



wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Aug 21, 2017 – 03:54 PM EDT

PDB ID : 5M5X
EMDB ID: : EMD-3447
Title : RNA Polymerase I elongation complex 1
Authors : Tafur, L.; Sadian, Y.; Hoffmann, N.A.; Jakobi, A.J.; Wetzels, R.; Hagen, W.J.H.; Sachse, C.; Muller, C.W.
Deposited on : unknown
Resolution : 4.00 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
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

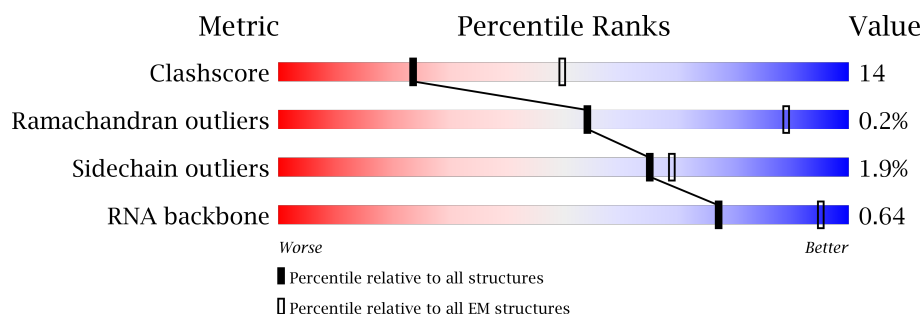
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 4.00 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.








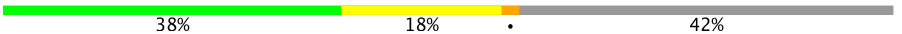
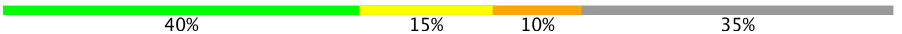
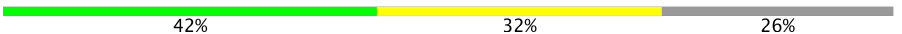
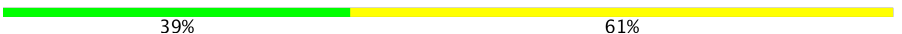
Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026
RNA backbone	3398	335

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 ≥ 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	1664	61% 26% • 13%
2	B	1203	65% 31% • •
3	C	335	66% 24% • 9%
4	D	137	29% 14% 57%
5	E	215	72% 26% •
6	F	155	47% 17% 35%
7	G	326	43% 14% 43%
8	H	146	68% 22% • 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
9	I	125	
10	J	70	
11	K	142	
12	L	70	
13	M	415	
14	N	233	
15	R	20	
16	S	38	
17	T	38	

2 Entry composition

There are 18 unique types of molecules in this entry. The entry contains 34491 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 I subunit RPA190.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	1455	Total	C	N	O	S	0	0
			11488	7259	1998	2169	62		

- Molecule 2 is a protein called DNA-directed RNA polymerase I subunit RPA135.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	1170	Total	C	N	O	S	0	0
			9301	5888	1625	1737	51		

- Molecule 3 is a protein called DNA-directed RNA polymerases I and III subunit RPAC1.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	304	Total	C	N	O	S	0	0
			2418	1536	414	460	8		

- Molecule 4 is a protein called DNA-directed RNA polymerase I subunit RPA14.

Mol	Chain	Residues	Atoms				AltConf	Trace
4	D	59	Total	C	N	O	0	0
			467	293	80	94		

- Molecule 5 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	214	Total	C	N	O	S	0	0
			1751	1111	309	320	11		

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	100	Total	C	N	O	S	0	0
			823	522	144	154	3		

- Molecule 7 is a protein called DNA-directed RNA polymerase I subunit RPA43.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	187	Total	C	N	O	S	0	0
			1486	959	255	267	5		

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	134	Total	C	N	O	S	0	0
			1075	677	182	212	4		

- Molecule 9 is a protein called DNA-directed RNA polymerase I subunit RPA12.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	64	Total	C	N	O	S	0	0
			472	295	78	95	4		

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

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

- Molecule 11 is a protein called DNA-directed RNA polymerases I and III subunit RPAC2.

Mol	Chain	Residues	Atoms					AltConf	Trace
11	K	100	Total	C	N	O	S	0	0
			785	491	129	160	5		

- Molecule 12 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

Mol	Chain	Residues	Atoms					AltConf	Trace
12	L	43	Total	C	N	O	S	0	0
			344	211	69	60	4		

- Molecule 13 is a protein called DNA-directed RNA polymerase I subunit RPA49.

Mol	Chain	Residues	Atoms					AltConf	Trace
13	M	101	Total	C	N	O		0	0
			802	508	132	162			

- Molecule 14 is a protein called DNA-directed RNA polymerase I subunit RPA34.

Mol	Chain	Residues	Atoms					AltConf	Trace
14	N	135	Total	C	N	O	S	0	0
			1070	685	175	206	4		

- Molecule 15 is a RNA chain called RNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
15	R	13	Total	C	N	O	P	0	0
			281	126	56	86	13		

- Molecule 16 is a DNA chain called Non-template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
16	S	28	Total	C	N	O	P	0	0
			579	276	108	167	28		

- Molecule 17 is a DNA chain called Template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
17	T	38	Total	C	N	O	P	0	0
			774	370	137	229	38		

- Molecule 18 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		AltConf
18	B	1	Total	Zn	0
			1	1	
18	A	2	Total	Zn	0
			2	2	
18	L	1	Total	Zn	0
			1	1	
18	J	1	Total	Zn	0
			1	1	
18	I	1	Total	Zn	0
			1	1	

These plots are drawn for all protein, RNA and DNA 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. 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. 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.

- Chain A: 61% 26% 13%

The sequence logo displays the conservation of amino acids at each position. The y-axis represents the information content in bits, with a scale from 0.00 to 0.40. The x-axis represents the position in the sequence, from 1 to 250. The color scale at the top indicates the percentage of positions with a conservation level of 61% (yellow), 26% (orange), and 13% (red).

Key features of the sequence logo include:

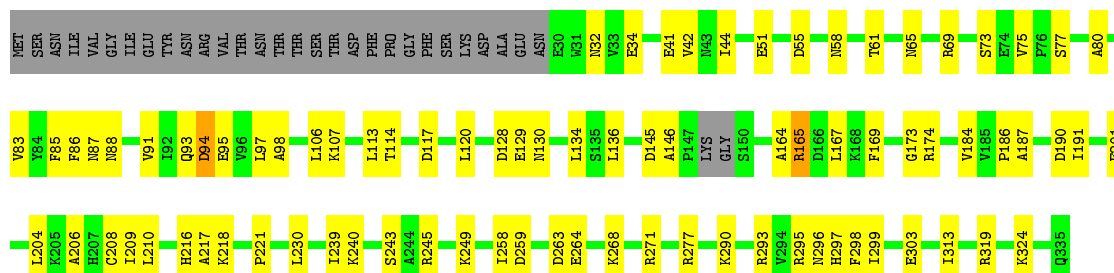
 - High Conservation (61%):** Positions 1-100 show high conservation, with many positions having an information content above 0.20 bits.
 - Medium Conservation (26%):** Positions 101-150 show moderate conservation, with information content generally between 0.10 and 0.20 bits.
 - Low Conservation (13%):** Positions 151-250 show low conservation, with information content generally below 0.10 bits.

The sequence logo is a visual representation of the conservation of amino acids in Chain A. The y-axis represents the information content in bits, and the x-axis represents the position in the sequence. The color scale at the top indicates the percentage of positions with a conservation level of 61% (yellow), 26% (orange), and 13% (red).



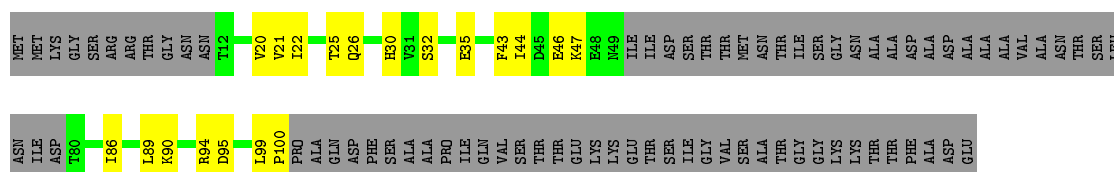
- Molecule 3: DNA-directed RNA polymerases I and III subunit RPAC1

Chain C: 66% 24% 9%



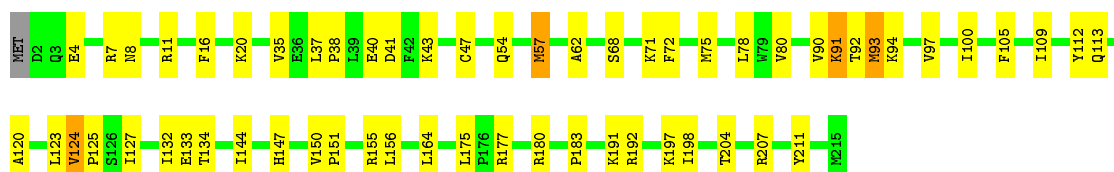
- Molecule 4: DNA-directed RNA polymerase I subunit RPA14

Chain D: 29% 14% 57%



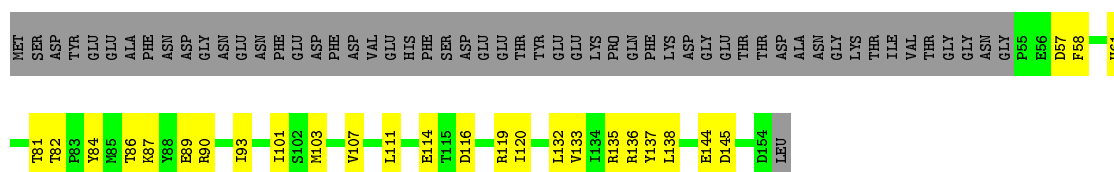
- Molecule 5: DNA-directed RNA polymerases I, II, and III subunit RPABC1

Chain E: 72% 26% 2%



- Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC2

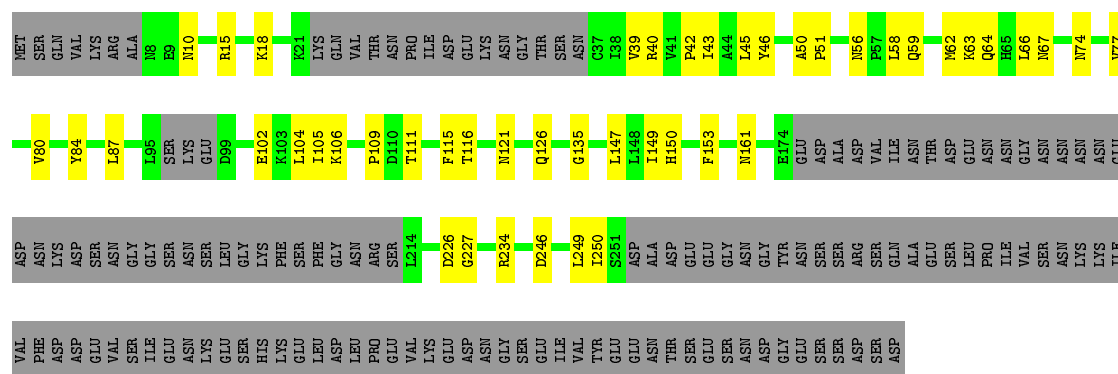
Chain F: 47% 17% 35%



- Molecule 7: DNA-directed RNA polymerase I subunit RPA43

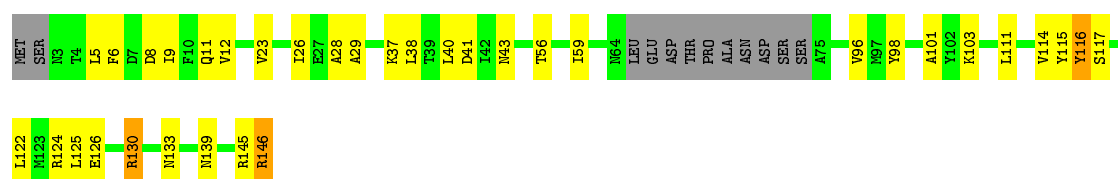
Chain G: 43% 14% 43%





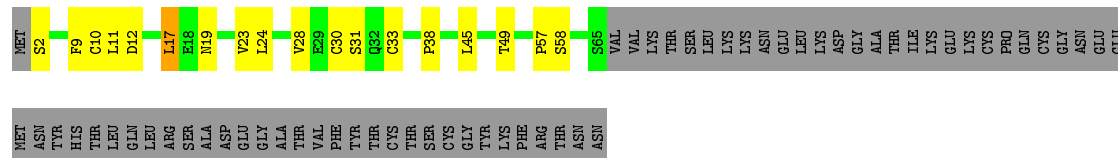
• Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H: 68% 22% 8%



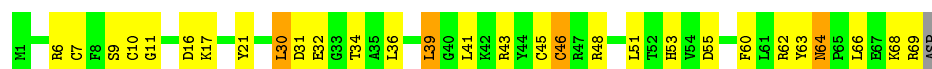
• Molecule 9: DNA-directed RNA polymerase I subunit RPA12

Chain I: 37% 14% 49%



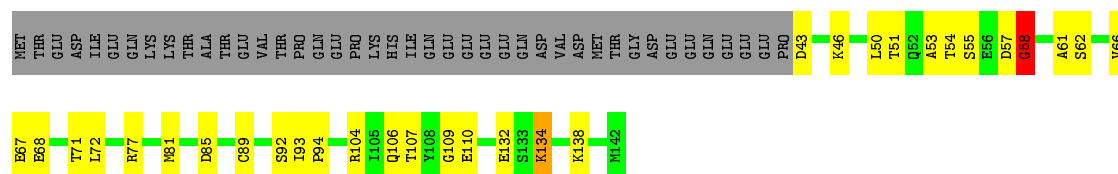
• Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J: 57% 36% 6%



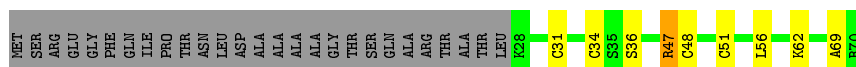
• Molecule 11: DNA-directed RNA polymerases I and III subunit RPAC2

Chain K: 49% 20% 30%



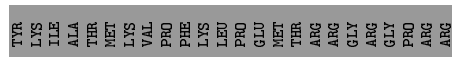
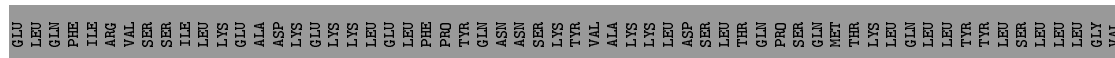
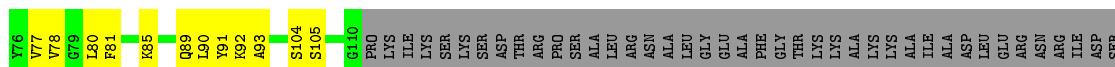
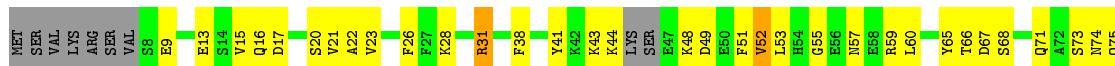
• Molecule 12: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L: 49% 11% 39%

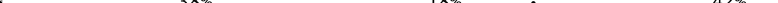


- Molecule 13: DNA-directed RNA polymerase I subunit RPA49

Chain M:  13% 10% 76%



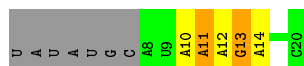
- Molecule 14: DNA-directed RNA polymerase I subunit RPA34

Chain N:  38% 18% 0% 42%



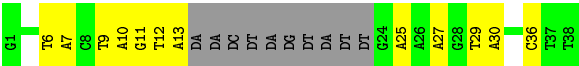
- Molecule 15: RNA

Chain R:



- Molecule 16: Non-template DNA

Chain S: 42% 32% 26%



- Molecule 17: Template DNA

Chain T: 39% 61%



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	83787	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	40	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

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	0.24	1/11698 (0.0%)	0.44	3/15798 (0.0%)
10	J	0.23	0/578	0.43	0/775
11	K	0.54	3/795 (0.4%)	0.81	3/1072 (0.3%)
12	L	0.22	0/346	0.40	0/457
13	M	0.25	0/816	0.47	0/1095
14	N	0.24	0/1090	0.54	1/1466 (0.1%)
15	R	0.12	0/315	0.64	0/489
16	S	0.44	0/649	0.87	0/998
17	T	0.44	0/866	0.89	0/1333
2	B	0.24	1/9506 (0.0%)	0.45	3/12847 (0.0%)
3	C	0.24	0/2469	0.42	0/3347
4	D	0.22	0/473	0.41	0/641
5	E	0.24	0/1787	0.39	0/2406
6	F	0.23	0/838	0.39	0/1129
7	G	0.23	0/1521	0.42	0/2067
8	H	0.24	0/1093	0.47	0/1480
9	I	0.24	0/478	0.46	0/647
All	All	0.26	5/35318 (0.0%)	0.49	10/48047 (0.0%)

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

Mol	Chain	#Chirality outliers	#Planarity outliers
11	K	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	K	57	ASP	N-CA	9.73	1.65	1.46

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	K	58	GLY	CA-C	-8.01	1.39	1.51
1	A	763	GLY	N-CA	6.47	1.55	1.46
11	K	57	ASP	CA-CB	5.88	1.66	1.53
2	B	1064	LYS	N-CA	5.18	1.56	1.46

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
11	K	57	ASP	N-CA-C	-13.91	73.44	111.00
11	K	57	ASP	CB-CA-C	-12.25	85.90	110.40
2	B	1064	LYS	N-CA-C	-12.04	78.50	111.00
1	A	763	GLY	N-CA-C	-8.98	90.65	113.10
11	K	57	ASP	N-CA-CB	-8.30	95.66	110.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
11	K	58	GLY	Mainchain

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	11488	0	11571	377	0
2	B	9301	0	9194	310	0
3	C	2418	0	2401	64	0
4	D	467	0	468	16	0
5	E	1751	0	1776	38	0
6	F	823	0	841	24	0
7	G	1486	0	1488	37	0
8	H	1075	0	1046	25	0
9	I	472	0	473	13	0
10	J	569	0	585	23	0
11	K	785	0	782	28	0
12	L	344	0	363	7	0
13	M	802	0	787	37	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
14	N	1070	0	1085	43	0
15	R	281	0	143	3	0
16	S	579	0	318	9	0
17	T	774	0	430	23	0
18	A	2	0	0	0	0
18	B	1	0	0	0	0
18	I	1	0	0	0	0
18	J	1	0	0	0	0
18	L	1	0	0	0	0
All	All	34491	0	33751	940	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 14.

The worst 5 of 940 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:250:LYS:NZ	1:A:314:TYR:HE2	1.28	1.29
1:A:250:LYS:CE	1:A:314:TYR:HE2	1.49	1.25
1:A:250:LYS:HG3	1:A:252:PHE:CE1	1.77	1.18
1:A:250:LYS:NZ	1:A:314:TYR:CE2	2.04	1.17
1:A:250:LYS:CG	1:A:252:PHE:CE1	2.29	1.14

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	1437/1664 (86%)	1345 (94%)	85 (6%)	7 (0%)	32	73
2	B	1160/1203 (96%)	1098 (95%)	62 (5%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	C	300/335 (90%)	293 (98%)	7 (2%)	0	100	100
4	D	55/137 (40%)	55 (100%)	0	0	100	100
5	E	212/215 (99%)	206 (97%)	6 (3%)	0	100	100
6	F	98/155 (63%)	97 (99%)	1 (1%)	0	100	100
7	G	179/326 (55%)	179 (100%)	0	0	100	100
8	H	130/146 (89%)	125 (96%)	5 (4%)	0	100	100
9	I	62/125 (50%)	57 (92%)	5 (8%)	0	100	100
10	J	67/70 (96%)	60 (90%)	7 (10%)	0	100	100
11	K	98/142 (69%)	93 (95%)	5 (5%)	0	100	100
12	L	41/70 (59%)	40 (98%)	1 (2%)	0	100	100
13	M	97/415 (23%)	86 (89%)	11 (11%)	0	100	100
14	N	127/233 (54%)	101 (80%)	25 (20%)	1 (1%)	22	66
All	All	4063/5236 (78%)	3835 (94%)	220 (5%)	8 (0%)	54	84

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	250	LYS
1	A	227	LEU
1	A	248	PHE
1	A	251	ILE
14	N	148	ILE

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	1284/1465 (88%)	1263 (98%)	21 (2%)	68	86
2	B	1025/1053 (97%)	1009 (98%)	16 (2%)	68	86
3	C	269/296 (91%)	266 (99%)	3 (1%)	78	89
4	D	56/116 (48%)	56 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
5	E	196/197 (100%)	190 (97%)	6 (3%)	45	74
6	F	90/137 (66%)	90 (100%)	0	100	100
7	G	166/291 (57%)	164 (99%)	2 (1%)	75	89
8	H	117/128 (91%)	114 (97%)	3 (3%)	51	78
9	I	56/110 (51%)	54 (96%)	2 (4%)	40	72
10	J	64/65 (98%)	59 (92%)	5 (8%)	15	50
11	K	90/130 (69%)	89 (99%)	1 (1%)	78	89
12	L	38/57 (67%)	37 (97%)	1 (3%)	51	78
13	M	91/371 (24%)	87 (96%)	4 (4%)	33	67
14	N	125/220 (57%)	120 (96%)	5 (4%)	36	69
All	All	3667/4636 (79%)	3598 (98%)	69 (2%)	65	83

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	1063	ARG
3	C	165	ARG
13	M	53	LEU
2	B	1084	THR
2	B	1134	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 42 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	492	ASN
2	B	824	HIS
9	I	32	GLN
2	B	547	HIS
2	B	682	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
15	R	12/20 (60%)	3 (25%)	2 (16%)

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
15	R	10	A
15	R	11	A
15	R	13	G

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
15	R	10	A
15	R	12	A

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 6 ligands modelled in this entry, 6 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.