WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 21, 2023 – 03:06 AM JST

PDB ID	:	7EH1
Title	:	Thermus thermophilus transcription initiation complex containing a template-strand purine at position TSS-2, GpG RNA primer, and CMPcPP
Authors	:	Li, L.; Zhang, Y.
Deposited on	:	2021-03-27
Resolution	:	2.90 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1   Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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.

Percentile relative to all X-ray structures

Percentile relative to X-ray structures of similar resolution
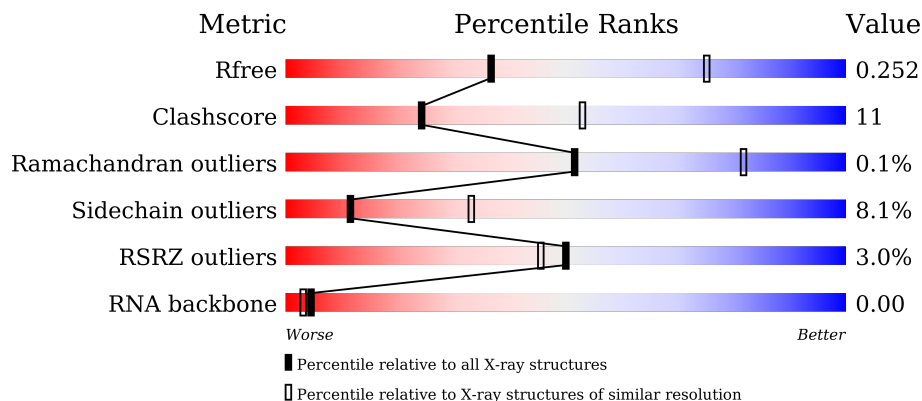
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)
RNA backbone	3102	1007 (3.16-2.64)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower 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 electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	315	2% 50% 21% • 27%
1	B	315	% 48% 23% • 28%
2	C	1119	2% 68% 29% ••
3	D	1524	4% 68% 27% ••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	68% 23% • 5%
5	F	443	% 56% 21% • 22%
6	G	27	44% 44% • 7%
7	H	19	5% 58% 26% 5% 11%
8	I	2	100%

2 Entry composition ⓘ

There are 13 unique types of molecules in this entry. The entry contains 28630 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, 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 alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	231	Total	C	N	O	S	0	0	0
			1803	1152	312	337	2			
1	B	227	Total	C	N	O	S	0	0	0
			1758	1124	301	331	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1112	Total	C	N	O	S	0	0	0
			8750	5538	1553	1635	24			

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1486	Total	C	N	O	S	0	1	0
			11714	7428	2061	2190	35			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	94	Total	C	N	O	S	0	0	0
			759	484	132	139	4			

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	346	Total	C	N	O	S	0	0	0
			2803	1767	508	524	4			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-19	MET	-	expression tag	UNP Q5SKW1
F	-18	GLY	-	expression tag	UNP Q5SKW1
F	-17	SER	-	expression tag	UNP Q5SKW1
F	-16	SER	-	expression tag	UNP Q5SKW1
F	-15	HIS	-	expression tag	UNP Q5SKW1
F	-14	HIS	-	expression tag	UNP Q5SKW1
F	-13	HIS	-	expression tag	UNP Q5SKW1
F	-12	HIS	-	expression tag	UNP Q5SKW1
F	-11	HIS	-	expression tag	UNP Q5SKW1
F	-10	HIS	-	expression tag	UNP Q5SKW1
F	-9	SER	-	expression tag	UNP Q5SKW1
F	-8	SER	-	expression tag	UNP Q5SKW1
F	-7	GLY	-	expression tag	UNP Q5SKW1
F	-6	LEU	-	expression tag	UNP Q5SKW1
F	-5	VAL	-	expression tag	UNP Q5SKW1
F	-4	PRO	-	expression tag	UNP Q5SKW1
F	-3	ARG	-	expression tag	UNP Q5SKW1
F	-2	GLY	-	expression tag	UNP Q5SKW1
F	-1	SER	-	expression tag	UNP Q5SKW1
F	0	HIS	-	expression tag	UNP Q5SKW1

- Molecule 6 is a DNA chain called DNA (27-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	G	25	516	246	99	147	24	0	0	0

- Molecule 7 is a DNA chain called DNA (5'-D(*CP*C*TP*GP*CP*AP*TP*CP*CP*GP*T P*GP*AP*GP*CP*CP*AP*AP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
7	H	17	346	165	66	99	16	0	0	0

- Molecule 8 is a RNA chain called RNA (5'-R(*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
8	I	2	43	20	10	12	1	0	0	0

- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	B	2	Total 2	Mg 2	0	0
9	C	1	Total 1	Mg 1	0	0
9	D	1	Total 1	Mg 1	0	0
9	F	1	Total 1	Mg 1	0	0

- Molecule 10 is 1,4-BUTANEDIOL (three-letter code: BU1) (formula: $C_4H_{10}O_2$).
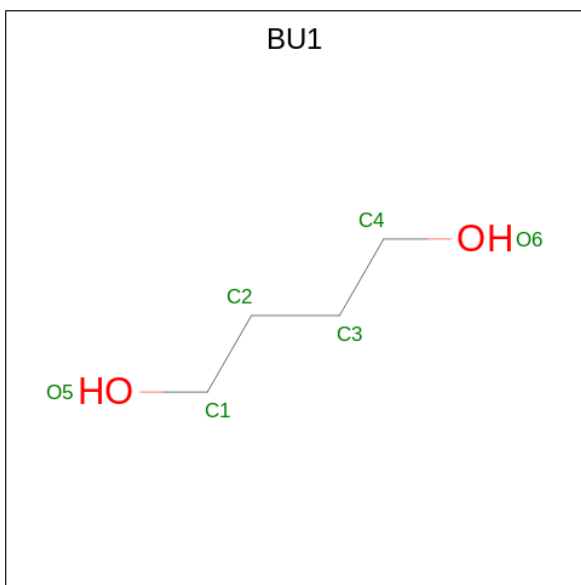
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
10	B	1	Total 6	C 4	O 2	0	0
10	D	1	Total 6	C 4	O 2	0	0

- Molecule 11 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	D	2	Total 2	Zn 2	0	0

- Molecule 12 is 5'-O-[(S)-hydroxy{[(S)-hydroxy(phosphonooxy)phosphoryl]methyl}phosphoryl]cytidine (three-letter code: 2TM) (formula: $C_{10}H_{18}N_3O_{13}P_3$).

2TM
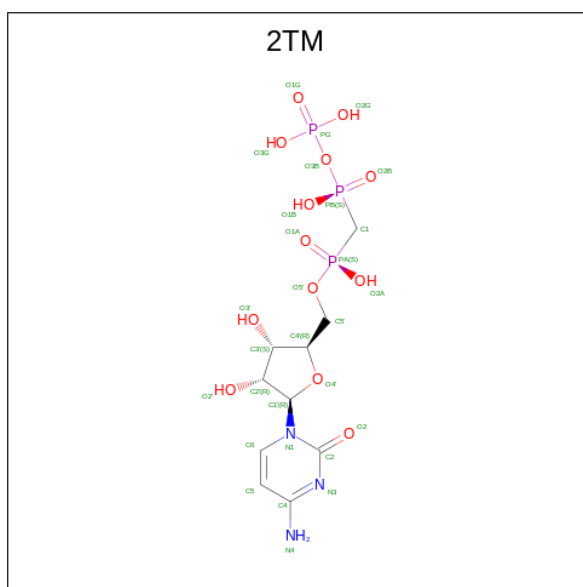
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
12	I	1	Total 29	C 10	N 3	O 13	P 3	0	0

- Molecule 13 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
13	A	4	Total 4	O 4	0	0
13	B	4	Total 4	O 4	0	0
13	C	28	Total 28	O 28	0	0
13	D	42	Total 42	O 42	0	0
13	E	3	Total 3	O 3	0	0
13	F	5	Total 5	O 5	0	0
13	H	1	Total 1	O 1	0	0
13	I	3	Total 3	O 3	0	0

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 electron 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 dot above a residue indicates a poor fit to the electron density (RSRZ > 2). 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: DNA-directed RNA polymerase subunit alpha

• Molecule 1: DNA-directed RNA polymerase subunit alpha

• Molecule 2: DNA-directed RNA polymerase subunit beta

M1 E2 I3 K4 R5 V12 I13 P17 E20 E24 S25 Y26 R27 P35 K38 E50 E56 GLU ASP LYS GLY LYS GLY G63 L64 V65 R72 D81 E82 C83 D87 L88 T89 Y90 Q91 R97 L98 Q99 L100 I101 H102 K103 D104 T105 G106 L107 I108 K109

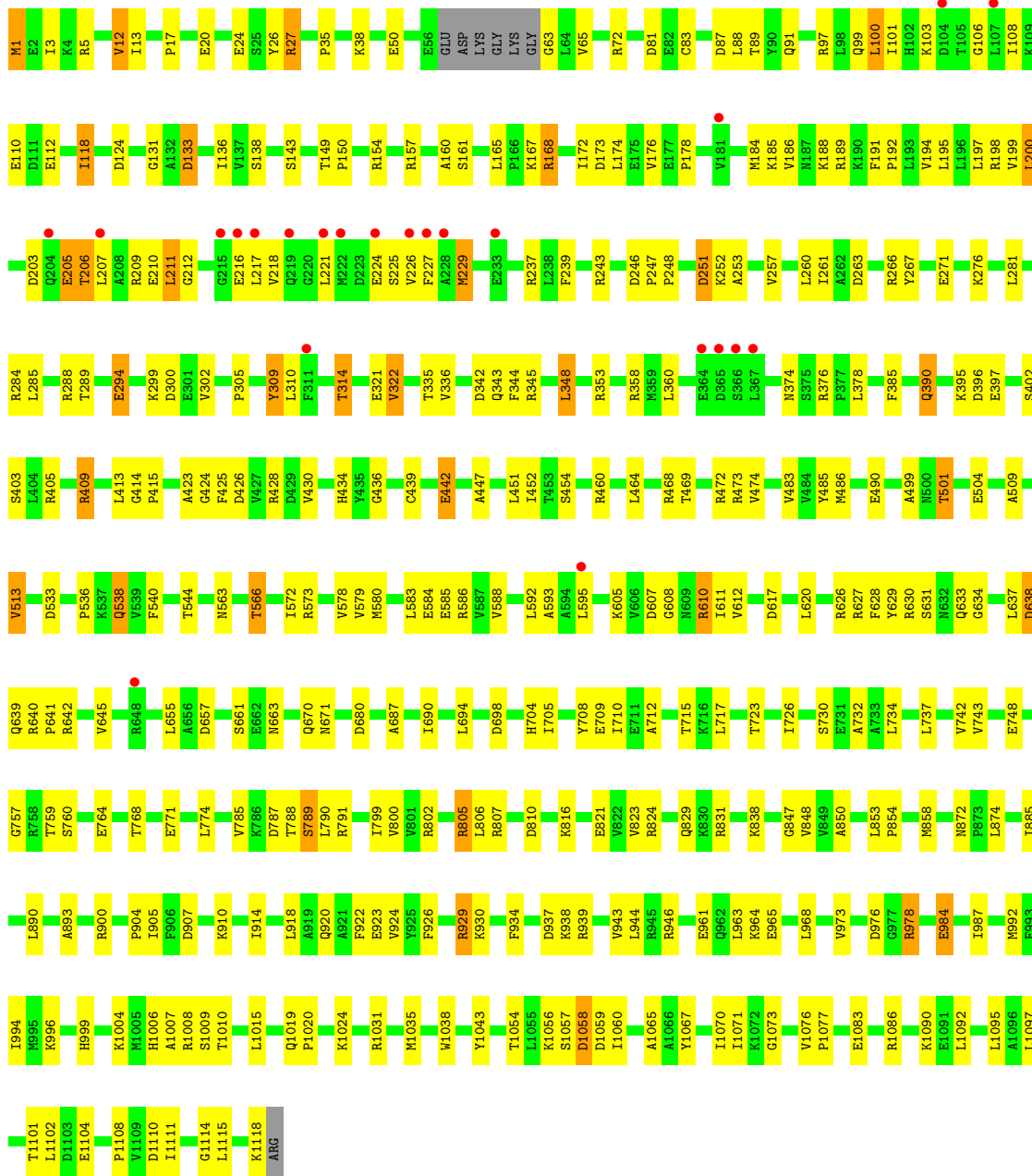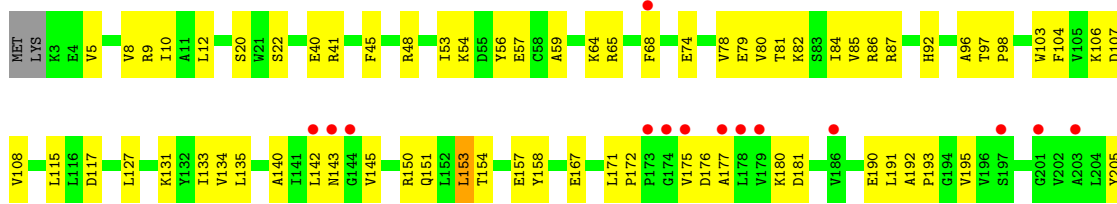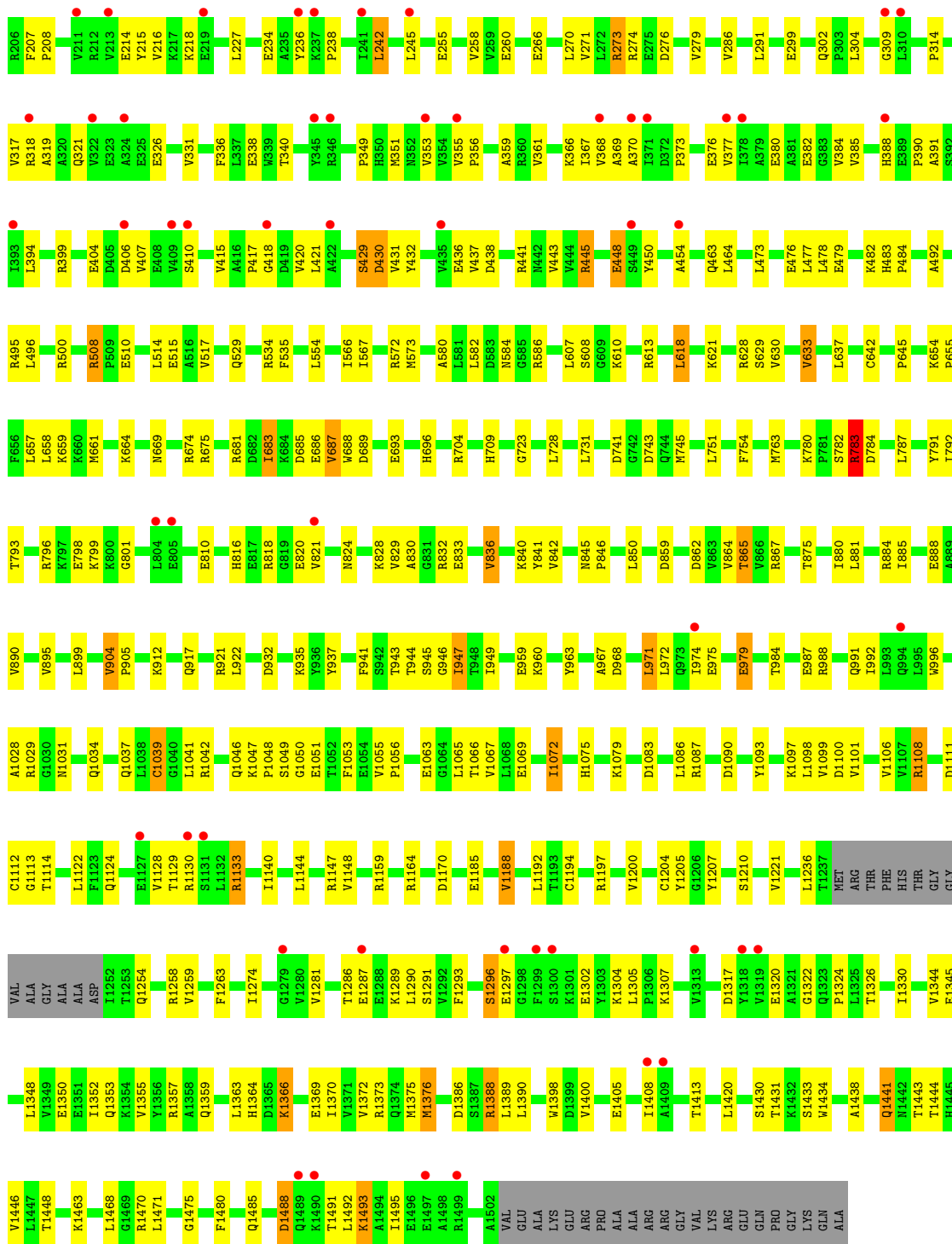E110 D111 E112 I118 D124 G131 A132 D133 I136 V137 S138 S143 T149 P150 R154 R157 A160 S161 L165 P166 K167 R168 I172 D173 L174 E175 V176 E177 P178 V181 M184 K185 V186 N187 K188 R189 K190 F191 P192 L193 V194 L195 L196 L197 R198 V199 L200

D203 Q204 E205 T206 L207 A208 R209 E210 L211 G212 G215 E216 L217 V218 Q219 G220 L221 M222 D223 E224 S225 V226 F227 A228 M229 E233 R237 L238 F239 R243 D246 P247 P248 D251 K252 A253 V257 L260 I261 A262 D263 R266 Y267 E271 K276 L281

R284 L285 R288 T289 E294 K299 D300 E301 V302 P305 Y309 L310 F311 T314 E321 V322 T335 V336 D342 Q343 F344 R345 L348 R353 R358 M359 L360 E364 D365 S366 L367 N374 S375 R376 P377 L378 F385 Q390 K395 D396 E397 S402

S403 L404 R405 R409 L413 G414 P415 A423 G424 F425 D426 V427 R428 D429 V430 H434 Y435 G436 C439 E442 A447 L451 I452 T453 S454 R460 L464 R468 T469 R472 R473 V474 V483 V484 Y485 M486 E490 A499 N500 T501 E504 A509

V513 D533 P536 K537 Q538 V539 F540 T544 N563 T566 I572 R573 V578 V579 M580 L583 E584 E585 R586 V587 V588 L592 A593 A594 L595 K605 V606 D607 G608 N609 R610 I611 V612 D617 L620 R626 R627 F628 Y629 R630 S631 N632 Q633 G634 L637 D638

Q639 R640 P641 R642 V645 R648 L655 A656 D657 S661 E662 N663 Q670 N671 D680 A687 I690 L694 D698 H704 I705 Y708 E709 I710 E711 A712 T715 K716 L717 T723 I726 S730 E731 A732 A733 L734 L737 V742 V743 E748

G757 R758 T759 S760 E764 T768 E771 L774 V785 K786 D787 T788 S789 L790 R791 I799 V800 V801 R802 R805 L806 R807 D810 K816 E821 V822 V823 R824 Q829 K830 R831 K838 G847 V848 V849 A850 L853 P854 M858 N872 P873 L874 I885

L890 A893 R900 P904 I905 F906 D907 K910 I914 L918 A919 Q920 A921 F922 E923 V924 Y925 F926 R929 K930 F934 D937 K938 R939 V943 L944 R945 R946 E961 Q962 L963 K964 E965 L968 V973 D976 G977 R978 E984 I987 M992 F993

I994 M995 K996 H999 K1004 M1005 H1006 A1007 R1008 S1009 T1010 L1015 Q1019 P1020 K1024 R1031 M1035 W1038 Y1043 T1054 L1055 K1056 S1057 D1058 D1059 I1060 A1065 A1066 Y1067 I1070 I1071 K1072 G1073 V1076 P1077 E1083 R1086 K1090 E1091 L1092 L1095 A1096 L1097

T1101 L1102 D1103 E1104 P1108 V1109 D1110 I1111 G1114 L1115 K1118 ARG

• **Molecule 3: DNA-directed RNA polymerase subunit beta'**

Chain D:

4%
68% 27%

MET LYS K3 E4 V5 V8 R9 I10 A11 L12 S20 W21 S22 E40 R41 F45 R48 I53 K54 D55 Y56 E57 C58 A59 K64 R65 F68 E74 V78 E79 V80 T81 K82 S83 I84 V85 R86 R87 H92 A96 T97 P98 W103 F104 V105 K106 D107

V108 L115 L116 D117 L127 K131 Y132 I133 V134 L135 A140 I141 N143 G144 V145 R150 Q151 L152 L153 T154 E157 Y158 E167 L171 P172 P173 G174 V175 D176 A177 L178 V179 K180 D181 V186 E190 L191 A192 P193 G194 V195 V196 S197 G201 V202 A203 L204 Y205

• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:

• Molecule 5: RNA polymerase sigma factor SigA

Chain F:

% 56% 21% · 22%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET LYS LYS SER LYS ARG LYS ASN ALA GLN ALA GLN GLU GLN ALA GLN GLU THR GLU VAL LEU VAL GLN GLU GLU ALA GLU GLU LEU PRO GLU PHE PRO GLU GLY GLU PRO ASP PRO ASP LEU

GLU ASP PRO ASP LEU THR LEU GLU ASP ASP LEU LEU ASP LEU PRO GLU GLU GLY GLU GLY LEU ASP LEU GLU GLU GLU GLU ASP ASP LEU PRO ILE PRO LYS ILE SER THR **S78 D79 P80** Q83 **E87 I88** V91 P92 L93 L94 T95 L96 E97 E98 E99 V100 E101 L102 V106

G109 M110 I113 K114 K115 T120 G121 L122 D123 P124 I127 R128 E129 V130 V131 R132 I135 S138 I144 L147 T150 L151 D152 P153 K154 T155 V156 E157 D160 Q161 K162 L163 L166 H170 K171 R172 Y173 L174 H175 R178 R184 E189 A190 N191 L192

R193 L194 V195 L210 G216 N217 Q218 R222 Y229 K230 T237 W241 K270 L271 S272 R276 Q279 Q280 P286 I291 M295 T307 I310 E313 P314 I321 G322 D323 E324 K325 D326 S327 G330 D331 F332 D335 L338 P339 S340 A345

T346 Q347 S348 L349 L350 L354 L361 S362 E365 K370 L371 I376 D377 G378 L383 I397 R398 Q399 I400 K403 S413 D420 D423

• Molecule 6: DNA (27-MER)

Chain G:

44% 44% · 7%

T1 A4 A5 T6 G7 G8 G9 A10 G11 C12 T15 C16 G19 G20 G23 C24 A25 DG DG

• Molecule 7: DNA (5'-D(*CP*C*TP*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*GP*CP*CP*AP*AP*G)-3')

Chain H:

5% 58% 26% 5% 11%

DC DC T3 G4 C5 A13 G14 C15 C16 A17 A18 G19

• Molecule 8: RNA (5'-R(*GP*G)-3')

Chain I:

100%

There are no outlier residues recorded for this chain.

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	184.12Å 103.32Å 295.65Å 90.00° 98.96° 90.00°	Depositor
Resolution (Å)	25.27 – 2.90 48.88 – 2.90	Depositor EDS
% Data completeness (in resolution range)	82.1 (25.27-2.90) 82.1 (48.88-2.90)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.86 (at 2.91Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.191 , 0.253 0.191 , 0.252	Depositor DCC
R_{free} test set	1971 reflections (1.97%)	wwPDB-VP
Wilson B-factor (Å^2)	49.9	Xtriage
Anisotropy	0.054	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 41.2	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.019 for 1/2*h-3/2*k,-1/2*h-1/2*k,-1/2*h +1/2*k-l 0.024 for 1/2*h+3/2*k,1/2*h-1/2*k,-1/2*h-1/2*k-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	28630	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.16% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 2TM, MG, BU1, 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	#$|Z|$ >5	RMSZ	#$|Z|$ >5
1	A	0.45	0/1835	0.64	0/2497
1	B	0.43	0/1790	0.62	0/2440
2	C	0.48	0/8917	0.63	0/12064
3	D	0.49	1/11923 (0.0%)	0.66	3/16126 (0.0%)
4	E	0.48	0/773	0.65	0/1042
5	F	0.43	0/2848	0.55	0/3833
6	G	1.03	0/580	1.03	1/895 (0.1%)
7	H	0.94	0/388	0.99	2/597 (0.3%)
8	I	0.79	0/48	1.26	0/74
All	All	0.50	1/29102 (0.0%)	0.65	6/39568 (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
3	D	0	3
5	F	0	1
All	All	0	4

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	1039	CYS	CB-SG	-5.85	1.72	1.81

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	H	16	DC	O4'-C4'-C3'	-7.19	101.62	104.50

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	783	ARG	NE-CZ-NH1	-6.26	117.17	120.30
3	D	1108	ARG	NE-CZ-NH1	6.08	123.34	120.30
3	D	1108	ARG	NE-CZ-NH2	-5.91	117.35	120.30
7	H	13	DA	O4'-C4'-C3'	-5.70	102.22	104.50

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	1207	TYR	Peptide
3	D	782	SER	Peptide
3	D	829	VAL	Peptide
5	F	330	GLY	Peptide

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	1803	0	1852	47	0
1	B	1758	0	1778	50	0
2	C	8750	0	8833	229	0
3	D	11714	0	11926	275	0
4	E	759	0	771	18	0
5	F	2803	0	2871	67	0
6	G	516	0	283	14	0
7	H	346	0	192	3	0
8	I	43	0	23	0	0
9	B	2	0	0	0	0
9	C	1	0	0	0	0
9	D	1	0	0	0	0
9	F	1	0	0	0	0
10	B	6	0	10	3	0
10	D	6	0	10	1	0
11	D	2	0	0	0	0
12	I	29	0	14	0	0
13	A	4	0	0	0	0
13	B	4	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
13	C	28	0	0	3	0
13	D	42	0	0	2	0
13	E	3	0	0	0	0
13	F	5	0	0	1	0
13	H	1	0	0	0	0
13	I	3	0	0	0	0
All	All	28630	0	28563	628	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 11.

The worst 5 of 628 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:C:12:VAL:HG21	2:C:472:ARG:HD3	1.46	0.97
2:C:805:ARG:HH21	2:C:807:ARG:HD3	1.26	0.97
2:C:628:PHE:H	2:C:638:ASP:HB2	1.32	0.93
1:B:176:ARG:NH1	3:D:888:GLU:OE2	2.04	0.88
2:C:715:THR:HG22	2:C:717:LEU:H	1.35	0.88

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 X-ray entries followed by that with respect to entries of similar resolution.

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	229/315 (73%)	211 (92%)	18 (8%)	0	100	100
1	B	225/315 (71%)	212 (94%)	13 (6%)	0	100	100
2	C	1108/1119 (99%)	1034 (93%)	74 (7%)	0	100	100
3	D	1483/1524 (97%)	1389 (94%)	92 (6%)	2 (0%)	51	82
4	E	92/99 (93%)	86 (94%)	6 (6%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
5	F	344/443 (78%)	335 (97%)	9 (3%)	0	100	100
All	All	3481/3815 (91%)	3267 (94%)	212 (6%)	2 (0%)	51	82

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	830	ALA
3	D	783	ARG

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 X-ray entries followed by that with respect to entries of similar resolution.

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	199/273 (73%)	182 (92%)	17 (8%)	10	31
1	B	193/273 (71%)	177 (92%)	16 (8%)	11	32
2	C	932/941 (99%)	851 (91%)	81 (9%)	10	30
3	D	1247/1279 (98%)	1151 (92%)	96 (8%)	13	35
4	E	82/88 (93%)	73 (89%)	9 (11%)	6	19
5	F	300/388 (77%)	280 (93%)	20 (7%)	16	43
All	All	2953/3242 (91%)	2714 (92%)	239 (8%)	11	33

5 of 239 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	40	GLU
5	F	97	GLU
3	D	443	VAL
5	F	88	ILE
5	F	371	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	1/2 (50%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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 10 ligands modelled in this entry, 7 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	#$|Z| > 2$
12	2TM	I	2101	-	27,30,30	5.01	18 (66%)	39,47,47	1.21	2 (5%)
10	BU1	D	2004	-	5,5,5	0.38	0	4,4,4	0.17	0
10	BU1	B	2002	-	5,5,5	0.43	0	4,4,4	0.30	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
12	2TM	I	2101	-	-	6/19/38/38	0/2/2/2

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
10	BU1	D	2004	-	-	1/3/3/3	-
10	BU1	B	2002	-	-	1/3/3/3	-

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
12	I	2101	2TM	PB-O3B	15.00	1.75	1.58
12	I	2101	2TM	C2'-C3'	-7.91	1.31	1.53
12	I	2101	2TM	O4'-C4'	-7.42	1.28	1.45
12	I	2101	2TM	C2-N3	7.30	1.51	1.36
12	I	2101	2TM	C6-C5	6.84	1.51	1.35

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
12	I	2101	2TM	PB-O3B-PG	-3.16	121.50	132.62
12	I	2101	2TM	C2'-C3'-C4'	2.11	106.75	102.64

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
12	I	2101	2TM	C5'-O5'-PA-O2A
12	I	2101	2TM	PA-C1-PB-O3B
12	I	2101	2TM	PA-C1-PB-O1B
12	I	2101	2TM	PA-C1-PB-O2B
10	B	2002	BU1	O5-C1-C2-C3

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	D	2004	BU1	1	0
10	B	2002	BU1	3	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be

highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and 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 2TM I 2101

Bond lengths

Bond angles

Torsions

Rings
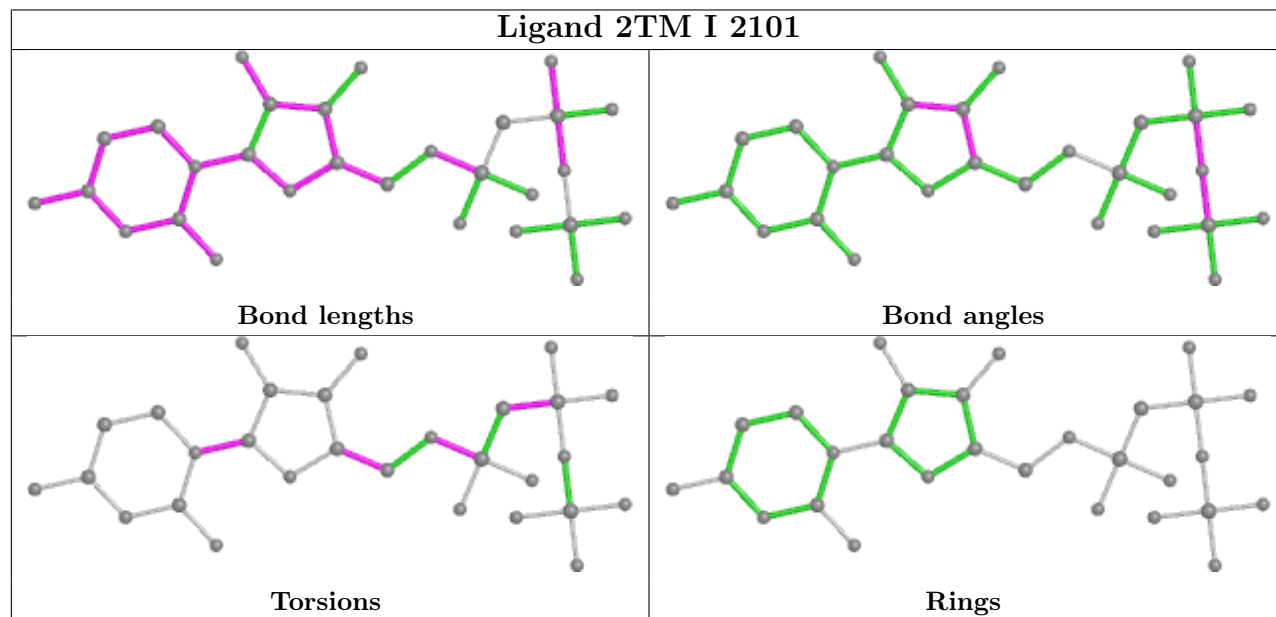
5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å²)	Q<0.9
1	A	231/315 (73%)	-0.25	5 (2%)	62	59	23, 43, 65, 96	0
1	B	227/315 (72%)	-0.17	4 (1%)	68	67	27, 53, 82, 118	0
2	C	1112/1119 (99%)	-0.24	23 (2%)	63	61	12, 37, 87, 121	0
3	D	1486/1524 (97%)	-0.05	67 (4%)	33	29	8, 41, 97, 125	0
4	E	94/99 (94%)	-0.36	0	100	100	20, 40, 71, 82	0
5	F	346/443 (78%)	-0.08	5 (1%)	75	75	24, 56, 94, 114	0
6	G	25/27 (92%)	-0.01	0	100	100	57, 74, 136, 146	0
7	H	17/19 (89%)	0.14	1 (5%)	22	18	31, 60, 140, 144	0
8	I	2/2 (100%)	-0.14	0	100	100	27, 27, 27, 28	2 (100%)
All	All	3540/3863 (91%)	-0.14	105 (2%)	50	45	8, 43, 94, 146	2 (0%)

The worst 5 of 105 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	3	ASP	5.0
2	C	365	ASP	4.9
1	B	2	LEU	4.5
3	D	422	ALA	4.1
2	C	107	LEU	4.1

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

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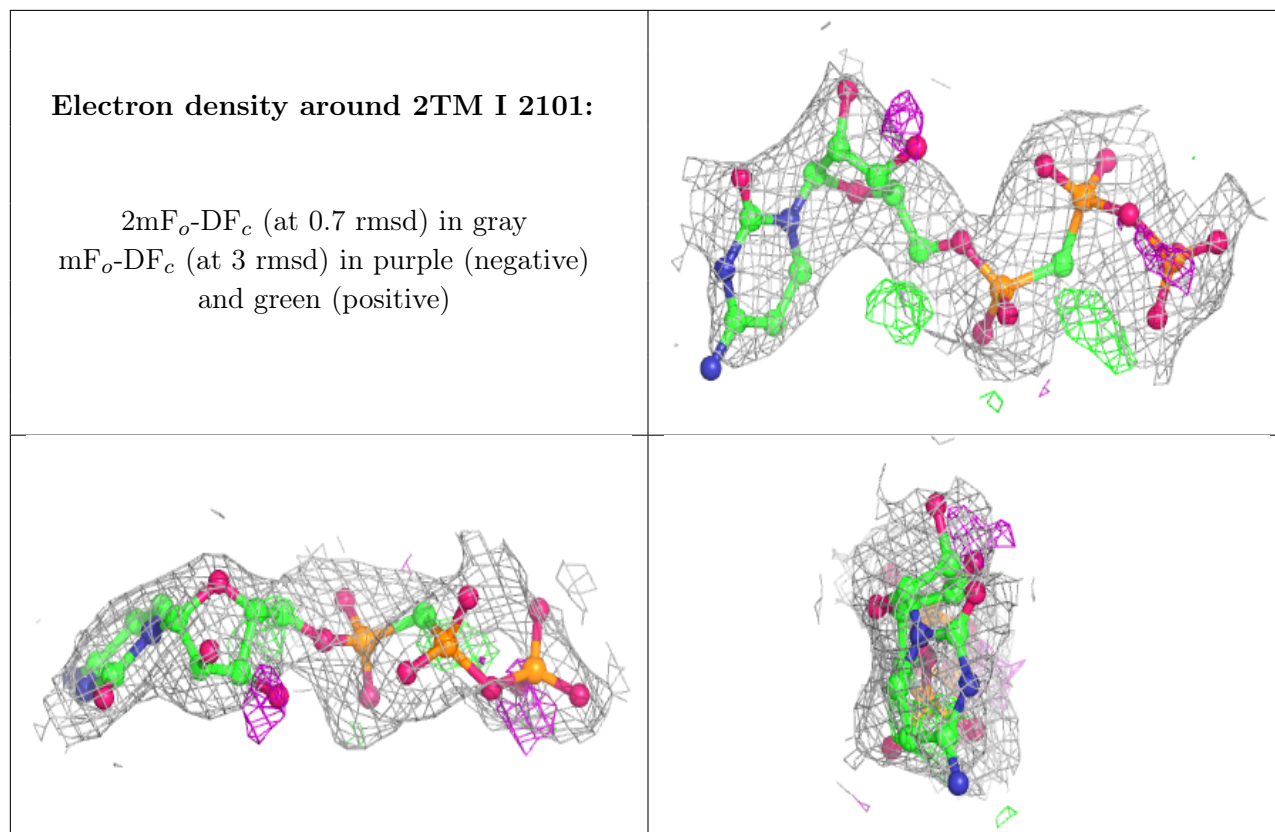
6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
9	MG	B	2001	1/1	0.87	0.11	57,57,57,57	0
12	2TM	I	2101	29/29	0.87	0.24	21,38,72,76	29
10	BU1	B	2002	6/6	0.88	0.29	34,40,48,49	0
10	BU1	D	2004	6/6	0.91	0.33	27,30,37,39	0
9	MG	B	2003	1/1	0.92	0.23	41,41,41,41	0
9	MG	C	2001	1/1	0.94	0.27	20,20,20,20	0
9	MG	F	2001	1/1	0.98	0.10	28,28,28,28	0
11	ZN	D	2001	1/1	0.99	0.11	24,24,24,24	0
11	ZN	D	2002	1/1	0.99	0.08	61,61,61,61	0
9	MG	D	2003	1/1	0.99	0.22	8,8,8,8	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around 2TM I 2101:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative) and green (positive)

6.5 Other polymers ⓘ

There are no such residues in this entry.