WORLD WIDE

wwPDB

PROTEIN DATA BANK

wwPDB EM Validation Summary Report (i)

Jul 3, 2024 – 09:20 am BST

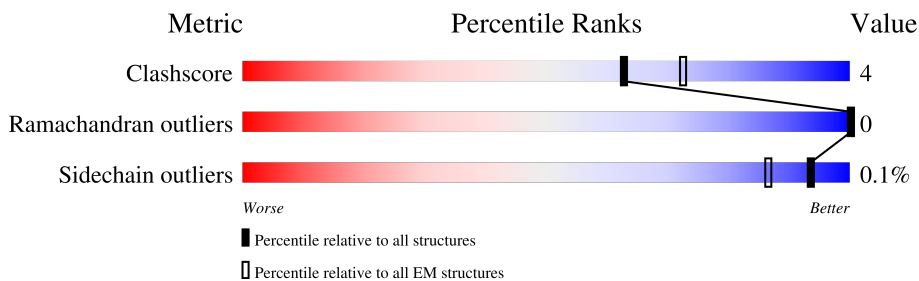|		
PDB ID	:	7O4I
EMDB ID	:	EMD-12719
Title	:	Yeast RNA polymerase II transcription pre-initiation complex with initial transcription bubble
Authors	:	Schilbach, S.; Aibara, S.; Dienemann, C.; Grabbe, F.; Cramer, P.
Deposited on	:	2021-04-06
Resolution	:	3.20 Å(reported)
Based on initial model	:	5OQJ

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.
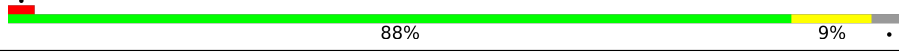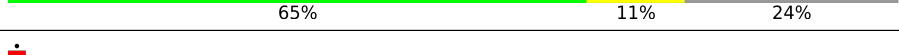
Metric	Percentile Ranks	Value
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		0.1%

Worse   *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all EM structures
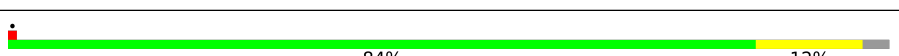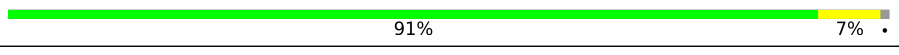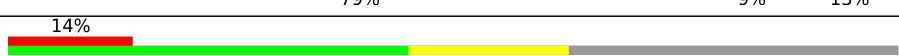
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

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 <=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.

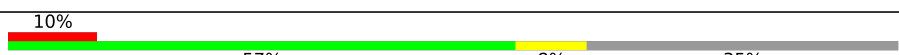| Mol	Chain	Length	Quality of chain
1	0	778	84% 13% •
2	1	645	76% 5% 19%
3	2	517	16% 76% 10% 14%
4	3	324	36% • 60%
5	4	341	72% 13% 14%
6	5	76	16% 80% 5% 14%
7	6	464	68% 9% 23%
8	7	843	15% 63% 10% 27%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	A	1733	75% 8% 17%
10	B	1224	88% 9%
11	C	347	69% 7% 23%
12	D	221	65% 11% 24%
13	E	215	87% 12%
14	F	155	50% 5% 45%
15	G	177	84% 12%
16	H	146	88% 8%
17	I	122	85% 10% 5%
18	J	70	91% 7%
19	K	120	87% 9%
20	L	70	59% 6% 36%
21	M	352	79% 9% 13%
22	N	106	14% 45% 18% 37%
23	O	247	63% 11% 27%
24	Q	738	8% 26% 70%
25	R	400	8% 59% 8% 33%
26	T	106	13% 48% 15% 37%
27	U	171	9% 54% 9% 37%
28	V	129	9% 74% 7% 19%
29	W	492	9% 53% 10% 37%
30	X	328	10% 57% 8% 35%

2   Entity composition ⓘ

There are 35 unique types of molecules in this entry. The entry contains 73949 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 General transcription and DNA repair factor IIH helicase subunit XPD.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	0	752	6091	3882	1029	1142	38	0	0

- Molecule 2 is a protein called General transcription and DNA repair factor IIH subunit TFB1.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	1	522	4214	2660	734	798	22	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
1	-2	GLY	-	expression tag	UNP P32776
1	-1	GLY	-	expression tag	UNP P32776
1	0	SER	-	expression tag	UNP P32776

- Molecule 3 is a protein called General transcription and DNA repair factor IIH subunit TFB2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	2	445	3597	2327	591	663	16	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
2	-3	GLY	-	expression tag	UNP Q02939
2	-2	PRO	-	expression tag	UNP Q02939
2	-1	GLY	-	expression tag	UNP Q02939
2	0	SER	-	expression tag	UNP Q02939

- Molecule 4 is a protein called RNA polymerase II transcription factor B subunit 3.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	3	131	Total	C	N	O	S	0	0
			1089	692	180	209	8		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
3	-2	GLY	-	expression tag	UNP Q03290
3	-1	PRO	-	expression tag	UNP Q03290
3	0	HIS	-	expression tag	UNP Q03290

- Molecule 5 is a protein called General transcription and DNA repair factor IIH subunit TFB4.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	4	293	Total	C	N	O	S	0	0
			2274	1454	377	429	14		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
4	-2	SER	-	expression tag	UNP Q12004
4	-1	ASN	-	expression tag	UNP Q12004
4	0	ALA	-	expression tag	UNP Q12004

- Molecule 6 is a protein called General transcription and DNA repair factor IIH subunit TFB5.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	5	65	Total	C	N	O	S	0	0
			514	326	90	95	3		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
5	-3	GLY	-	expression tag	UNP Q3E7C1
5	-2	PRO	-	expression tag	UNP Q3E7C1
5	-1	GLY	-	expression tag	UNP Q3E7C1
5	0	SER	-	expression tag	UNP Q3E7C1

- Molecule 7 is a protein called General transcription and DNA repair factor IIH subunit SSL1.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	6	356	Total	C	N	O	S	0	0
			2794	1771	482	513	28		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
6	-2	GLY	-	expression tag	UNP Q04673
6	-1	GLY	-	expression tag	UNP Q04673
6	0	SER	-	expression tag	UNP Q04673

- Molecule 8 is a protein called General transcription and DNA repair factor IIH helicase subunit XPB.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	7	615	Total	C	N	O	S	0	0
			4954	3153	860	914	27		

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB1.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	A	1430	Total	C	N	O	S	0	0
			11254	7088	1967	2137	62		

- Molecule 10 is a protein called DNA-directed RNA polymerase II subunit RPB2.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	B	1180	Total	C	N	O	S	0	0
			9404	5946	1643	1760	55		

- Molecule 11 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	C	266	Total	C	N	O	S	0	0
			2092	1315	348	416	13		

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-28	MET	-	initiating methionine	UNP P16370
C	-27	GLY	-	expression tag	UNP P16370
C	-26	SER	-	expression tag	UNP P16370
C	-25	HIS	-	expression tag	UNP P16370
C	-24	HIS	-	expression tag	UNP P16370
C	-23	HIS	-	expression tag	UNP P16370
C	-22	HIS	-	expression tag	UNP P16370
C	-21	HIS	-	expression tag	UNP P16370
C	-20	HIS	-	expression tag	UNP P16370

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-19	SER	-	expression tag	UNP P16370
C	-18	ASN	-	expression tag	UNP P16370
C	-17	SER	-	expression tag	UNP P16370
C	-16	GLY	-	expression tag	UNP P16370
C	-15	LEU	-	expression tag	UNP P16370
C	-14	ASN	-	expression tag	UNP P16370
C	-13	ASP	-	expression tag	UNP P16370
C	-12	ILE	-	expression tag	UNP P16370
C	-11	PHE	-	expression tag	UNP P16370
C	-10	GLU	-	expression tag	UNP P16370
C	-9	ALA	-	expression tag	UNP P16370
C	-8	GLN	-	expression tag	UNP P16370
C	-7	LYS	-	expression tag	UNP P16370
C	-6	ILE	-	expression tag	UNP P16370
C	-5	GLU	-	expression tag	UNP P16370
C	-4	TRP	-	expression tag	UNP P16370
C	-3	HIS	-	expression tag	UNP P16370
C	-2	GLU	-	expression tag	UNP P16370
C	-1	ASP	-	expression tag	UNP P16370
C	0	THR	-	expression tag	UNP P16370
C	1	GLY	-	expression tag	UNP P16370
C	2	SER	-	expression tag	UNP P16370
C	3	SER	-	expression tag	UNP P16370

- Molecule 12 is a protein called DNA-directed RNA polymerase II subunit RPB4.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	D	167	Total	C	N	O	S	0	0
			1343	829	242	270	2		

- Molecule 13 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

Mol	Chain	Residues	Atoms					AltConf	Trace
13	E	214	Total	C	N	O	S	0	0
			1752	1111	309	321	11		

- Molecule 14 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	F	86	Total	C	N	O	S	0	0
			697	445	118	131	3		

- Molecule 15 is a protein called DNA-directed RNA polymerase II subunit RPB7.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	G	171	Total	C	N	O	S	0	0
			1339	861	222	248	8		

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	172	HIS	-	expression tag	UNP P34087
G	173	HIS	-	expression tag	UNP P34087
G	174	HIS	-	expression tag	UNP P34087
G	175	HIS	-	expression tag	UNP P34087
G	176	HIS	-	expression tag	UNP P34087
G	177	HIS	-	expression tag	UNP P34087

- Molecule 16 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	H	140	Total	C	N	O	S	0	0
			1120	704	188	224	4		

- Molecule 17 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					AltConf	Trace
17	I	116	Total	C	N	O	S	0	0
			944	581	172	181	10		

- Molecule 18 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					AltConf	Trace
18	J	69	Total	C	N	O	S	0	0
			569	362	101	100	6		

- Molecule 19 is a protein called DNA-directed RNA polymerase II subunit RPB11.

Mol	Chain	Residues	Atoms					AltConf	Trace
19	K	115	Total	C	N	O	S	0	0
			924	593	157	172	2		

- Molecule 20 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

Mol	Chain	Residues	Atoms					AltConf	Trace
20	L	45	Total 359	C 221	N 71	O 63	S 4	0	0

- Molecule 21 is a protein called Transcription initiation factor IIB.

Mol	Chain	Residues	Atoms					AltConf	Trace
21	M	308	Total 2360	C 1489	N 406	O 447	S 18	0	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	346	LYS	-	expression tag	UNP P29055
M	347	HIS	-	expression tag	UNP P29055
M	348	HIS	-	expression tag	UNP P29055
M	349	HIS	-	expression tag	UNP P29055
M	350	HIS	-	expression tag	UNP P29055
M	351	HIS	-	expression tag	UNP P29055
M	352	HIS	-	expression tag	UNP P29055

- Molecule 22 is a DNA chain called Non-template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
22	N	67	Total 1374	C 657	N 246	O 404	P 67	0	0

- Molecule 23 is a protein called TATA-box-binding protein.

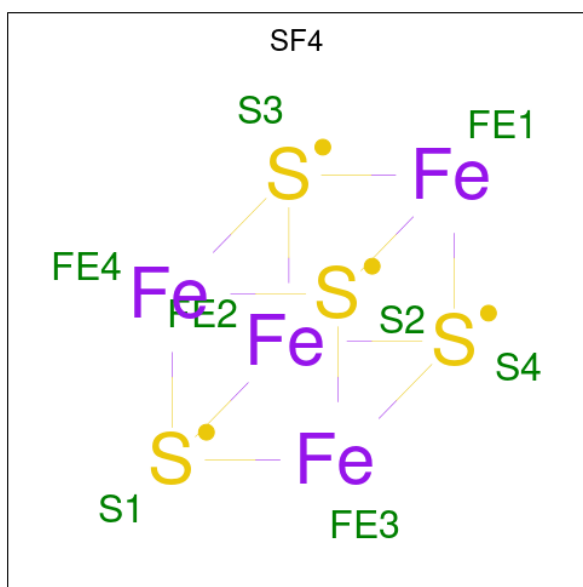| Mol	Chain	Residues	Atoms					AltConf	Trace
23	O	181	Total 1422	C 925	N 243	O 248	S 6	0	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
O	241	LYS	-	expression tag	UNP P13393
O	242	HIS	-	expression tag	UNP P13393
O	243	HIS	-	expression tag	UNP P13393
O	244	HIS	-	expression tag	UNP P13393
O	245	HIS	-	expression tag	UNP P13393
O	246	HIS	-	expression tag	UNP P13393
O	247	HIS	-	expression tag	UNP P13393

- Molecule 24 is a protein called Transcription initiation factor IIF subunit alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
24	Q	221	Total 1871	C 1179	N 346	O 339	S 7	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Q	-2	GLY	-	expression tag	UNP P41895
Q	-1	PRO	-	expression tag	UNP P41895
Q	0	GLY	-	expression tag	UNP P41895

- Molecule 25 is a protein called Transcription initiation factor IIF subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
25	R	268	Total 2230	C 1409	N 392	O 419	S 10	0	0

- Molecule 26 is a DNA chain called Template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
26	T	67	Total 1370	C 656	N 250	O 398	P 66	0	0

- Molecule 27 is a protein called Transcription initiation factor IIA large subunit,Transcription initiation factor IIA large subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
27	U	107	Total 885	C 559	N 147	O 176	S 3	0	0

- Molecule 28 is a protein called Transcription initiation factor IIA subunit 2.

Mol	Chain	Residues	Atoms					AltConf	Trace
28	V	104	Total 815	C 511	N 136	O 164	S 4	0	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
V	123	LYS	-	expression tag	UNP P32774
V	124	HIS	-	expression tag	UNP P32774

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
V	125	HIS	-	expression tag	UNP P32774
V	126	HIS	-	expression tag	UNP P32774
V	127	HIS	-	expression tag	UNP P32774
V	128	HIS	-	expression tag	UNP P32774
V	129	HIS	-	expression tag	UNP P32774

- Molecule 29 is a protein called Transcription initiation factor IIE subunit alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
29	W	312	2532	1595	442	488	7	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
W	483	ALA	-	expression tag	UNP P36100
W	484	ALA	-	expression tag	UNP P36100
W	485	ALA	-	expression tag	UNP P36100
W	486	LEU	-	expression tag	UNP P36100
W	487	GLU	-	expression tag	UNP P36100
W	488	HIS	-	expression tag	UNP P36100
W	489	HIS	-	expression tag	UNP P36100
W	490	HIS	-	expression tag	UNP P36100
W	491	HIS	-	expression tag	UNP P36100
W	492	HIS	-	expression tag	UNP P36100

- Molecule 30 is a protein called Transcription initiation factor IIE subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
30	X	213	1708	1091	292	319	6	0	0

- Molecule 31 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).

Mol	Chain	Residues	Atoms			AltConf
31	0	1	Total	Fe	S	0
			8	4	4	

- Molecule 32 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
32	3	2	Total	Zn	0
			2	2	
32	4	1	Total	Zn	0
			1	1	
32	6	4	Total	Zn	0
			4	4	
32	A	2	Total	Zn	0
			2	2	
32	B	1	Total	Zn	0
			1	1	
32	C	1	Total	Zn	0
			1	1	
32	I	2	Total	Zn	0
			2	2	
32	J	1	Total	Zn	0
			1	1	
32	L	1	Total	Zn	0
			1	1	
32	M	1	Total	Zn	0
			1	1	
32	W	1	Total	Zn	0
			1	1	

- Molecule 33 is BERYLLIUM TRIFLUORIDE ION (three-letter code: BEF) (formula: BeF_3).
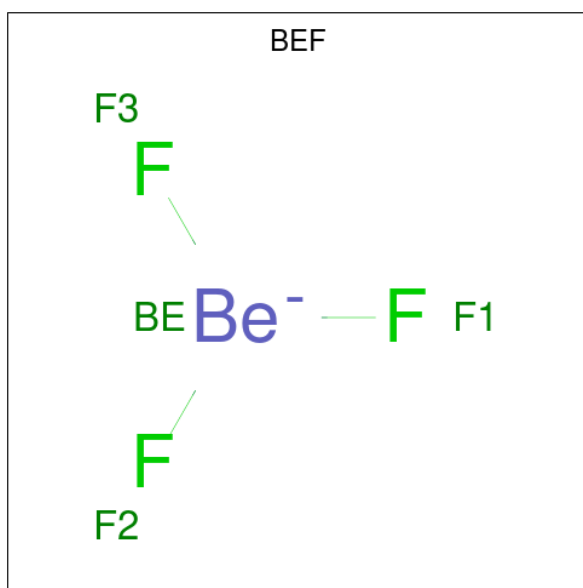
Mol	Chain	Residues	Atoms			AltConf
33	7	1	Total 4	Be 1	F 3	0

- Molecule 34 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
34	7	1	Total 1	Mg 1	0
34	A	1	Total 1	Mg 1	0

- Molecule 35 is A D E N O S I N E - 5'- D I P H O S P H A T E (three-letter code: ADP) (formula: $C_{10}H_{15}N_5O_{10}P_2$).

ADP
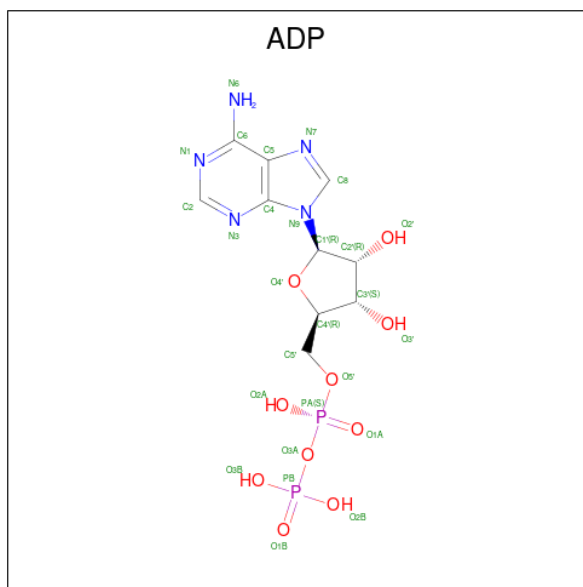
Mol	Chain	Residues	Atoms					AltConf
35	7	1	Total	C	N	O	P	0
			27	10	5	10	2	

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: General transcription and DNA repair factor IIH helicase subunit XPD
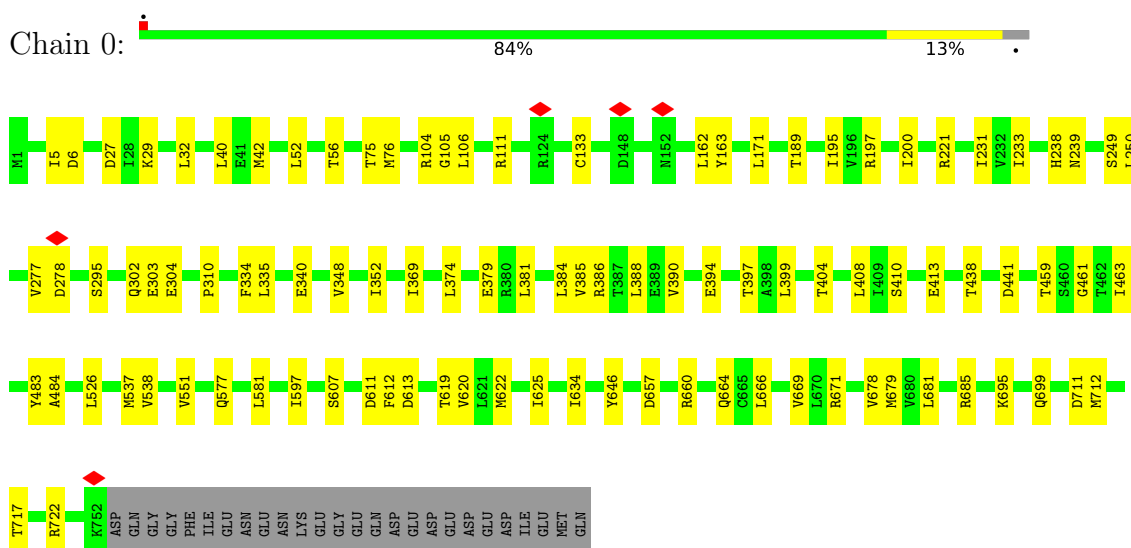
Chain 0:

• Molecule 2: General transcription and DNA repair factor IIH subunit TFB1
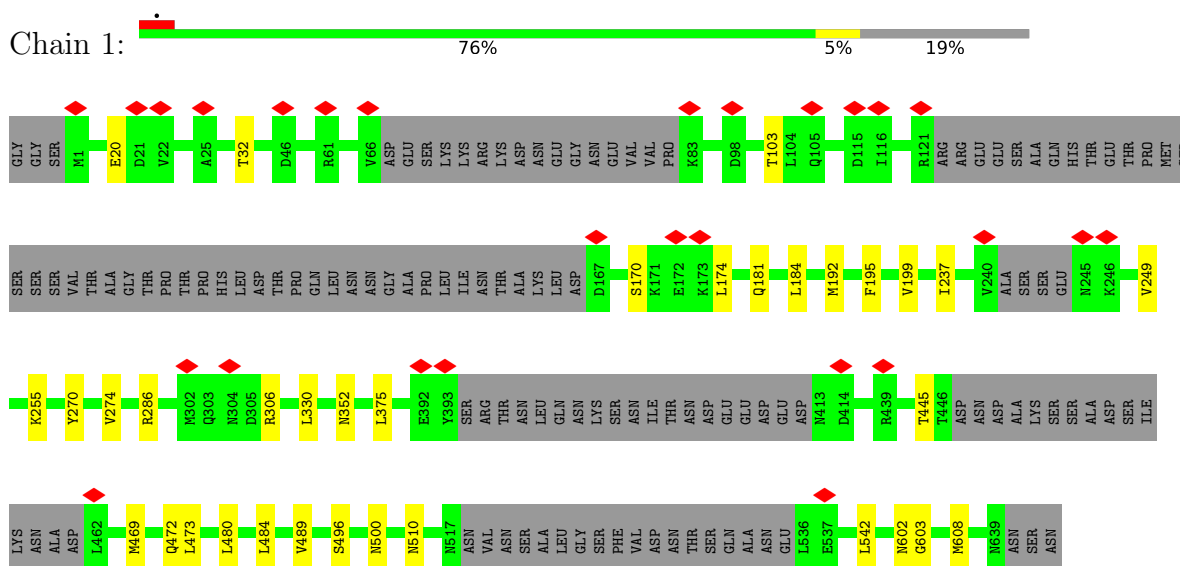
Chain 1:

• Molecule 3: General transcription and DNA repair factor IIH subunit TFB2

Chain 2:

• Molecule 4: RNA polymerase II transcription factor B subunit 3

Chain 3:

• Molecule 5: General transcription and DNA repair factor IIH subunit TFB4

Chain 4:

• Molecule 6: General transcription and DNA repair factor IIH subunit TFB5

Chain 5:

• Molecule 7: General transcription and DNA repair factor IIH subunit SSL1

Chain 6:

• Molecule 8: General transcription and DNA repair factor IIH helicase subunit XPB
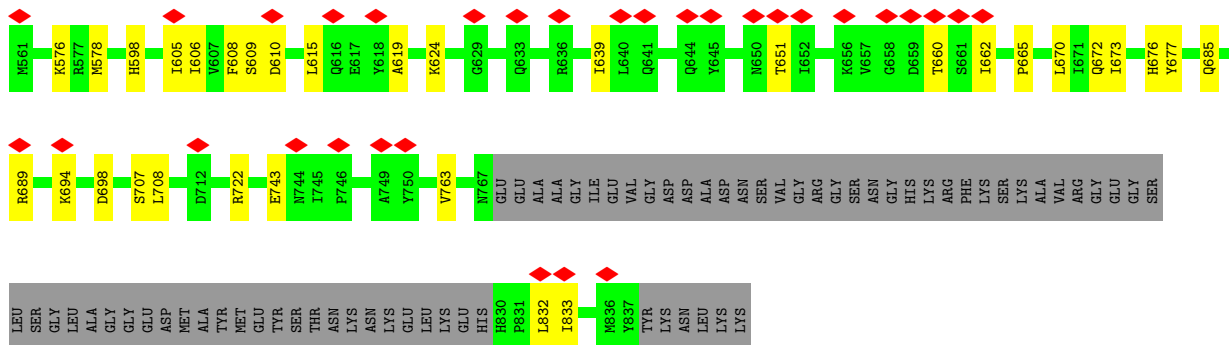
Chain 7:

• Molecule 9: DNA-directed RNA polymerase II subunit RPB1

Chain A:
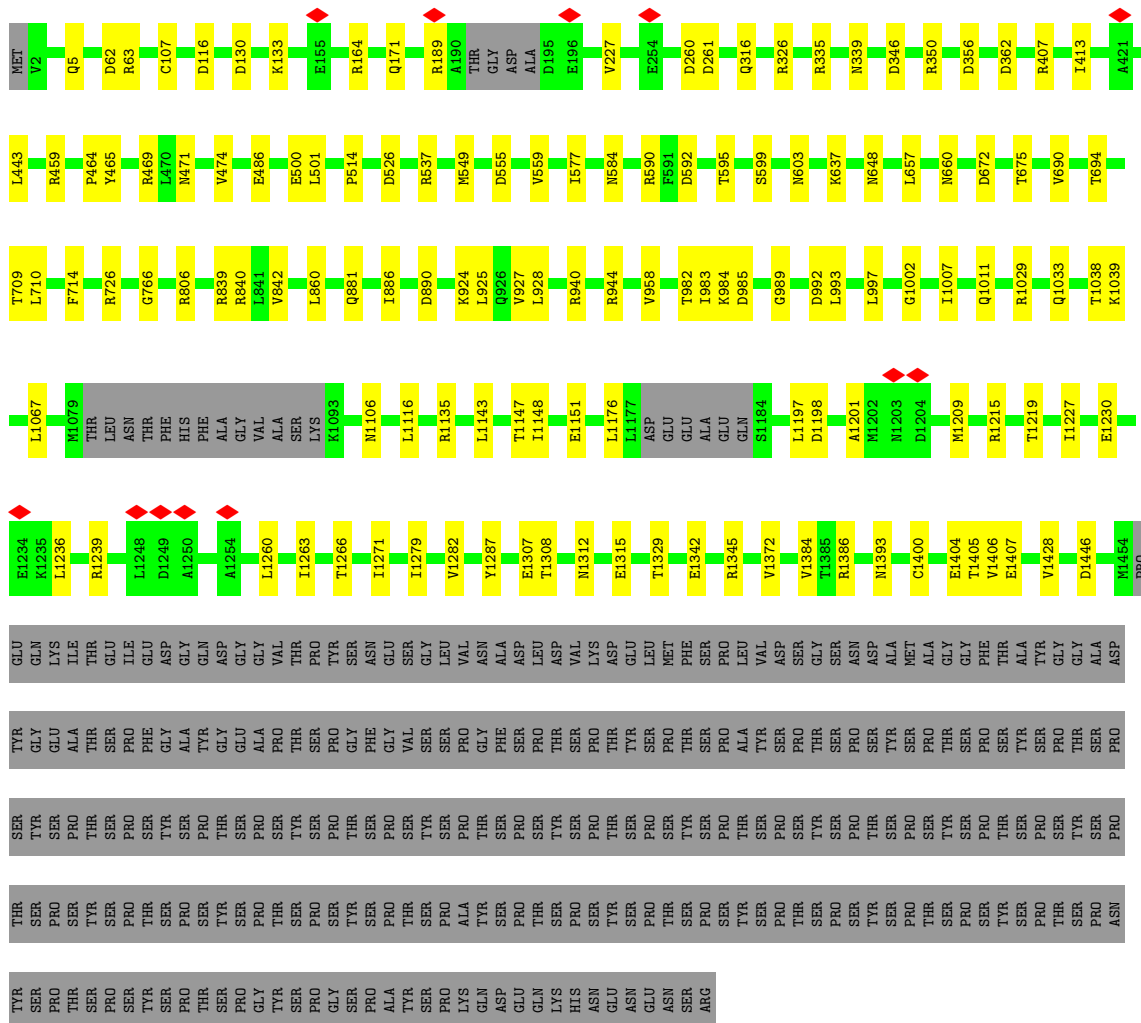
• Molecule 10: DNA-directed RNA polymerase II subunit RPB2

Chain B:
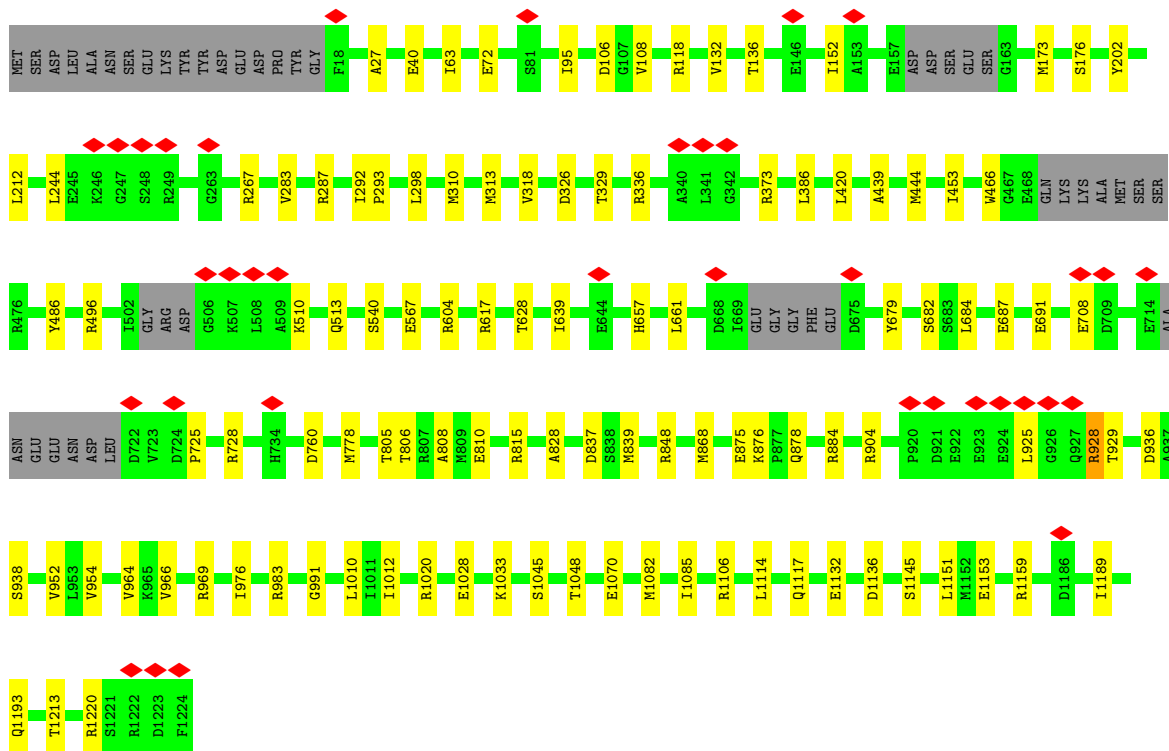
• Molecule 11: DNA-directed RNA polymerase II subunit RPB3

Chain C:

69% 7% 23%
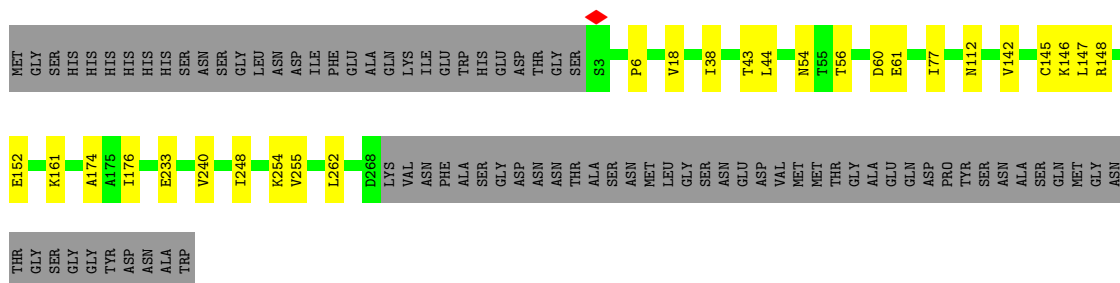
• Molecule 12: DNA-directed RNA polymerase II subunit RPB4

Chain D:

65% 11% 24%
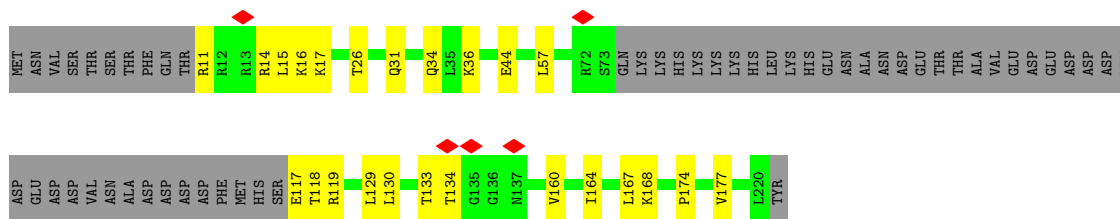
• Molecule 13: DNA-directed RNA polymerases I, II, and III subunit RPABC1

Chain E:

87% 12%
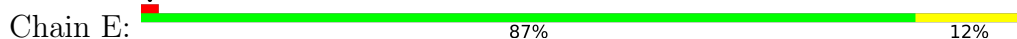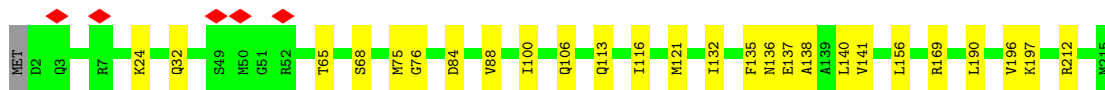
• Molecule 14: DNA-directed RNA polymerases I, II, and III subunit RPABC2

Chain F:
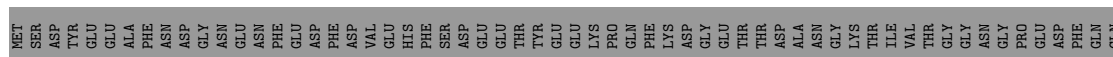
• Molecule 15: DNA-directed RNA polymerase II subunit RPB7

Chain G:
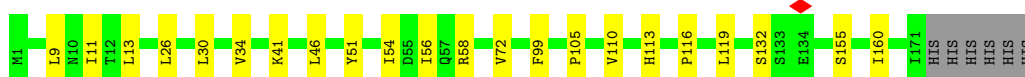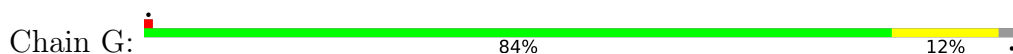
• Molecule 16: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H:
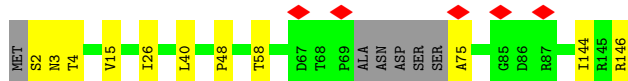
• Molecule 17: DNA-directed RNA polymerase II subunit RPB9
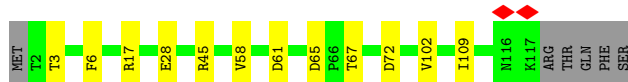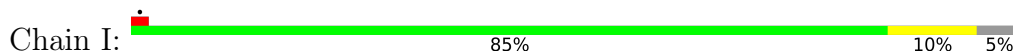
Chain I:

• Molecule 18: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J:

• Molecule 19: DNA-directed RNA polymerase II subunit RPB11
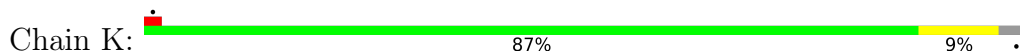
Chain K:

• Molecule 20: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L:

59%   6%   36%

MET SER ARG GLU GLY PHE GLN ILE PRO THR ASN LEU ASP ALA ALA ALA GLY THR SER GLN ALA ARG THR ALA **T26** **L40** **S41** **R42** **L56** **Q66** **R70**

• Molecule 21: Transcription initiation factor IIB

Chain M:

79%   9%   13%

MET MET THR ARG GLU SER ILE ASP LYS ARG ALA GLY ARG **R14** **L22** **E26** **Y30** **E36** **D58** THR ARG SER GLU TRP ARG THR PHE SER ASN ASP ASP HIS ASN GLY ASP PRO SER **R78** **N91** **E100** **M119** **D120** **D123** **D159** **K166** **L177**

R181 **R182** **T200** **D220** **G221** PHE LEU **K224** **I225** **D226** **T227** **D228** **N229** **S244** **P249** **V252** **C263** **I266** **I269** **L288** **I291** **S311** **I315** **L316** **D321** **K322** **L323** **V324** **D325** **P326** **Q327** **L328** **I329** **V334** **V342** GLU LYS LYS HIS HIS HIS HIS HIS

• Molecule 22: Non-template DNA

Chain N:

14%

45%   18%   37%

DC DG DA DG DA DA DC DA DG DT **A11** **T18** **G19** **T20** **G21** **T22** **A23** **T24** **A25** **T26** **A27** **A28** **G31** **T42** **T43** **C49** **A50** **C51** **C52** **T53** **C54** **C55** **G56** **A57** **T58** **G59** **T60** **G61** **T62** **G63** **T64** **T65** **T71** **A72** **C73** **A74** **T75** **A76** **A77** DA DA DA DT DA DT

DC DA DT DA DG DC DA DC DA DA DC DT DG DC DG DC DT DG DT DG DT DC DA
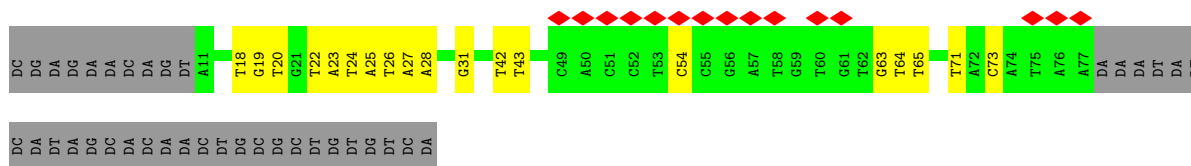
• Molecule 23: TATA-box-binding protein

Chain O:

63%   11%   27%

MET ALA ASP GLU GLU ARG LEU LYS GLU PHE LYS GLU ALA ASN LYS ILE VAL PHE ASP PRO ASN THR ARG GLN VAL TRP GLU ASN GLN ASN ARG ASP GLY THR LYS PRO ALA THR THR PHE GLN SER GLU GLU ASP ILE LYS ARG ALA ALA PRO GLU SER GLU LYS ASP THR SER ALA **T60**

S61 **G62** **I63** **V64** **L67** **V74** **T75** **K83** **L87** **R90** **R107** **E108** **P109** **I115** **I143** **Q144** **A150** **T153** **D154** **I157** **I160** **V166** **I170** **L189** **R196** **I202** **V203** **I223** **L234** **M240** LYS HIS HIS HIS HIS HIS

• Molecule 24: Transcription initiation factor IIF subunit alpha

Chain Q:

8%

26%   .   70%

GLY PRO GLY MET ARG ARG ASN PRO PRO GLY SER ARG ASN GLY GLY GLY PRO THR **N17** **A18** **S19** **I22** **K23** **R24** **D25** **R26** **M27** **R28** **R29** **N30** **F31** **L32** **R33** **M34** **R35** MET GLY GLN ASN GLY SER ASN GLY SER SER SER PRO GLY VAL PRO ASN GLY ASP ASN SER ARG GLY SER LEU

VAL LYS LYS ASP PRO GLU TYR ALA GLU GLU ARG GLU LYS MET LEU LEU GLN ILE GLY VAL GLU ALA ASP ALA GLY ARG SER ASN VAL LYS VAL ASP GLU **D94** **P95** **N96** **E97** **L111** **L118** **R141** **K142** **D143** **T144** **R145** **N146** **L147** **Q148** **F149** **R159** **E162** **I163** **S164**

• Molecule 25: Transcription initiation factor IIF subunit beta

Chain R:

| 8% | 59% | 8% | 33% |
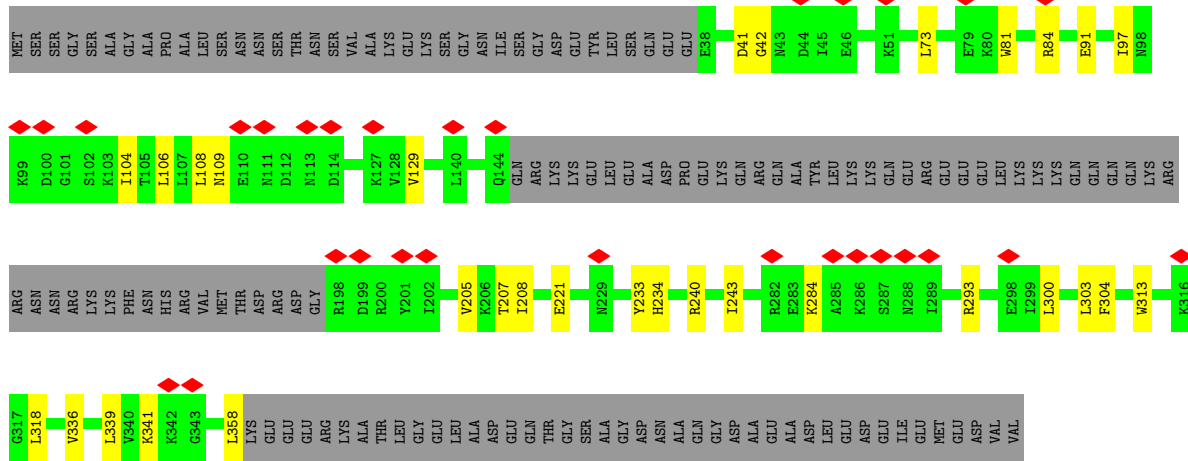
• Molecule 26: Template DNA

Chain T:

| 13% | 48% | 15% | 37% |

• Molecule 27: Transcription initiation factor IIA large subunit,Transcription initiation factor IIA large subunit

Chain U:

• Molecule 28: Transcription initiation factor IIA subunit 2

Chain V:

• Molecule 29: Transcription initiation factor IIE subunit alpha

Chain W:

• Molecule 30: Transcription initiation factor IIE subunit beta

Chain X:

10%

57%   8%   35%
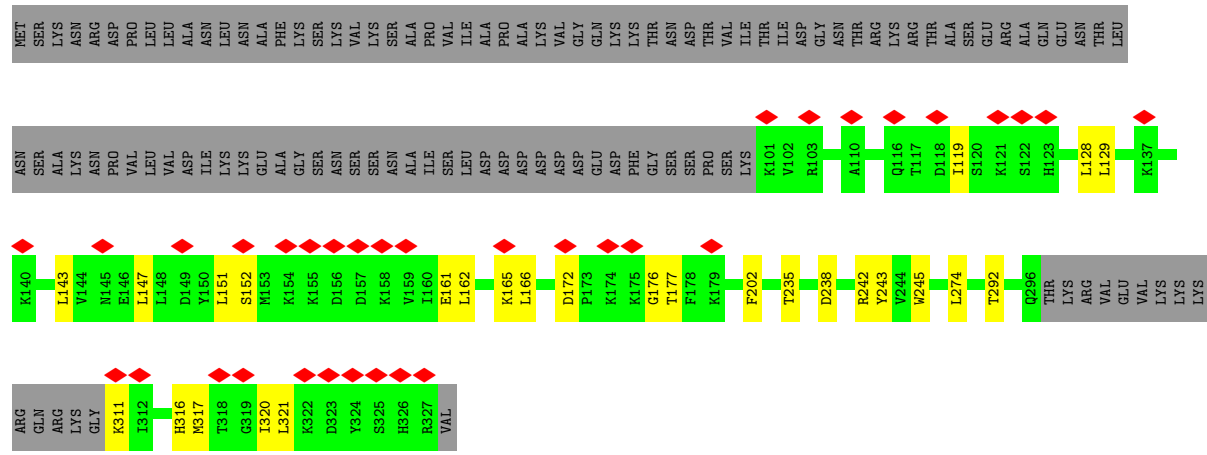
4　Experimental information ⓘ

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	
Number of particles used	16495	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 $(e^-/\text{Å}^2)$	43.6	Depositor
Minimum defocus (nm)	500	Depositor
Maximum defocus (nm)	4000	Depositor
Magnification	130000	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor
Maximum map value	0.233	Depositor
Minimum map value	-0.110	Depositor
Average map value	0.000	Depositor
Map value standard deviation	0.004	Depositor
Recommended contour level	0.011	Depositor
Map size (Å)	398.99997, 398.99997, 398.99997	wwPDB
Map dimensions	380, 380, 380	wwPDB
Map angles (°)	90.0, 90.0, 90.0	wwPDB
Pixel spacing (Å)	1.05, 1.05, 1.05	Depositor

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, SF4, BEF, ADP, 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	0	0.25	0/6209	0.41	0/8384
2	1	0.24	0/4277	0.38	0/5755
3	2	0.24	0/3664	0.39	0/4952
4	3	0.24	0/1109	0.38	0/1492
5	4	0.25	0/2312	0.42	0/3127
6	5	0.23	0/520	0.39	0/701
7	6	0.24	0/2851	0.40	0/3856
8	7	0.24	0/5059	0.41	0/6841
9	A	0.24	0/11455	0.41	0/15490
10	B	0.25	0/9589	0.42	0/12934
11	C	0.24	0/2130	0.41	0/2887
12	D	0.24	0/1351	0.40	0/1811
13	E	0.25	0/1788	0.40	0/2406
14	F	0.24	0/709	0.40	0/956
15	G	0.26	0/1367	0.42	0/1844
16	H	0.24	0/1139	0.43	0/1544
17	I	0.26	0/962	0.44	0/1295
18	J	0.25	0/578	0.39	0/775
19	K	0.25	0/942	0.41	0/1272
20	L	0.24	0/361	0.44	0/478
21	M	0.25	0/2387	0.40	0/3211
22	N	0.57	0/1540	0.98	1/2375 (0.0%)
23	O	0.26	0/1449	0.44	0/1952
24	Q	0.25	0/1907	0.44	0/2556
25	R	0.24	0/2270	0.43	0/3052
26	T	0.58	0/1537	0.94	0/2370
27	U	0.23	0/898	0.41	0/1212
28	V	0.23	0/822	0.41	0/1109
29	W	0.24	0/2575	0.39	0/3473
30	X	0.24	0/1740	0.39	0/2343
All	All	0.27	0/75497	0.45	1/102453 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
22	N	22	DT	N3-C4-O4	5.34	123.10	119.90

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	0	6091	0	6155	65	0
2	1	4214	0	4288	30	0
3	2	3597	0	3687	40	0
4	3	1089	0	1069	9	0
5	4	2274	0	2332	35	0
6	5	514	0	541	2	0
7	6	2794	0	2815	29	0
8	7	4954	0	4946	63	0
9	A	11254	0	11319	91	0
10	B	9404	0	9398	82	0
11	C	2092	0	2050	20	0
12	D	1343	0	1366	15	0
13	E	1752	0	1776	17	0
14	F	697	0	720	6	0
15	G	1339	0	1357	18	0
16	H	1120	0	1086	6	0
17	I	944	0	899	9	0
18	J	569	0	585	5	0
19	K	924	0	934	9	0
20	L	359	0	381	3	0
21	M	2360	0	2467	23	0
22	N	1374	0	759	16	0
23	O	1422	0	1500	19	0
24	Q	1871	0	1883	19	0
25	R	2230	0	2254	24	0
26	T	1370	0	758	13	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
27	U	885	0	866	11	0
28	V	815	0	822	8	0
29	W	2532	0	2529	37	0
30	X	1708	0	1752	19	0
31	0	8	0	0	0	0
32	3	2	0	0	0	0
32	4	1	0	0	0	0
32	6	4	0	0	0	0
32	A	2	0	0	0	0
32	B	1	0	0	0	0
32	C	1	0	0	0	0
32	I	2	0	0	0	0
32	J	1	0	0	0	0
32	L	1	0	0	0	0
32	M	1	0	0	0	0
32	W	1	0	0	0	0
33	7	4	0	0	0	0
34	7	1	0	0	0	0
34	A	1	0	0	0	0
35	7	27	0	12	0	0
All	All	73949	0	73306	645	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 4.

The worst 5 of 645 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)
1:0:438:THR:HG21	1:0:634:ILE:HD12	1.41	1.01
10:B:326:ASP:OD1	10:B:329:THR:OG1	1.89	0.91
29:W:284:LYS:NZ	29:W:288:ASN:OD1	2.04	0.91
22:N:63:DG:OP1	30:X:311:LYS:N	2.06	0.88
8:7:673:ILE:HG22	8:7:708:LEU:HD12	1.55	0.86

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	0	750/778 (96%)	733 (98%)	17 (2%)	0	100	100
2	1	508/645 (79%)	502 (99%)	6 (1%)	0	100	100
3	2	435/517 (84%)	430 (99%)	5 (1%)	0	100	100
4	3	129/324 (40%)	127 (98%)	2 (2%)	0	100	100
5	4	287/341 (84%)	287 (100%)	0	0	100	100
6	5	63/76 (83%)	61 (97%)	2 (3%)	0	100	100
7	6	352/464 (76%)	349 (99%)	3 (1%)	0	100	100
8	7	609/843 (72%)	588 (97%)	21 (3%)	0	100	100
9	A	1422/1733 (82%)	1387 (98%)	35 (2%)	0	100	100
10	B	1168/1224 (95%)	1131 (97%)	37 (3%)	0	100	100
11	C	264/347 (76%)	258 (98%)	6 (2%)	0	100	100
12	D	163/221 (74%)	163 (100%)	0	0	100	100
13	E	212/215 (99%)	207 (98%)	5 (2%)	0	100	100
14	F	84/155 (54%)	82 (98%)	2 (2%)	0	100	100
15	G	169/177 (96%)	166 (98%)	3 (2%)	0	100	100
16	H	136/146 (93%)	133 (98%)	3 (2%)	0	100	100
17	I	114/122 (93%)	106 (93%)	8 (7%)	0	100	100
18	J	67/70 (96%)	66 (98%)	1 (2%)	0	100	100
19	K	113/120 (94%)	112 (99%)	1 (1%)	0	100	100
20	L	43/70 (61%)	42 (98%)	1 (2%)	0	100	100
21	M	302/352 (86%)	294 (97%)	8 (3%)	0	100	100
23	O	179/247 (72%)	172 (96%)	7 (4%)	0	100	100
24	Q	215/738 (29%)	204 (95%)	11 (5%)	0	100	100
25	R	264/400 (66%)	256 (97%)	8 (3%)	0	100	100
27	U	101/171 (59%)	97 (96%)	4 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
28	V	100/129 (78%)	100 (100%)	0	0	100	100
29	W	304/492 (62%)	296 (97%)	8 (3%)	0	100	100
30	X	209/328 (64%)	199 (95%)	10 (5%)	0	100	100
All	All	8762/11445 (77%)	8548 (98%)	214 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	0	684/707 (97%)	684 (100%)	0	100	100
2	1	483/590 (82%)	483 (100%)	0	100	100
3	2	408/470 (87%)	408 (100%)	0	100	100
4	3	125/305 (41%)	125 (100%)	0	100	100
5	4	260/302 (86%)	260 (100%)	0	100	100
6	5	59/68 (87%)	59 (100%)	0	100	100
7	6	323/419 (77%)	322 (100%)	1 (0%)	92	96
8	7	547/737 (74%)	547 (100%)	0	100	100
9	A	1251/1520 (82%)	1251 (100%)	0	100	100
10	B	1024/1061 (96%)	1023 (100%)	1 (0%)	93	98
11	C	234/299 (78%)	234 (100%)	0	100	100
12	D	149/200 (74%)	148 (99%)	1 (1%)	84	94
13	E	196/197 (100%)	196 (100%)	0	100	100
14	F	76/137 (56%)	76 (100%)	0	100	100
15	G	152/158 (96%)	152 (100%)	0	100	100
16	H	123/128 (96%)	123 (100%)	0	100	100
17	I	110/116 (95%)	110 (100%)	0	100	100
18	J	64/65 (98%)	64 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
19	K	99/102 (97%)	99 (100%)	0	100	100
20	L	40/57 (70%)	40 (100%)	0	100	100
21	M	265/306 (87%)	265 (100%)	0	100	100
23	O	153/212 (72%)	153 (100%)	0	100	100
24	Q	204/642 (32%)	204 (100%)	0	100	100
25	R	252/363 (69%)	252 (100%)	0	100	100
27	U	99/154 (64%)	99 (100%)	0	100	100
28	V	94/115 (82%)	94 (100%)	0	100	100
29	W	281/436 (64%)	280 (100%)	1 (0%)	91	95
30	X	191/295 (65%)	191 (100%)	0	100	100
All	All	7946/10161 (78%)	7942 (100%)	4 (0%)	93	98

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	6	450	ASN
10	B	928	ARG
12	D	14	ARG
29	W	116	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
7	6	450	ASN
9	A	451	HIS
9	A	659	HIS
10	B	794	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5　Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6　Ligand geometry ⓘ

Of 22 ligands modelled in this entry, 19 are monoatomic - leaving 3 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).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
33	BEF	7	901	35	0,3,3	-	-	-		
31	SF4	0	801	1	0,12,12	-	-	-		
35	ADP	7	903	34,33	24,29,29	0.67	0	29,45,45	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
31	SF4	0	801	1	-	-	0/6/5/5
35	ADP	7	903	34,33	-	2/12/32/32	0/3/3/3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
35	7	903	ADP	PA-O3A-PB-O3B
35	7	903	ADP	PA-O3A-PB-O2B

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 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 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.

Ligand ADP 7 903

Bond lengths

Bond angles

Torsions

Rings
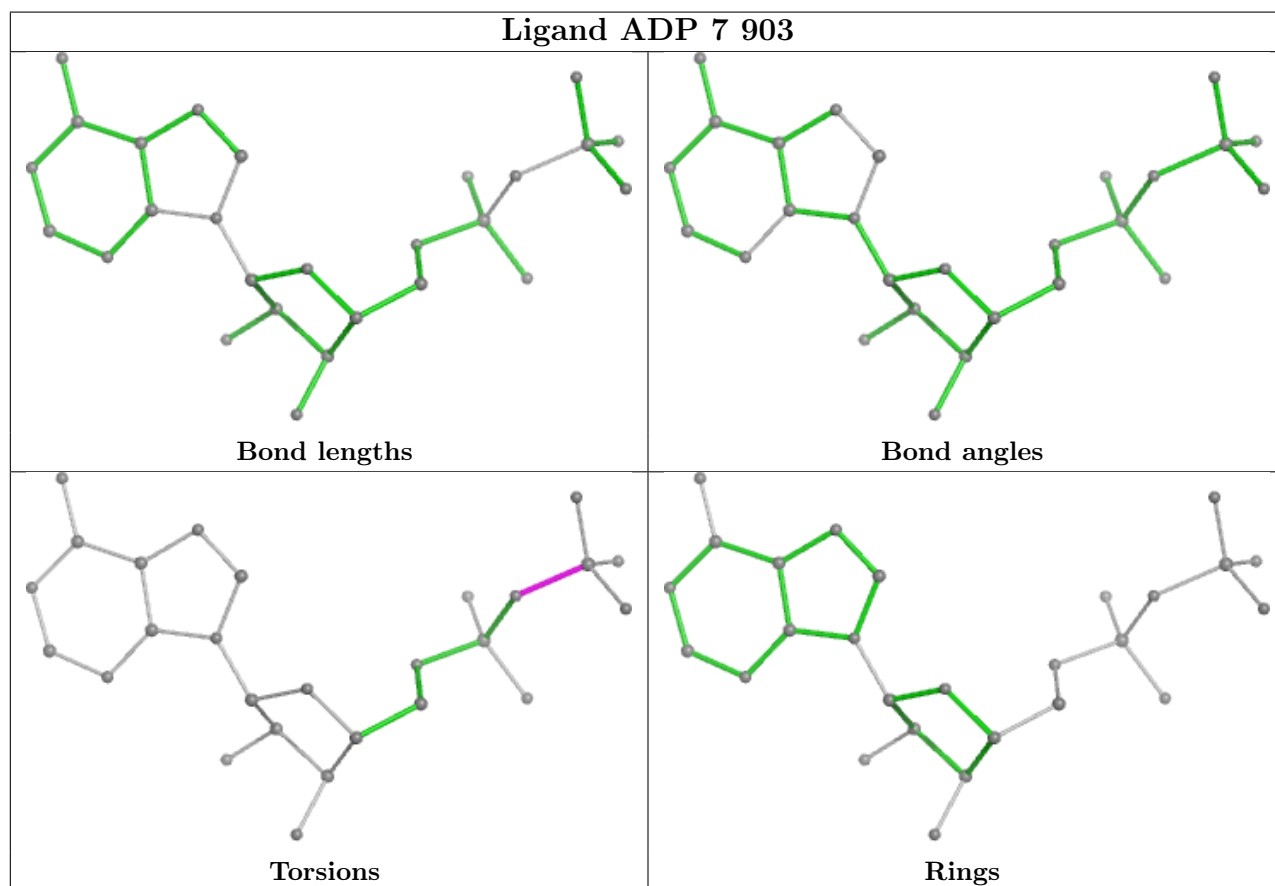
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Map visualisation (i)

This section contains visualisations of the EMDB entry EMD-12719. 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
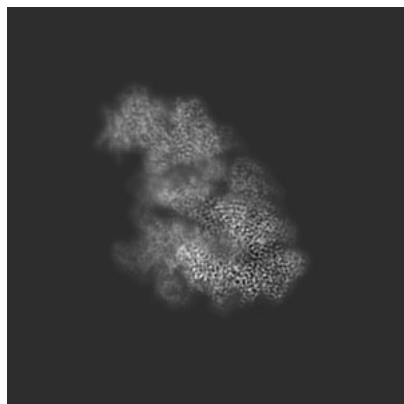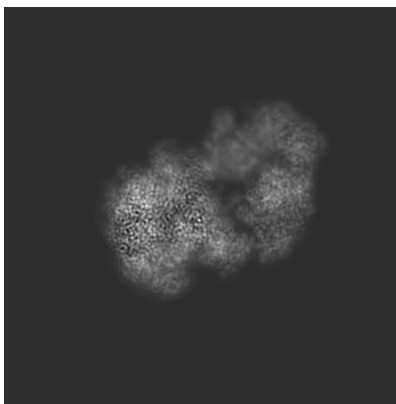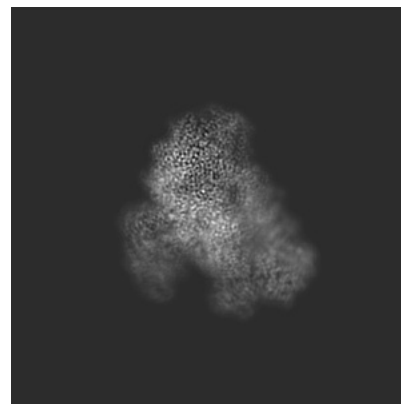
X   Y   Z

6.1.2 Raw map

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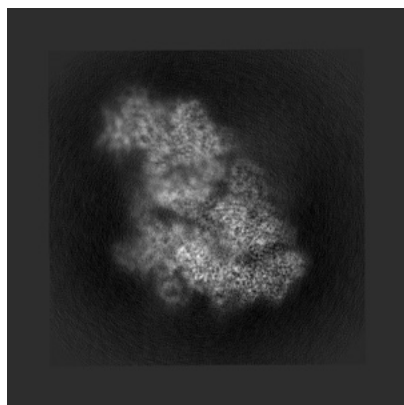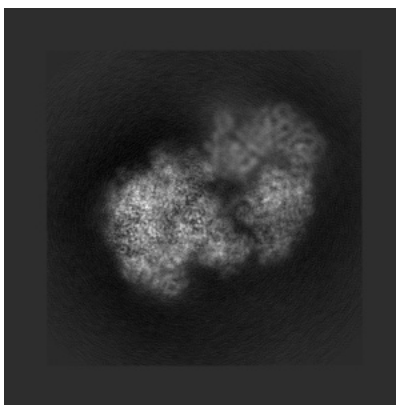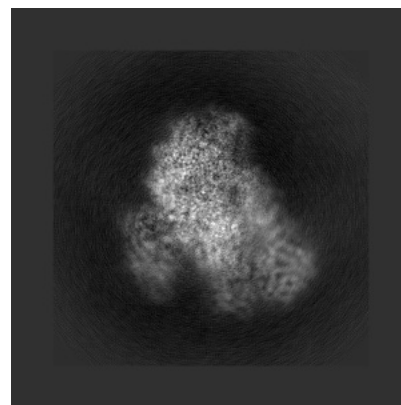
The images above show the map projected in three orthogonal directions.

6.2   Central slices   ⓘ
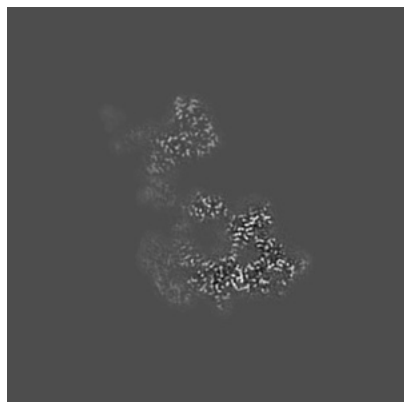
6.2.1   Primary map
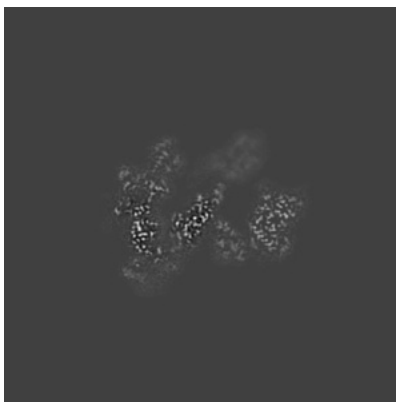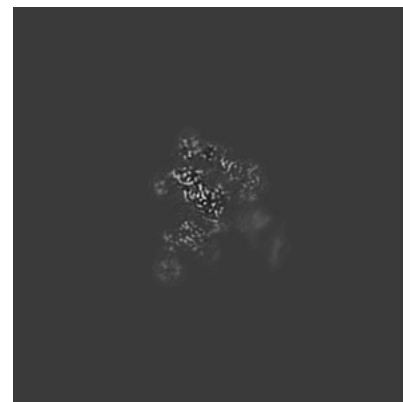
X Index: 190   Y Index: 190   Z Index: 190

6.2.2   Raw map
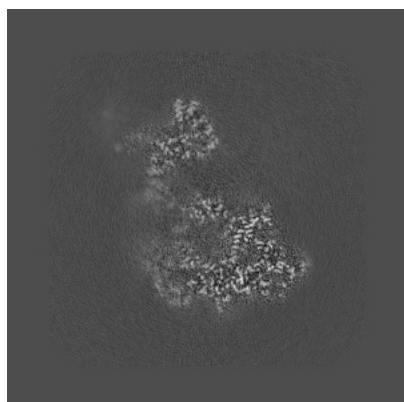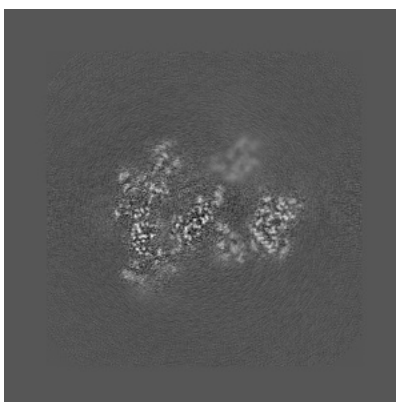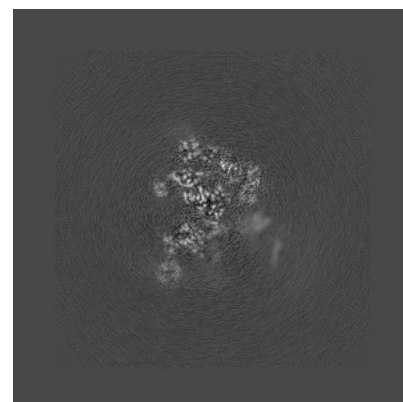
X Index: 190   Y Index: 190   Z Index: 190

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

6.3 Largest variance slices ⓘ

6.3.1 Primary map
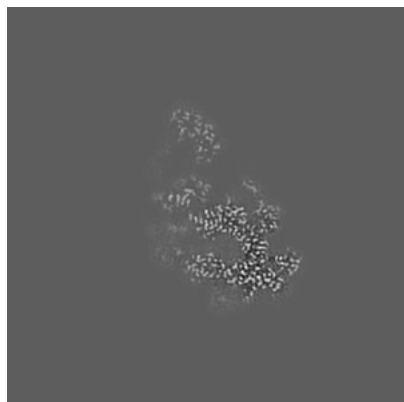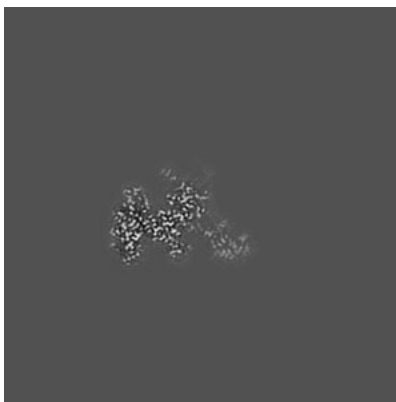
X Index: 172	Y Index: 234	Z Index: 127
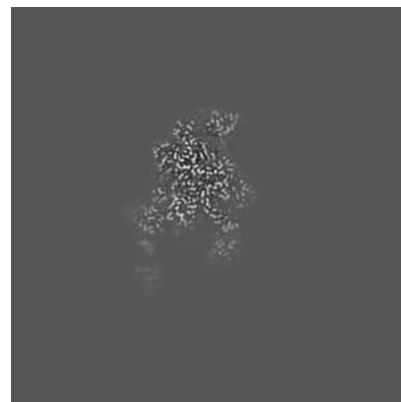
6.3.2 Raw map
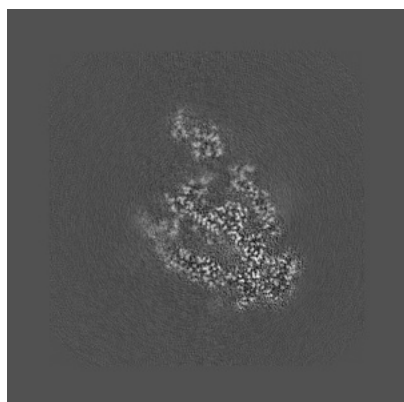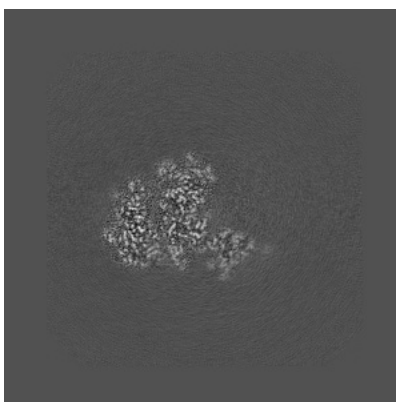
X Index: 163	Y Index: 223	Z Index: 127
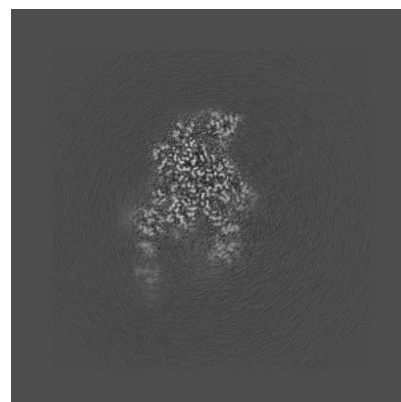
The images above show the largest variance slices of the map in three orthogonal directions.

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

6.4.1 Primary map
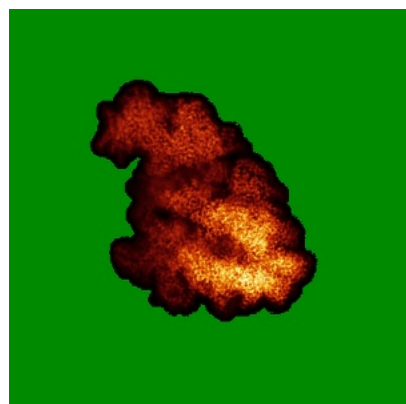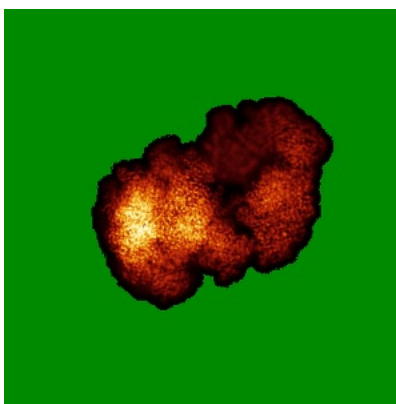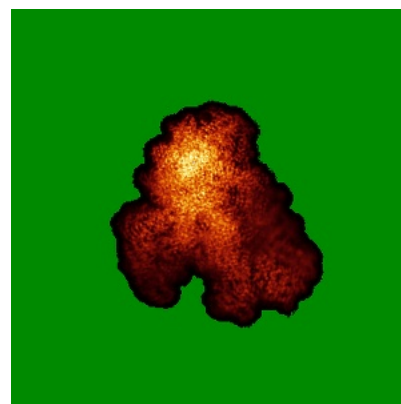
| X | Y | Z |

6.4.2 Raw map

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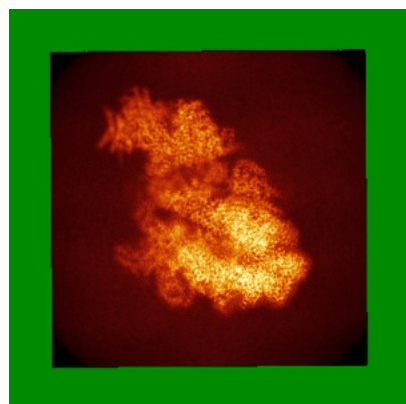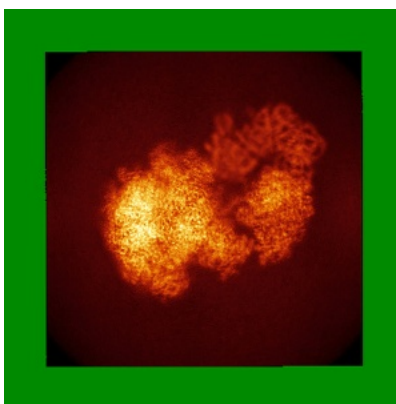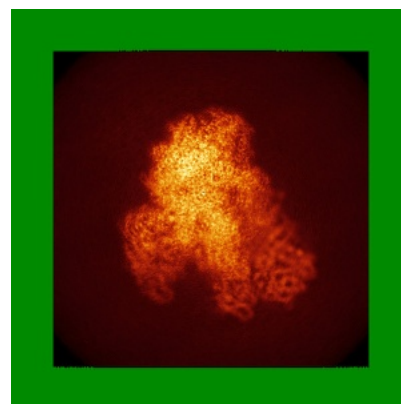
The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

6.5 Orthogonal surface views ⓘ

6.5.1 Primary map

X	Y	Z
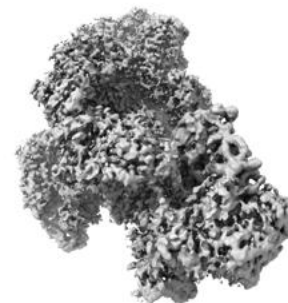
The images above show the 3D surface view of the map at the recommended contour level 0.011. These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

6.5.2 Raw map
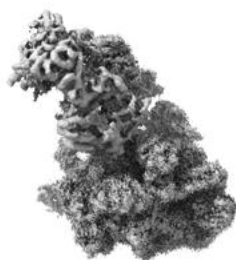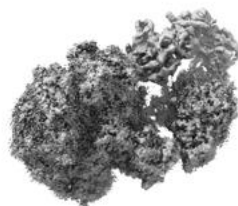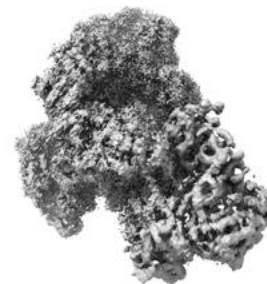
X	Y	Z

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.

6.6 Mask visualisation (i)

This section shows the 3D surface view of the primary map at 50% transparency overlaid with the specified mask at 0% transparency

A mask typically either:

- Encompasses the whole structure

- Separates out a domain, a functional unit, a monomer or an area of interest from a larger structure

6.6.1 emd_12719_msk_1.map (i)

| X | Y | Z |
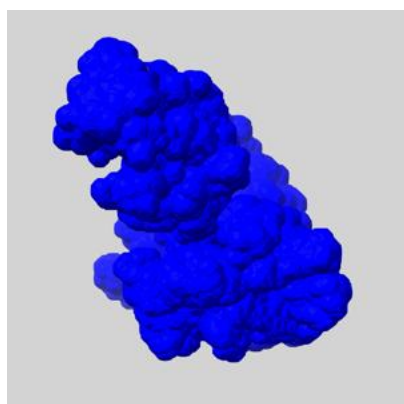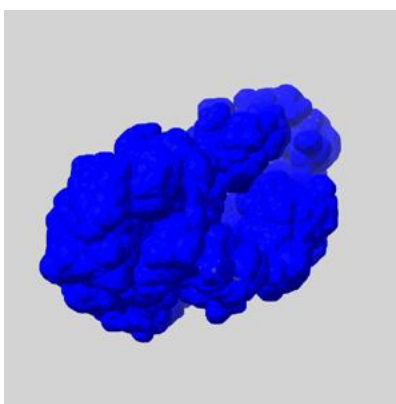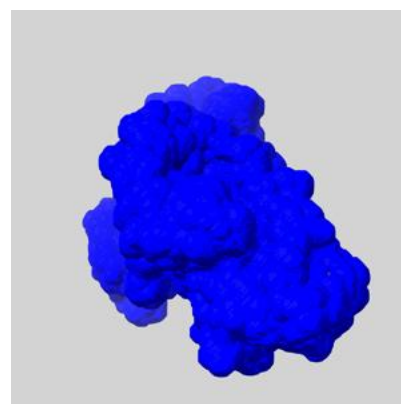
7 Map analysis ⓘ

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

7.1 Map-value distribution ⓘ
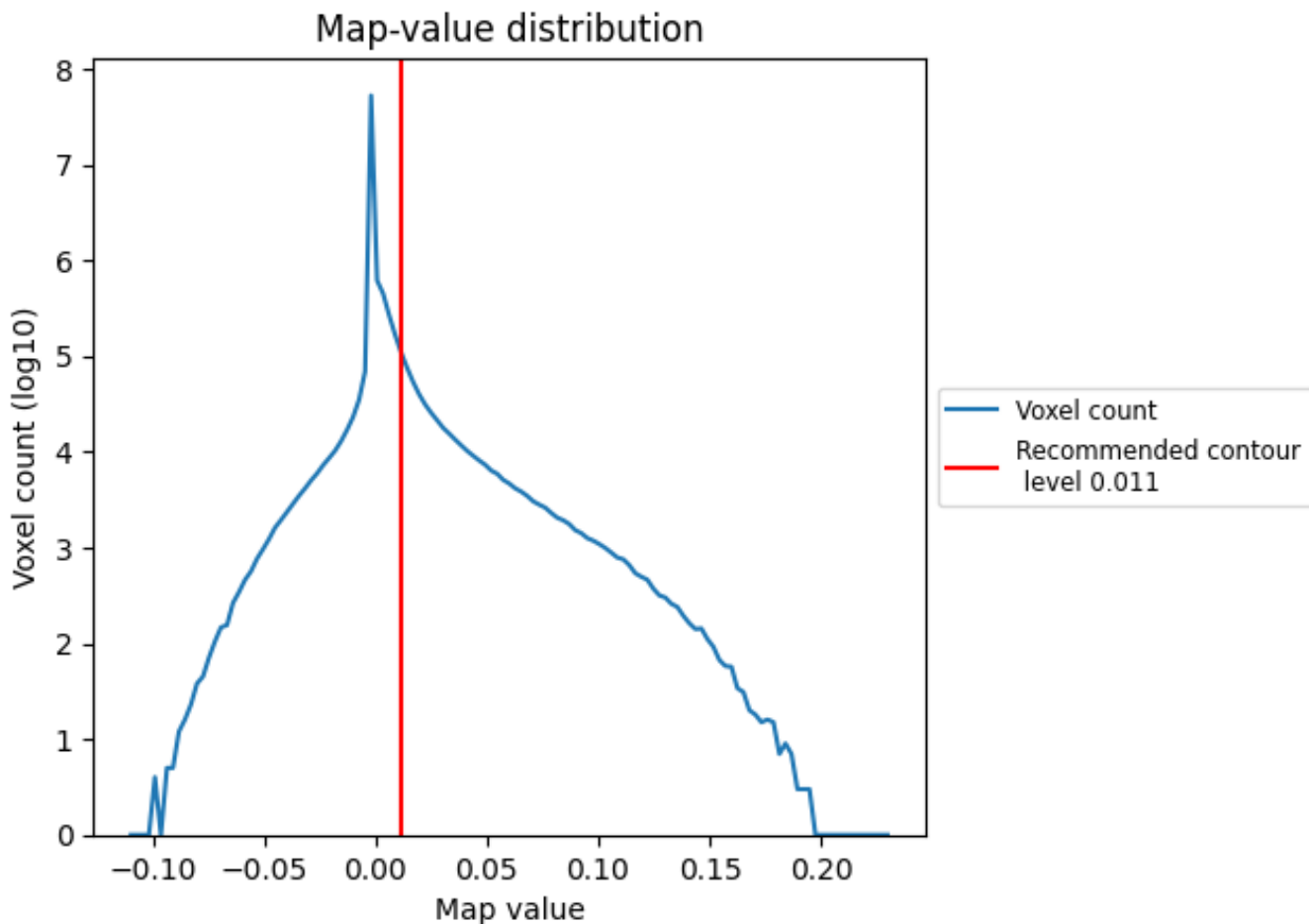
The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

7.2 Volume estimate ⓘ
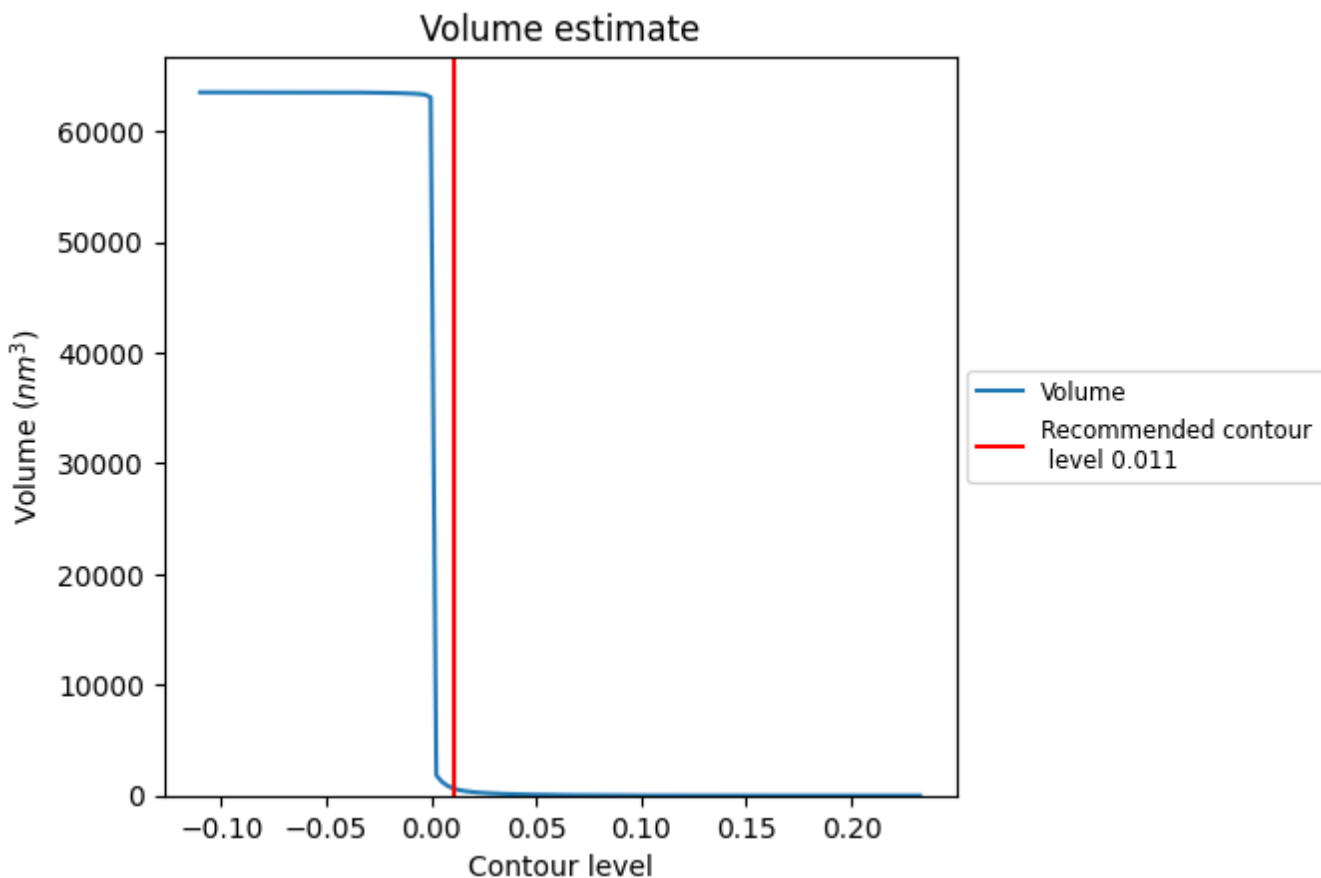
The volume at the recommended contour level is 626 nm^3; this corresponds to an approximate mass of 565 kDa.

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 ⓘ
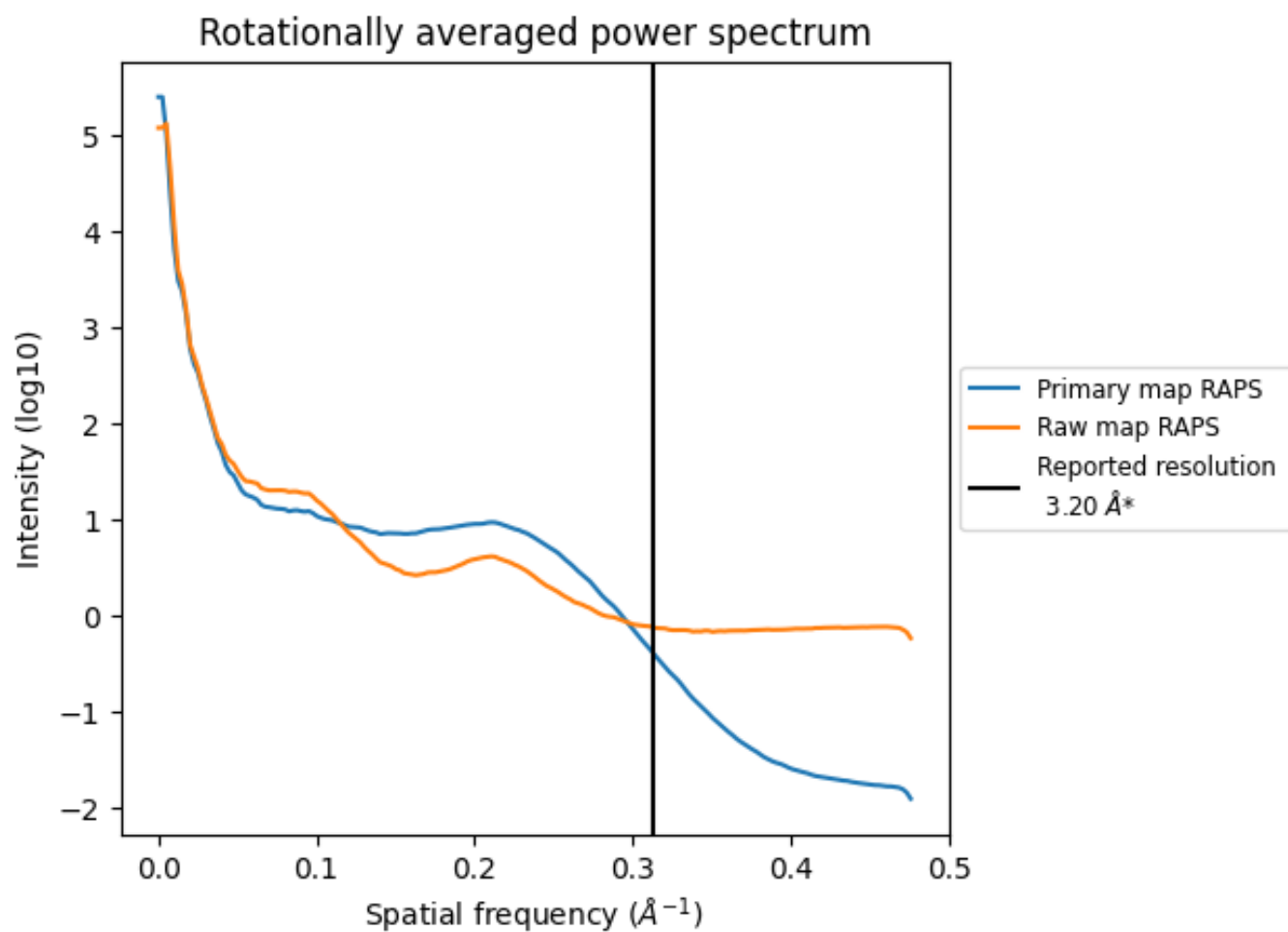
*Reported resolution corresponds to spatial frequency of 0.312 Å^{-1}

8   Fourier-Shell correlation ⓘ

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 ⓘ
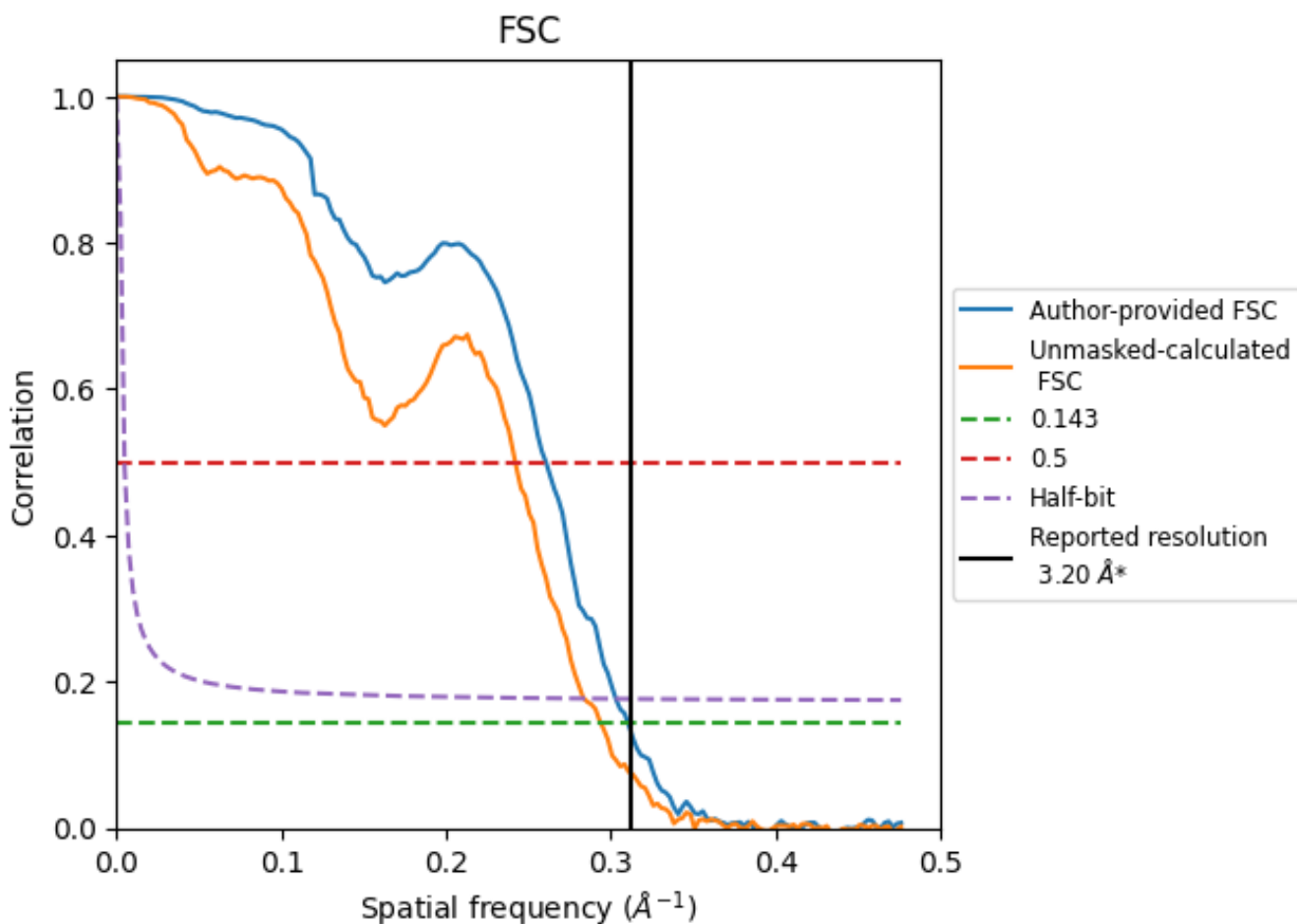
*Reported resolution corresponds to spatial frequency of 0.312 Å^{-1}

8.2   Resolution estimates (i)

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	0.5	Half-bit
Reported by author	3.20	-	-
Author-provided FSC curve	3.22	3.84	3.30
Unmasked-calculated*	3.40	4.13	3.52

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

9 Map-model fit ⓘ

This section contains information regarding the fit between EMDB map EMD-12719 and PDB model 7O4I. Per-residue inclusion information can be found in section 3 on page 15.

9.1 Map-model overlay ⓘ

X   Y   Z
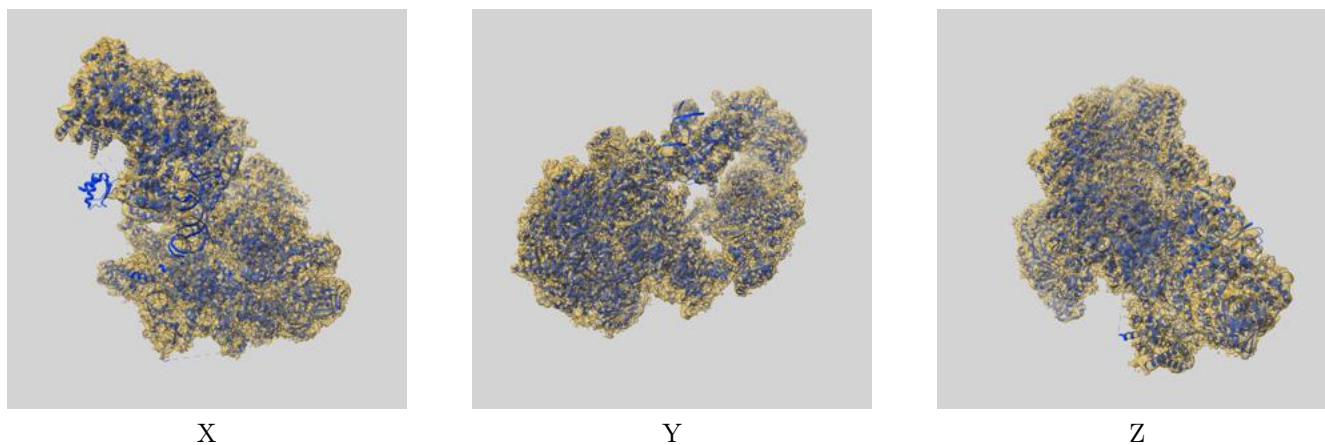
The images above show the 3D surface view of the map at the recommended contour level 0.011 at 50% transparency in yellow overlaid with a ribbon representation of the model coloured in blue. These images allow for the visual assessment of the quality of fit between the atomic model and the map.

9.2   Q-score mapped to coordinate model   ⓘ
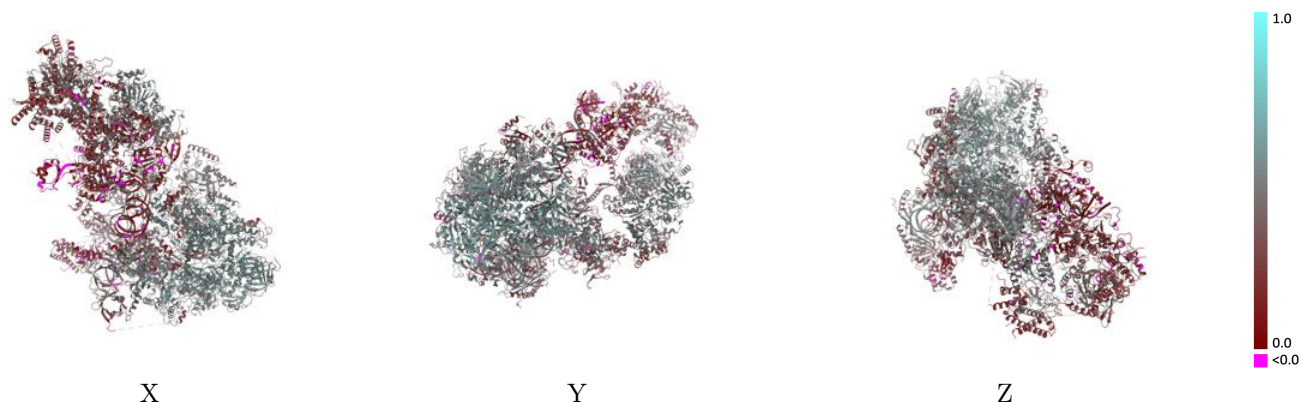
The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions 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   ⓘ
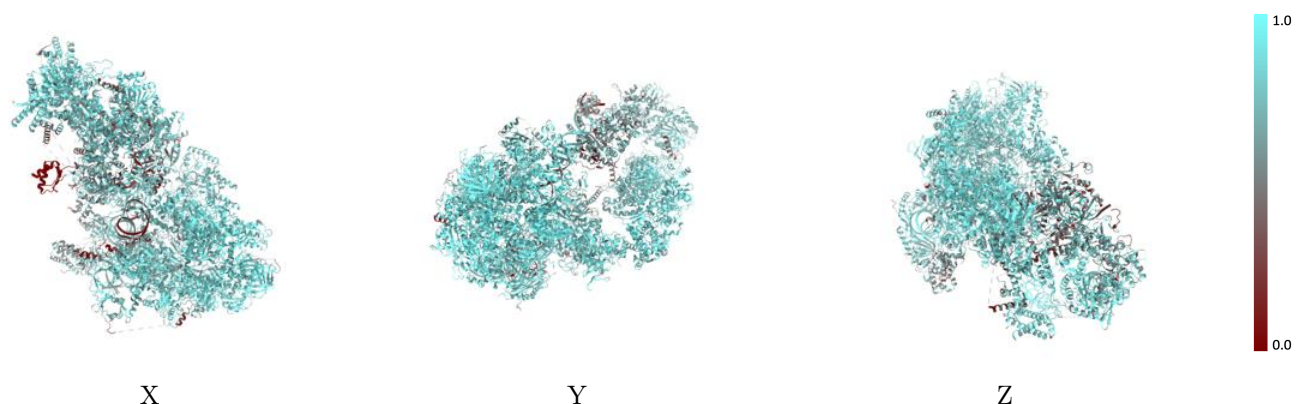
The images above show the model with each residue coloured according to its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (0.011).

9.4　Atom inclusion ⓘ
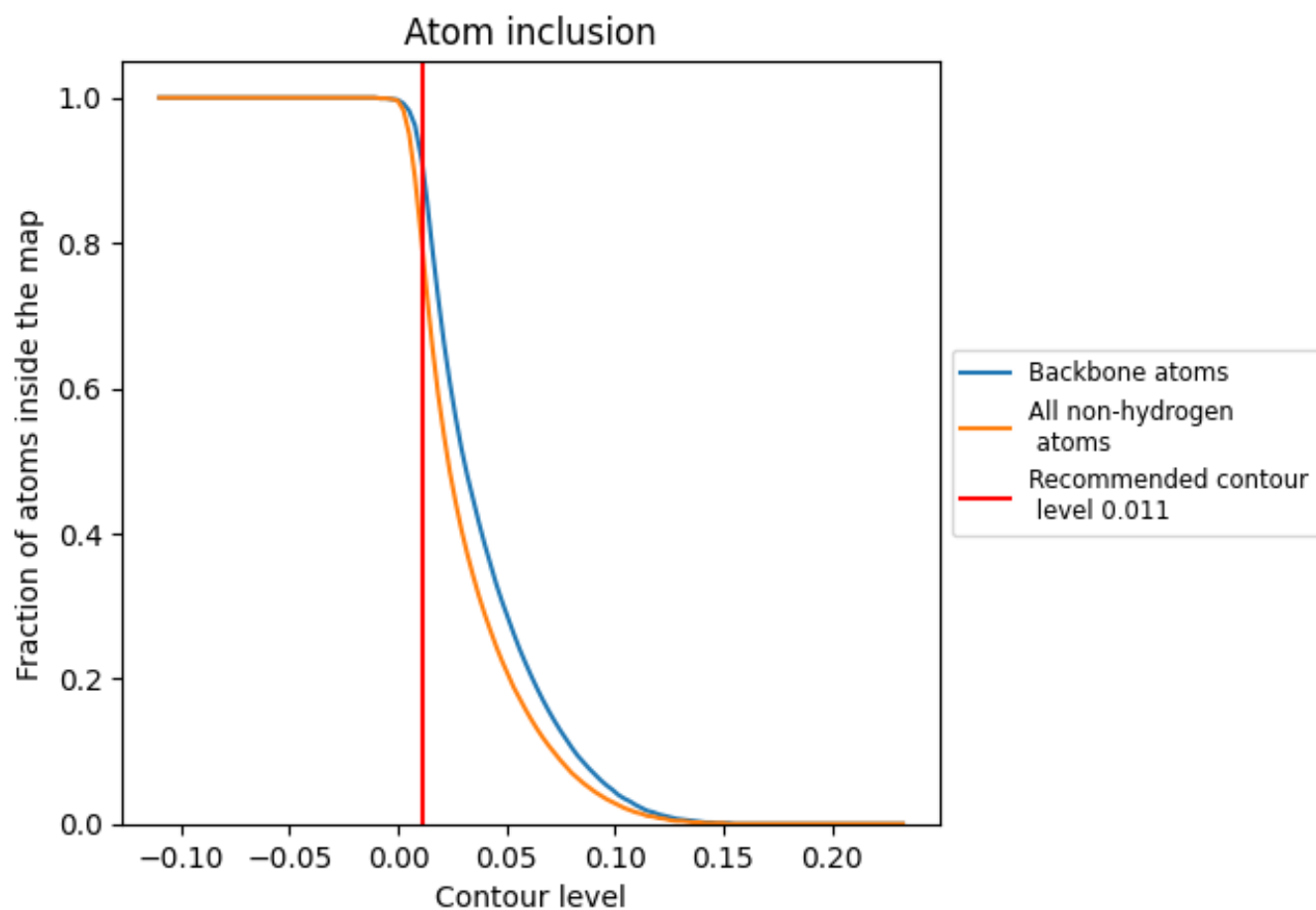
At the recommended contour level, 91% of all backbone atoms, 80% of all non-hydrogen atoms, are inside the map.

9.5   Map-model fit summary ⓘ

The table lists the average atom inclusion at the recommended contour level (0.011) and Q-score for the entire model and for each chain.

Chain	Atom inclusion	Q-score
All	0.8000	0.4000
0	0.8870	0.4730
1	0.7550	0.3260
2	0.6240	0.2160
3	0.8230	0.3730
4	0.8460	0.3630
5	0.6040	0.1700
6	0.8460	0.4220
7	0.6000	0.1730
A	0.8910	0.5020
B	0.8880	0.5180
C	0.9050	0.5360
D	0.7890	0.3740
E	0.8500	0.4410
F	0.8940	0.5380
G	0.8800	0.4660
H	0.8640	0.4830
I	0.8300	0.4260
J	0.8910	0.5490
K	0.9080	0.5390
L	0.9020	0.5190
M	0.8210	0.4420
N	0.6700	0.2170
O	0.8340	0.4000
Q	0.6210	0.3080
R	0.6950	0.3010
T	0.6830	0.2320
U	0.6480	0.2370
V	0.6410	0.2510
W	0.7330	0.3640
X	0.6820	0.3050

1.0

0.0
<0.0