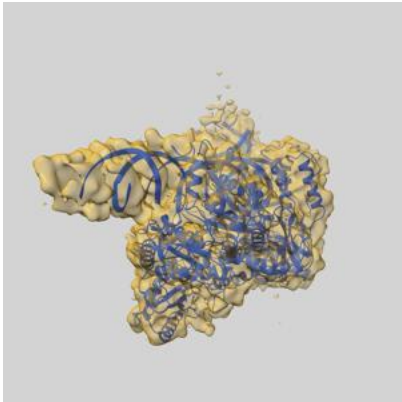
9.1 Map-model overlay (i)



X



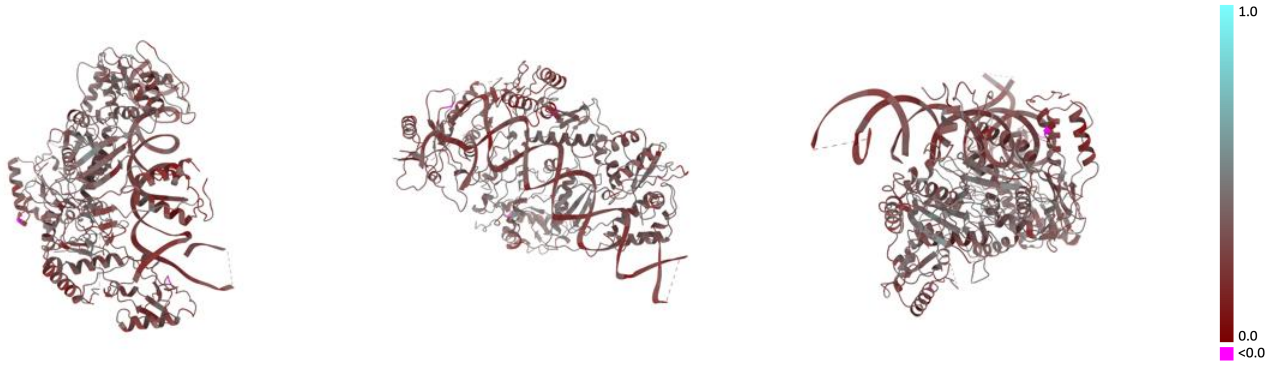
Y



Z

The images above show the 3D surface view of the map at the recommended contour level 0.013 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

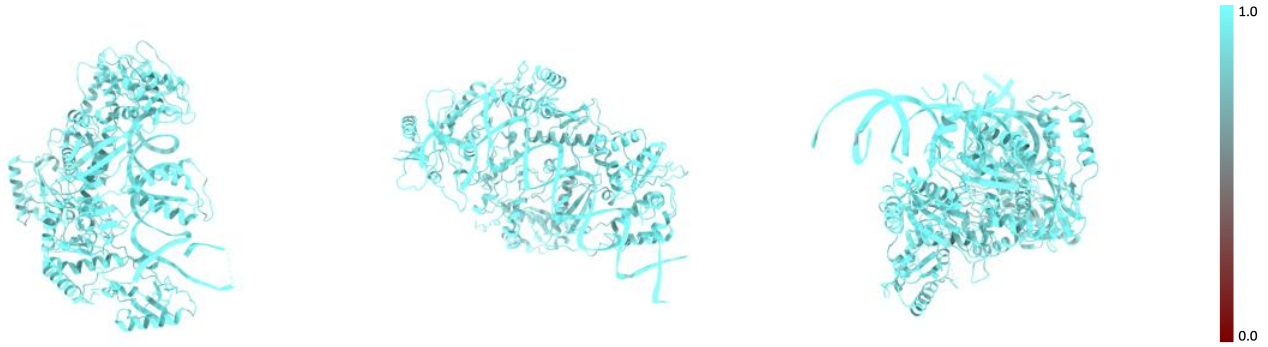
9.2 Q-score mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

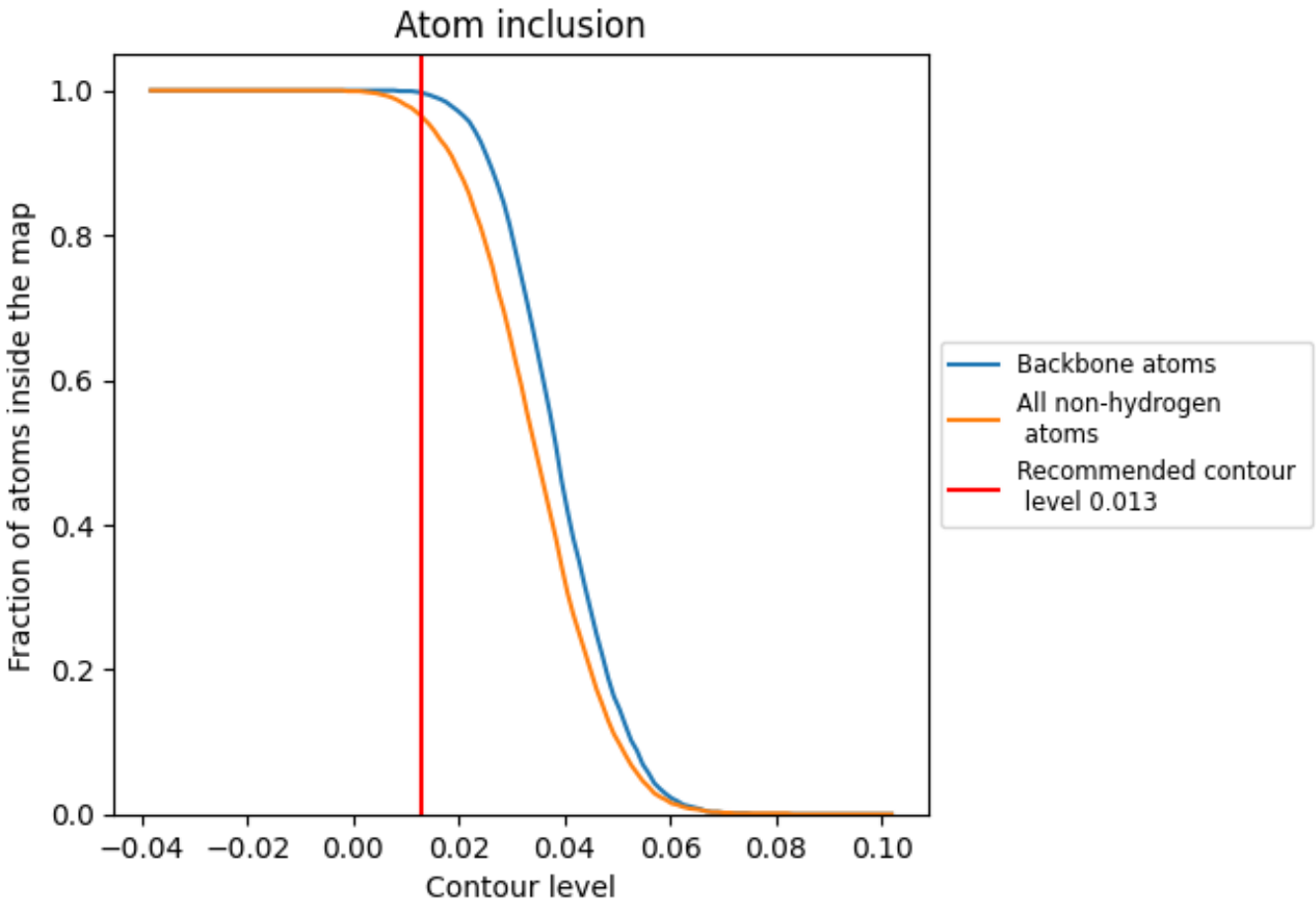
9.3 Atom inclusion mapped to coordinate model (i)



X Y Z

The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.013).

9.4 Atom inclusion (i)



At the recommended contour level, 100% of all backbone atoms, 96% 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 (0.013) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.9640	0.3320
A	0.9630	0.3370
B	0.9550	0.3660
C	0.9870	0.2710
D	0.9810	0.2480