



## wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Sep 5, 2017 – 04:53 AM EDT

PDB ID : 5XON  
EMDB ID: : EMD-6747  
Title : RNA Polymerase II elongation complex bound with Spt4/5 and TFIIS  
Authors : Ehara, H.; Yokoyama, T.; Shigematsu, H.; Shirouzu, M.; Sekine, S.  
Deposited on : unknown  
Resolution : 3.83 Å(reported)

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report  
for a publicly released PDB/EMDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/EMValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20029824

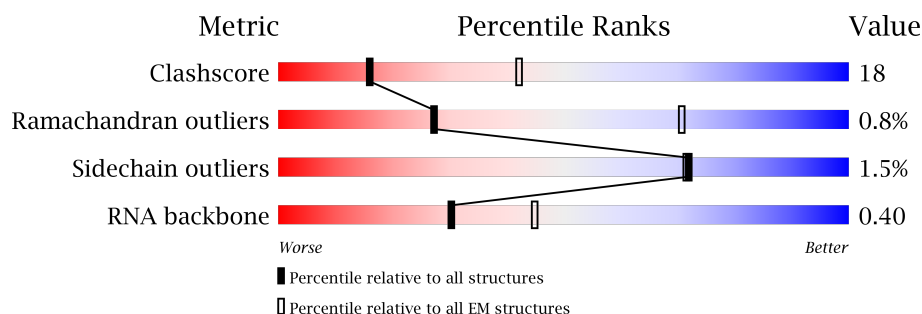
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.83 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.






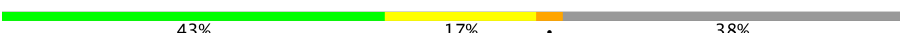

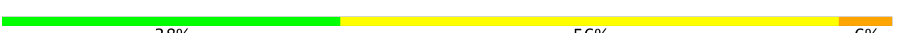
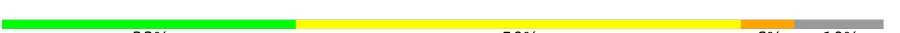



Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026
RNA backbone	3398	335

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	1743	61% 19% • 18%
2	B	1227	75% 19% • 5%
3	C	304	69% 16% • 13%
4	D	186	52% 37% • 10%
5	E	214	78% 20% •
6	F	155	47% 6% • 46%
7	G	171	49% 49% •
8	H	145	74% 17% 8%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
9	I	115	 78%17% . .
10	J	72	 72%19%8%
11	K	118	 82%14% .
12	L	72	 43%17%.38%
13	P	30	 23%10%20%47%
14	T	48	 38%56%6%
15	N	48	 33%50%6%10%
16	U	190	 39%33%8%.18%
17	V	108	 81%12%.6%
18	W	612	 35%16%.46%

## 2 Entry composition

There are 20 unique types of molecules in this entry. The entry contains 38369 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 DNA-directed RNA polymerase subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1427	Total	C	N	O	S	0	0
			11239	7089	1958	2122	70		

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	1161	Total	C	N	O	S	0	0
			9261	5835	1636	1732	58		

- Molecule 3 is a protein called RNA polymerase II third largest subunit B44, part of central core.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	263	Total	C	N	O	S	0	0
			2098	1319	354	413	12		

- Molecule 4 is a protein called RNA polymerase II subunit B32.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	168	Total	C	N	O	S	0	0
			1314	812	237	263	2		

- Molecule 5 is a protein called RNA polymerase subunit ABC27, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	213	Total	C	N	O	S	0	0
			1740	1094	312	324	10		

- Molecule 6 is a protein called RNA polymerase subunit ABC23, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	84	Total	C	N	O	S	0	0
			677	429	114	131	3		

- Molecule 7 is a protein called RNA polymerase II subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	171	Total	C	N	O	S	0	0
			1324	858	214	247	5		

- Molecule 8 is a protein called RNA polymerase subunit ABC14.5, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	133	Total	C	N	O	S	0	0
			1052	671	169	208	4		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	111	Total	C	N	O	S	0	0
			917	565	161	180	11		

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	1	MET	-	see sequence details	UNP C4QY79
I	2	ALA	-	see sequence details	UNP C4QY79
I	3	SER	MET	see sequence details	UNP C4QY79
I	4	PHE	THR	see sequence details	UNP C4QY79
I	5	ARG	ASN	see sequence details	UNP C4QY79
I	6	PHE	VAL	see sequence details	UNP C4QY79
I	7	CYS	ASN	see sequence details	UNP C4QY79
I	8	LEU	SER	see sequence details	UNP C4QY79
I	9	GLU	LEU	see sequence details	UNP C4QY79
I	10	CYS	SER	see sequence details	UNP C4QY79

- Molecule 10 is a protein called RNA polymerase subunit ABC10-beta, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	J	66	Total	C	N	O	S	0	0
			545	349	95	95	6		

- Molecule 11 is a protein called RNA polymerase II subunit B12.5.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	K	113	Total	C	N	O	S	0	0
			932	599	160	169	4		

- Molecule 12 is a protein called RNA polymerase subunit, found in RNA polymerase complexes I, II, and III.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	L	45	Total	C	N	O	S	0	0
			359	221	72	61	5		

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	67	ILE	SER	see sequence details	UNP C4QWA8
L	68	GLN	LYS	see sequence details	UNP C4QWA8
L	70	ASP	LEU	see sequence details	UNP C4QWA8
L	71	ALA	THR	see sequence details	UNP C4QWA8
L	72	ARG	THR	see sequence details	UNP C4QWA8

- Molecule 13 is a RNA chain called RNA (30-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
13	P	16	Total	C	N	O	P	0	0
			338	151	55	116	16		

- Molecule 14 is a DNA chain called DNA (48-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
14	T	48	Total	C	N	O	P	0	0
			975	463	179	285	48		

- Molecule 15 is a DNA chain called DNA (48-MER).

Mol	Chain	Residues	Atoms					AltConf	Trace
15	N	43	Total	C	N	O	P	0	0
			889	422	163	261	43		

- Molecule 16 is a protein called General transcription elongation factor TFIIS.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	U	155	Total	C	N	O	S	0	0
			1239	774	223	232	10		

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
U	96	GLY	-	expression tag	UNP C4R2R6
U	97	PRO	-	expression tag	UNP C4R2R6
U	98	GLY	-	expression tag	UNP C4R2R6
U	266	ALA	ASP	conflict	UNP C4R2R6
U	267	ALA	GLU	conflict	UNP C4R2R6

- Molecule 17 is a protein called Transcription elongation factor SPT4.

Mol	Chain	Residues	Atoms					AltConf	Trace
17	V	102	Total	C	N	O	S	0	0
			792	492	143	150	7		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
V	7	MET	-	expression tag	UNP C4R0E6

- Molecule 18 is a protein called Protein that forms a complex with Spt4p.

Mol	Chain	Residues	Atoms					AltConf	Trace
18	W	329	Total	C	N	O	S	0	0
			2667	1698	483	485	1		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
W	204	GLY	-	expression tag	UNP C4R370
W	205	PRO	-	expression tag	UNP C4R370

- Molecule 19 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
19	J	1	Total	Zn	0
			1	1	
19	B	1	Total	Zn	0
			1	1	

*Continued on next page...*

*Continued from previous page...*

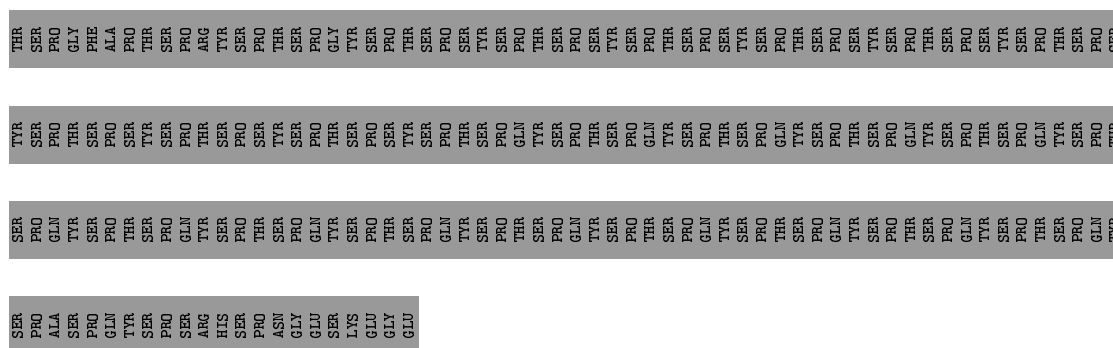
Mol	Chain	Residues	Atoms		AltConf
19	I	2	Total 2	Zn 2	0
19	C	1	Total 1	Zn 1	0
19	V	1	Total 1	Zn 1	0
19	A	2	Total 2	Zn 2	0
19	U	1	Total 1	Zn 1	0
19	L	1	Total 1	Zn 1	0

- Molecule 20 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

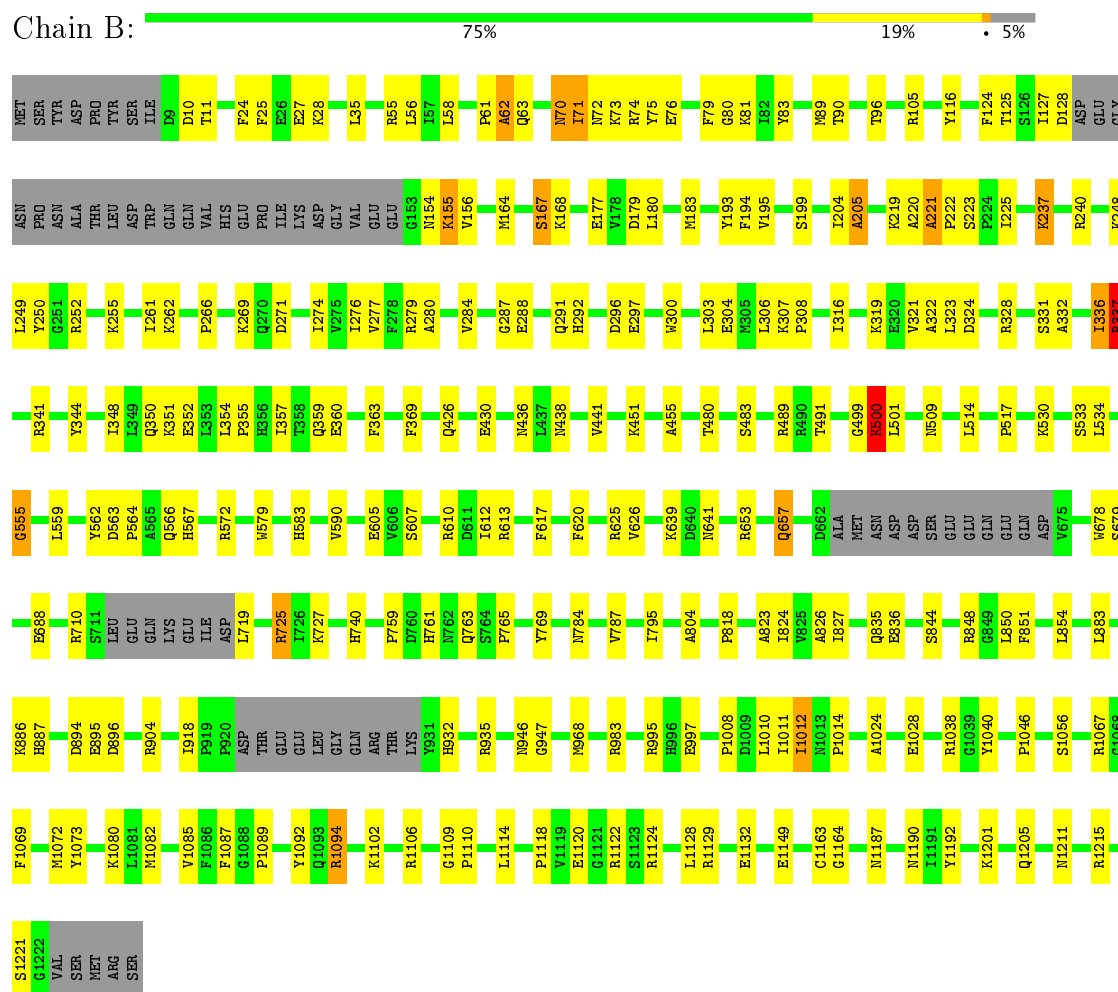
Mol	Chain	Residues	Atoms		AltConf
20	A	1	Total 1	Mg 1	0



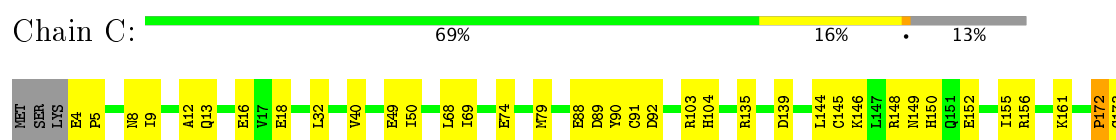




- Molecule 2: DNA-directed RNA polymerase subunit beta

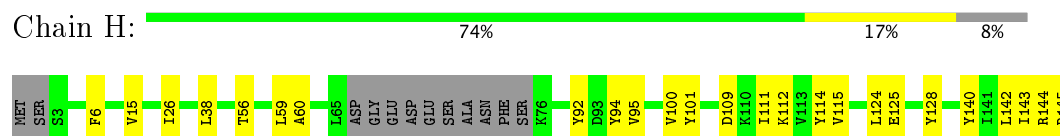


- Molecule 3: RNA polymerase II third largest subunit B44, part of central core

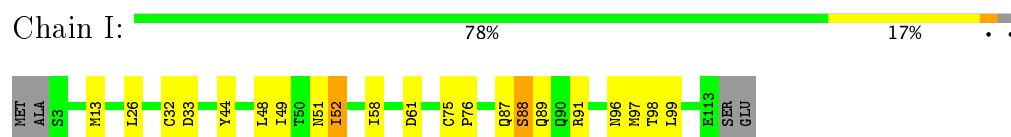




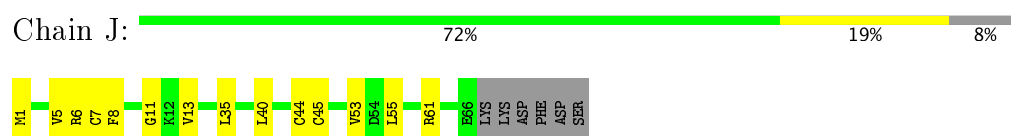
- Molecule 8: RNA polymerase subunit ABC14.5, common to RNA polymerases I, II, and III



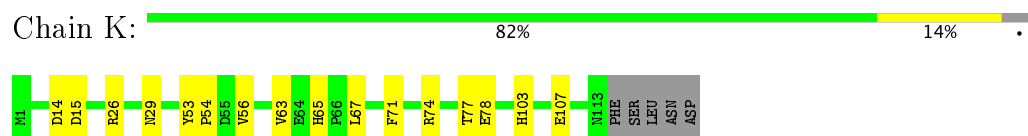
- Molecule 9: DNA-directed RNA polymerase subunit



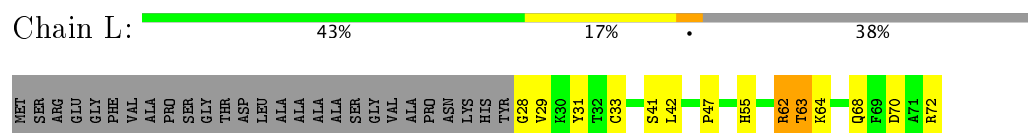
- Molecule 10: RNA polymerase subunit ABC10-beta, common to RNA polymerases I, II, and III



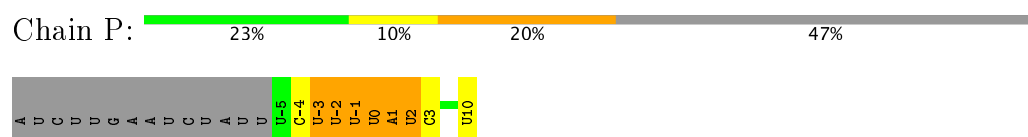
- Molecule 11: RNA polymerase II subunit B12.5



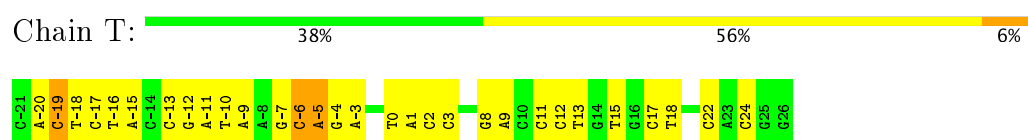
- Molecule 12: RNA polymerase subunit, found in RNA polymerase complexes I, II, and III



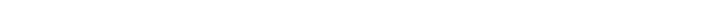
- Molecule 13: RNA (30-MER)



- Molecule 14: DNA (48-MER)



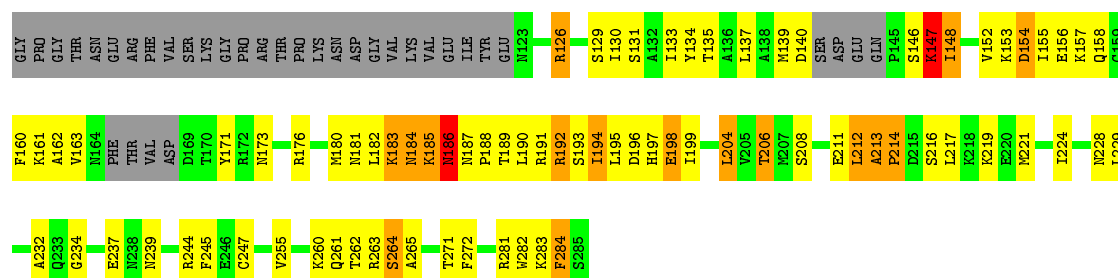
- Molecule 15: DNA (48-MER)

Chain N:  33% 50% 6% 10%

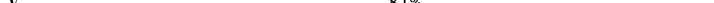


- Molecule 16: General transcription elongation factor TFIIS

Chain U:  39% 33% 8% 18%

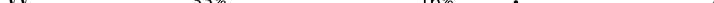


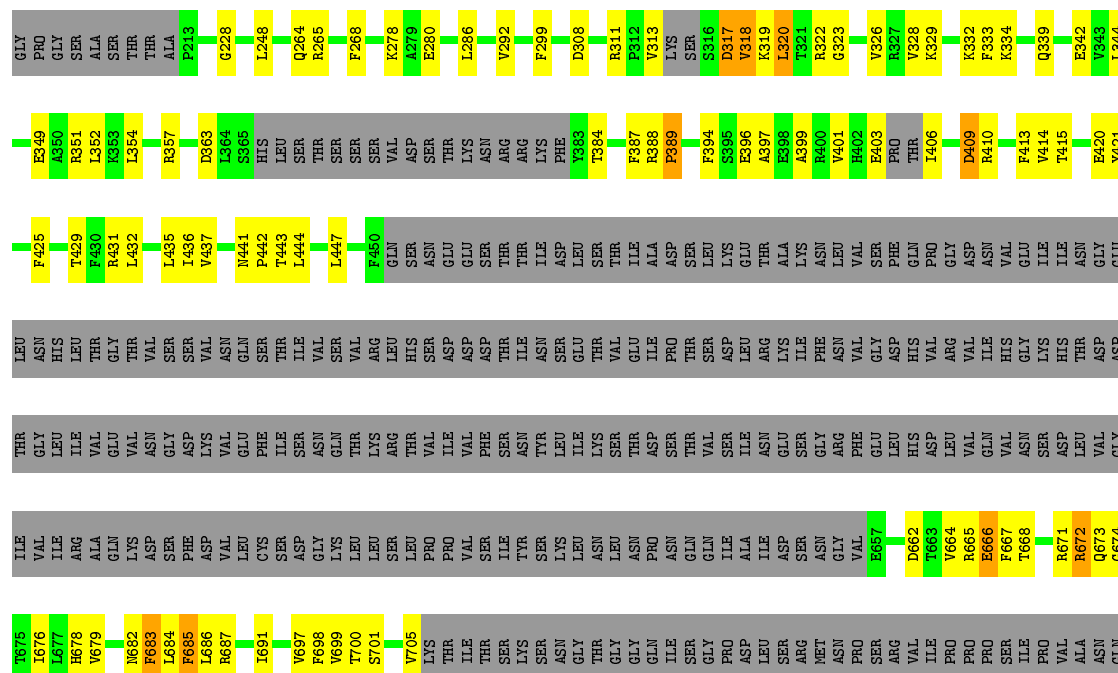
- Molecule 17: Transcription elongation factor SPT4

Chain V:  81% 12% • 6%



- Molecule 18: Protein that forms a complex with Spt4p

Chain W:  35% 16% . 46%



ARG	MET	THR	G747	R748	D749	L752	V756	K757	I758	R759	I767	G775	D776	R777	E781	N784	P785	I786	C793	L796	E799	H802	G803	Y804	Y805	P806	G807	E808	D809	PHE	VAL	ALA	SER	ASP	ARG
-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----

## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	682749	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	NONE	Depositor
Microscope	FEI TECNAI ARCTICA	Depositor
Voltage (kV)	200	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	50	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	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, 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  > 2$	RMSZ	# $ Z  > 2$
1	A	0.26	0/11449	0.69	42/15474 (0.3%)
10	J	0.25	0/554	0.44	0/742
11	K	0.26	0/953	0.44	0/1291
12	L	0.25	0/365	0.70	2/484 (0.4%)
13	P	0.28	0/376	0.85	0/583
14	T	0.99	9/1092 (0.8%)	1.22	3/1680 (0.2%)
15	N	0.89	10/996 (1.0%)	1.21	6/1535 (0.4%)
16	U	0.29	0/1255	0.71	7/1677 (0.4%)
17	V	0.46	0/808	0.60	0/1097
18	W	0.40	0/2713	0.61	5/3646 (0.1%)
2	B	0.26	0/9441	0.62	28/12732 (0.2%)
3	C	0.25	0/2139	0.50	2/2895 (0.1%)
4	D	0.31	0/1326	0.54	3/1788 (0.2%)
5	E	0.25	0/1772	0.56	1/2385 (0.0%)
6	F	0.24	0/687	0.52	1/931 (0.1%)
7	G	0.33	0/1353	0.60	4/1837 (0.2%)
8	H	0.25	0/1069	0.45	0/1444
9	I	0.22	0/934	0.40	0/1257
All	All	0.35	19/39282 (0.0%)	0.67	104/53478 (0.2%)

The worst 5 of 19 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
14	T	-6	DC	C3'-O3'	-6.37	1.35	1.44
15	N	2	DT	C1'-N1	6.00	1.57	1.49
15	N	-24	DC	C1'-N1	5.26	1.56	1.49
15	N	-19	DC	C1'-N1	5.24	1.56	1.49
15	N	-25	DC	C1'-N1	5.24	1.56	1.49

The worst 5 of 104 bond angle outliers are listed below:



Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	310	ALA	CB-CA-C	16.63	135.04	110.10
2	B	221	ALA	N-CA-CB	-14.50	89.80	110.10
2	B	167	SER	CB-CA-C	14.38	137.42	110.10
1	A	47	ARG	C-N-CD	-13.03	91.93	120.60
1	A	861	LEU	CB-CA-C	-12.61	86.24	110.20

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	11239	0	11263	452	0
2	B	9261	0	9267	204	0
3	C	2098	0	2058	32	0
4	D	1314	0	1314	153	0
5	E	1740	0	1754	28	0
6	F	677	0	693	11	0
7	G	1324	0	1342	222	0
8	H	1052	0	1050	15	0
9	I	917	0	865	30	0
10	J	545	0	560	9	0
11	K	932	0	944	9	0
12	L	359	0	360	16	0
13	P	338	0	169	25	0
14	T	975	0	538	43	0
15	N	889	0	487	35	0
16	U	1239	0	1263	264	0
17	V	792	0	757	7	0
18	W	2667	0	2712	250	0
19	A	2	0	0	0	0
19	B	1	0	0	0	0
19	C	1	0	0	0	0
19	I	2	0	0	0	0
19	J	1	0	0	0	0
19	L	1	0	0	0	0
19	U	1	0	0	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
19	V	1	0	0	0	0
20	A	1	0	0	0	0
All	All	38369	0	37396	1359	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 18.

The worst 5 of 1359 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:A:1363:GLY:CA	16:U:282:TRP:HA	1.30	1.62
7:G:100:PHE:HZ	18:W:698:PHE:CE1	1.21	1.54
7:G:100:PHE:CZ	18:W:698:PHE:CE1	1.94	1.53
16:U:133:ILE:HG21	16:U:155:ILE:CG2	1.38	1.49
1:A:1232:GLU:HG3	16:U:180:MET:CB	1.43	1.48

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

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

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

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	1417/1743 (81%)	1235 (87%)	169 (12%)	13 (1%)	20	63
2	B	1151/1227 (94%)	987 (86%)	159 (14%)	5 (0%)	38	77
3	C	261/304 (86%)	231 (88%)	27 (10%)	3 (1%)	17	60
4	D	162/186 (87%)	141 (87%)	20 (12%)	1 (1%)	28	70
5	E	211/214 (99%)	190 (90%)	19 (9%)	2 (1%)	20	63
6	F	82/155 (53%)	70 (85%)	12 (15%)	0	100	100
7	G	169/171 (99%)	150 (89%)	19 (11%)	0	100	100
8	H	129/145 (89%)	111 (86%)	18 (14%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
9	I	109/115 (95%)	93 (85%)	14 (13%)	2 (2%)	10	51
10	J	64/72 (89%)	62 (97%)	2 (3%)	0	100	100
11	K	111/118 (94%)	99 (89%)	12 (11%)	0	100	100
12	L	43/72 (60%)	37 (86%)	6 (14%)	0	100	100
16	U	149/190 (78%)	120 (80%)	24 (16%)	5 (3%)	4	39
17	V	100/108 (93%)	97 (97%)	3 (3%)	0	100	100
18	W	317/612 (52%)	288 (91%)	26 (8%)	3 (1%)	20	63
All	All	4475/5432 (82%)	3911 (87%)	530 (12%)	34 (1%)	27	65

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	47	ARG
1	A	48	PRO
1	A	743	ASN
1	A	1085	THR
2	B	62	ALA

### 5.3.2 Protein sidechains ⓘ

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

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

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	1238/1528 (81%)	1222 (99%)	16 (1%)	73	88
2	B	1016/1077 (94%)	1008 (99%)	8 (1%)	85	93
3	C	236/264 (89%)	235 (100%)	1 (0%)	93	96
4	D	143/160 (89%)	142 (99%)	1 (1%)	87	94
5	E	196/197 (100%)	193 (98%)	3 (2%)	70	87
6	F	75/137 (55%)	75 (100%)	0	100	100
7	G	148/148 (100%)	145 (98%)	3 (2%)	60	83
8	H	120/130 (92%)	120 (100%)	0	100	100
9	I	106/109 (97%)	103 (97%)	3 (3%)	49	76

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
10	J	60/66 (91%)	60 (100%)	0	100	100
11	K	104/109 (95%)	104 (100%)	0	100	100
12	L	38/56 (68%)	37 (97%)	1 (3%)	51	78
16	U	139/171 (81%)	127 (91%)	12 (9%)	12	47
17	V	86/92 (94%)	85 (99%)	1 (1%)	75	88
18	W	290/548 (53%)	279 (96%)	11 (4%)	38	70
All	All	3995/4792 (83%)	3935 (98%)	60 (2%)	72	87

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	E	120	ASN
9	I	49	ILE
18	W	777	ARG
7	G	125	ASN
12	L	62	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	426	GLN
3	C	8	ASN
18	W	434	ASN
2	B	651	HIS
2	B	794	ASN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
13	P	15/30 (50%)	7 (46%)	1 (6%)

5 of 7 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
13	P	-4	C
13	P	-3	U
13	P	-2	U
13	P	0	U

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
13	P	1	A

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
13	P	-1	U

## 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 11 ligands modelled in this entry, 11 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.