wwPDB X-ray Structure Validation Summary Report (i)

Oct 5, 2023 – 07:41 PM EDT

PDB ID	:	6UPY
Title	:	RNA polymerase II elongation complex with 5-guanidinohydantoin lesion in state 2E
Authors	:	Oh, J.; Wang, D.
Deposited on	:	2019-10-18
Resolution	:	3.40 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

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

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

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

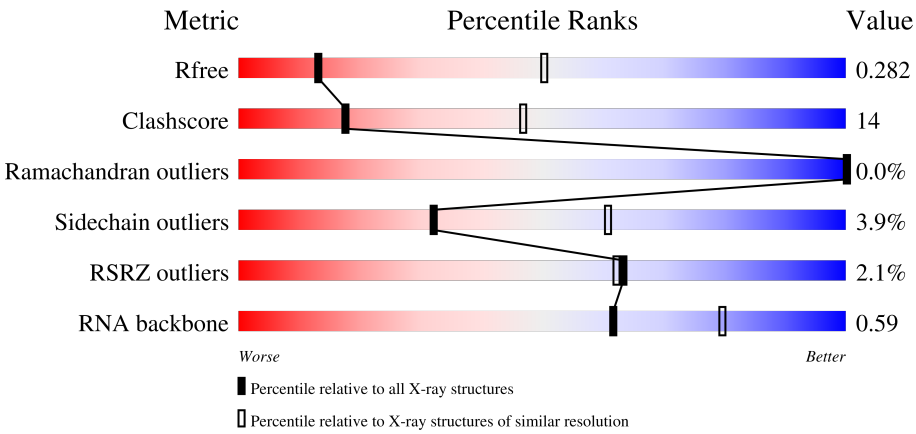
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.40 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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)
RNA backbone	3102	1006 (3.84-2.96)

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	R	9	56% 44%
2	T	29	38% 45% • 14%
3	N	18	22% 61% 17%
4	A	1733	2% 54% 25% • 20%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	B	1224	%, 61%, 28%, 9%
6	C	318	53%, 30%, 16%
7	E	215	9%, 64%, 33%
8	F	155	%, 43%, 11%, 45%
9	H	146	4%, 59%, 32%, 9%
10	I	122	69%, 25%
11	J	70	50%, 43%, 7%
12	K	120	69%, 22%, 5%
13	L	70	4%, 37%, 24%, 39%

2 Entry composition (i)

There are 16 unique types of molecules in this entry. The entry contains 29013 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 RNA chain called RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	R	9	199	88	40	62	9	0	0	0

- Molecule 2 is a DNA chain called Template strand DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	T	25	500	239	78	158	25	0	0	0

- Molecule 3 is a DNA chain called Non-template strand DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	N	15	317	148	71	83	15	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerase II subunit RPB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	A	1384	10851	6845	1898	2048	60	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerase II subunit RPB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	B	1109	8790	5566	1538	1633	53	0	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	C	267	2101	1320	349	419	13	0	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	E	213	1740	1104	307	318	11	0	0	0

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	F	86	684	437	115	129	3	0	0	0

- Molecule 9 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
9	H	133	1058	667	176	211	4	0	0	0

- Molecule 10 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
10	I	117	945	581	172	182	10	0	0	0

- Molecule 11 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
11	J	65	532	339	93	94	6	0	0	0

- Molecule 12 is a protein called DNA-directed RNA polymerase II subunit RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
12	K	114	919	590	156	171	2	0	0	0

- Molecule 13 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
13	L	43	337	208	66	59	4	0	0	0

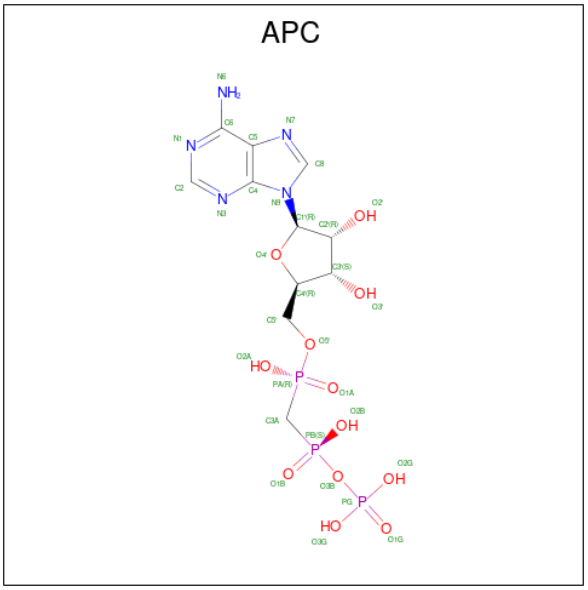
- Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
14	A	2	2	2	0	0
14	B	1	1	1	0	0
14	C	1	1	1	0	0
14	I	2	2	2	0	0
14	J	1	1	1	0	0
14	L	1	1	1	0	0

- Molecule 15 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
15	A	1	1	1	0	0

- Molecule 16 is DIPHOSPHOMETHYLPHOSPHONIC ACID ADENOSYL ESTER (three-letter code: APC) (formula: $C_{11}H_{18}N_5O_{12}P_3$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
16	B	1	31	11	5	12	3	0	0

3 Residue-property plots (i)

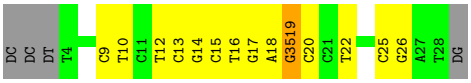
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: RNA

Chain R:

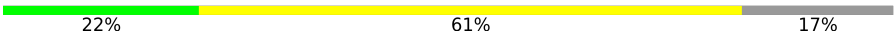
• Molecule 2: Template strand DNA

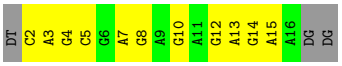
Chain T:



• Molecule 3: Non-template strand DNA

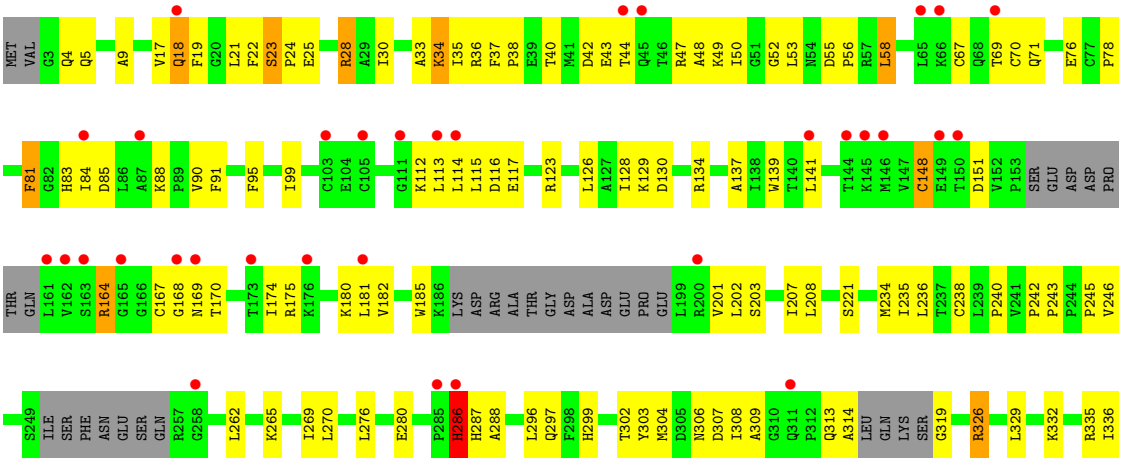
Chain N:

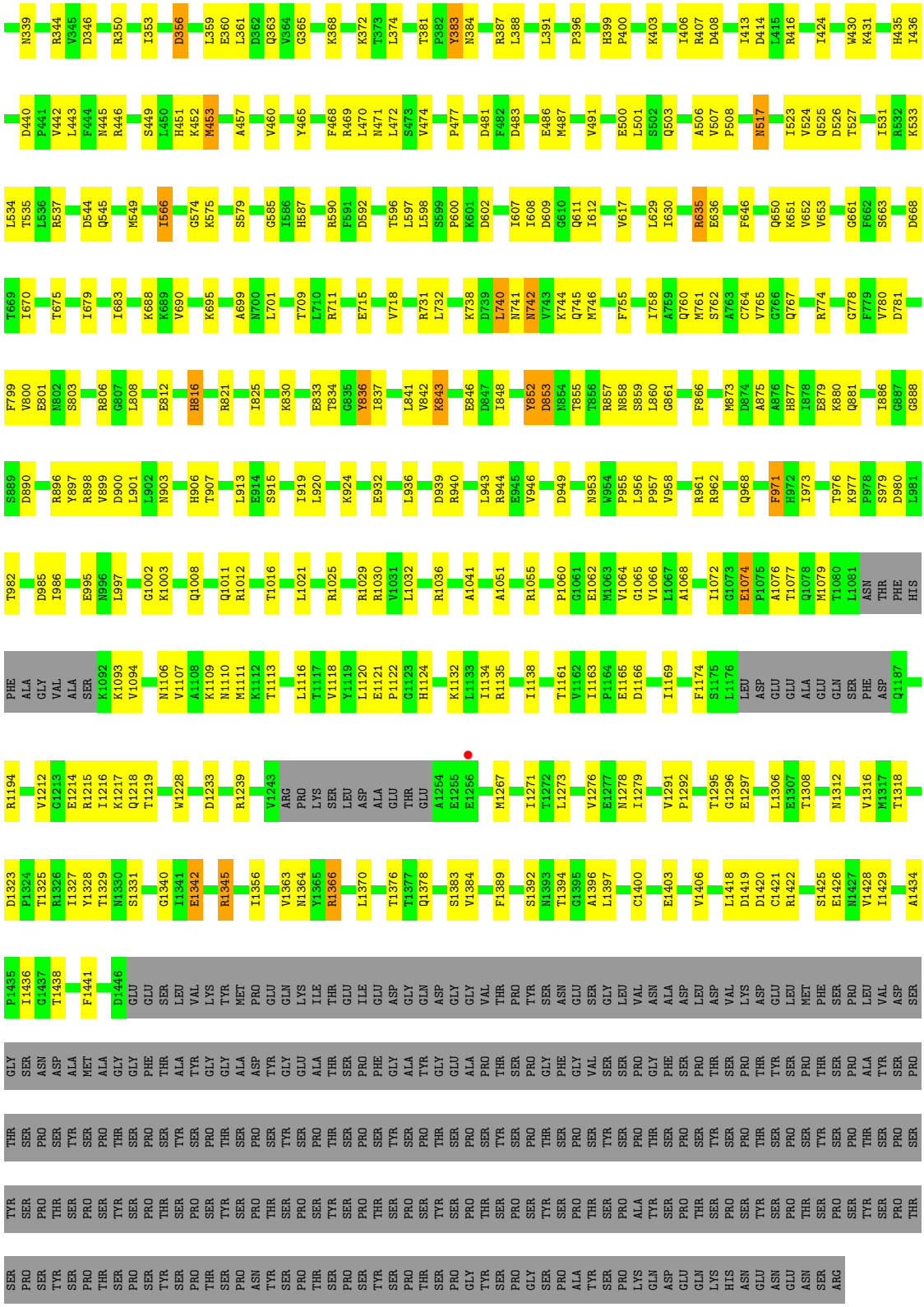




• Molecule 4: DNA-directed RNA polymerase II subunit RPB1

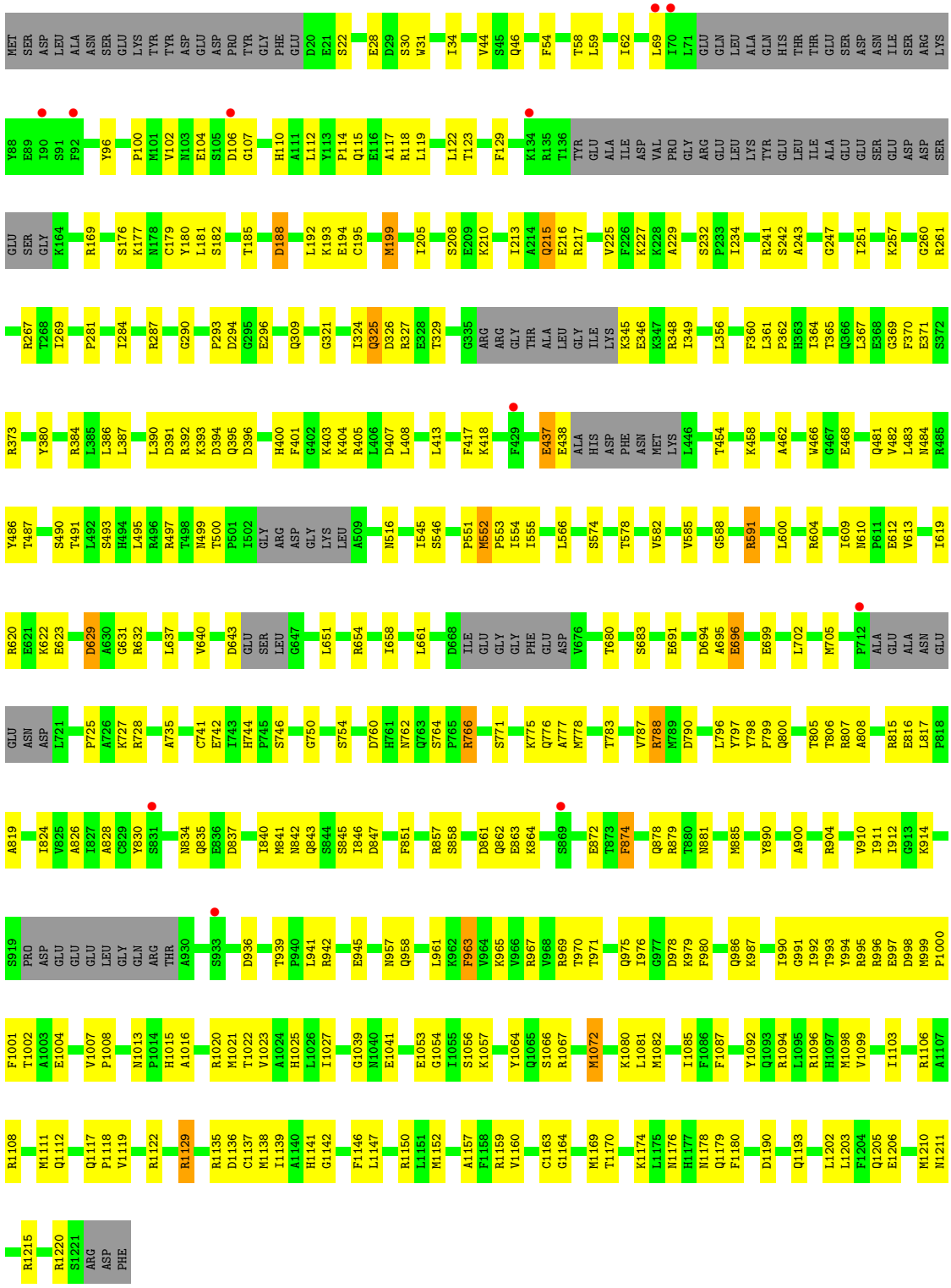
Chain A:





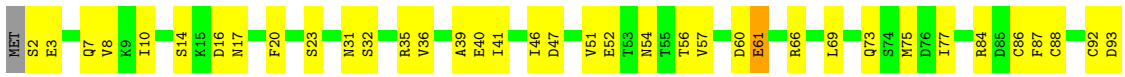
- Molecule 5: DNA-directed RNA polymerase II subunit RPB2

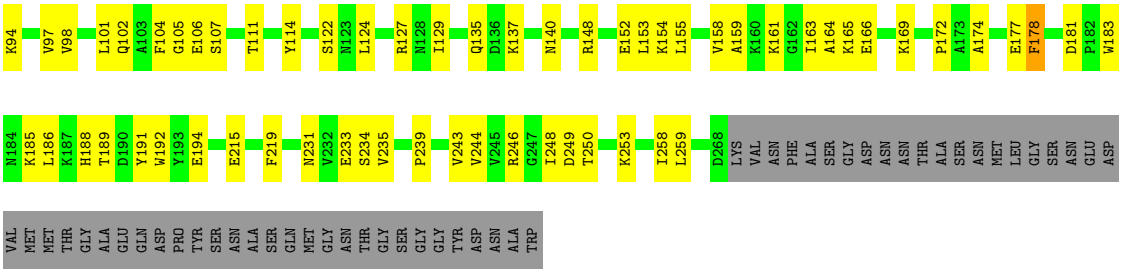
Chain B:



- Molecule 6: DNA-directed RNA polymerase II subunit RPB3

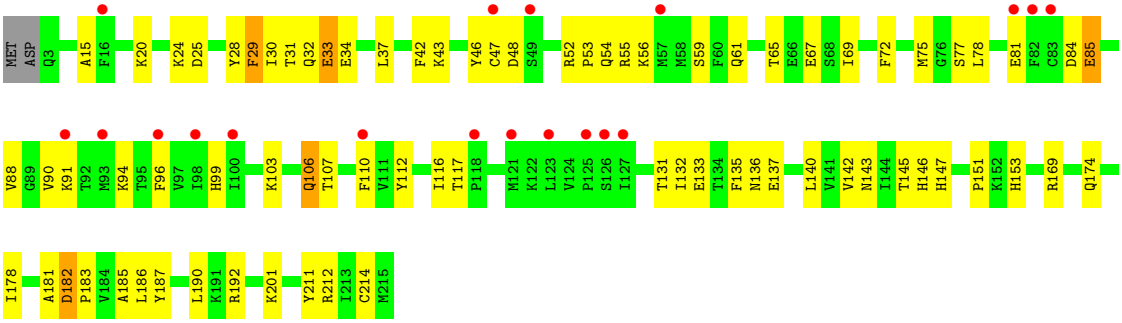
Chain C:





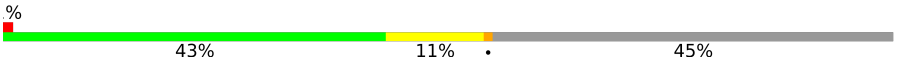
- Molecule 7: DNA-directed RNA polymerases I, II, and III subunit RPABC1

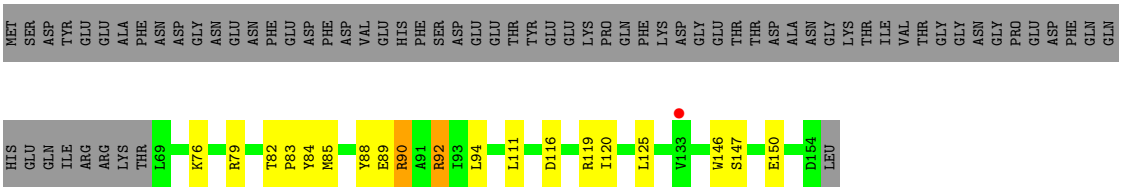
Chain E:



- Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC2

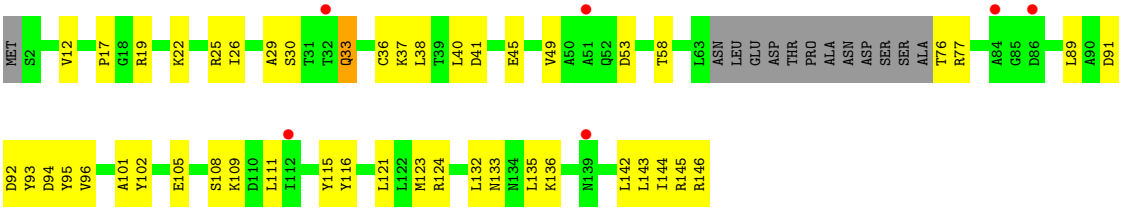
Chain F:





- Molecule 9: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H:

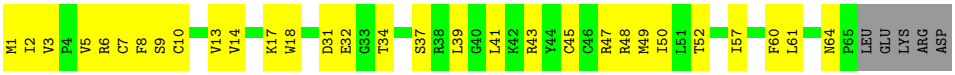


- Molecule 10: DNA-directed RNA polymerase II subunit RPB9

Chain I:

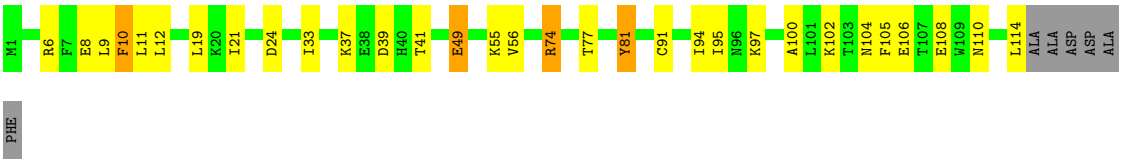
• Molecule 11: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J:



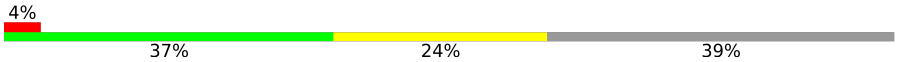
• Molecule 12: DNA-directed RNA polymerase II subunit RPB11

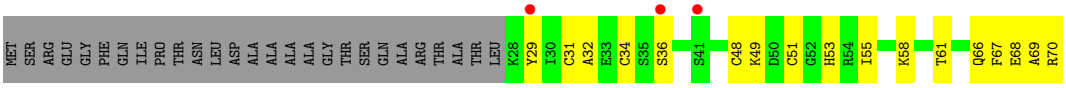
Chain K:



• Molecule 13: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	168.54Å 223.12Å 193.06Å 90.00° 100.92° 90.00°	Depositor
Resolution (Å)	49.22 – 3.40 49.22 – 3.40	Depositor EDS
% Data completeness (in resolution range)	99.5 (49.22-3.40) 99.5 (49.22-3.40)	Depositor EDS
R_{merge}	0.45	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.38 (at 3.40Å)	Xtriage
Refinement program	PHENIX 1.13_2998	Depositor
R, R_{free}	0.227 , 0.282 0.227 , 0.282	Depositor DCC
R_{free} test set	1809 reflections (1.89%)	wwPDB-VP
Wilson B-factor (Å^2)	89.9	Xtriage
Anisotropy	0.531	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 61.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	29013	wwPDB-VP
Average B, all atoms (Å^2)	110.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.07% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: G35, APC, 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	$\#\|Z\|>5$	RMSZ	$\#\|Z\|>5$
1	R	0.74	1/223 (0.4%)	0.75	0/345
2	T	0.55	0/529	1.04	0/809
3	N	0.55	0/359	0.82	0/553
4	A	0.25	0/11042	0.45	0/14932
5	B	0.25	0/8961	0.43	0/12090
6	C	0.25	0/2139	0.45	0/2899
7	E	0.25	0/1776	0.42	0/2391
8	F	0.24	0/696	0.42	0/943
9	H	0.25	0/1076	0.46	0/1459
10	I	0.26	0/963	0.44	0/1298
11	J	0.25	0/541	0.43	0/727
12	K	0.24	0/937	0.41	0/1265
13	L	0.28	0/339	0.61	0/450
All	All	0.27	1/29581 (0.0%)	0.47	0/40161

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
4	A	0	3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	R	1	A	OP3-P	-10.59	1.48	1.61

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	A	1106	ASN	Peptide
4	A	524	VAL	Peptide
4	A	566	ILE	Peptide

5.2 Too-close contacts (i)

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

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	R	199	0	98	7	0
2	T	500	0	287	16	0
3	N	317	0	166	11	0
4	A	10851	0	10916	332	0
5	B	8790	0	8781	259	0
6	C	2101	0	2056	84	0
7	E	1740	0	1761	54	0
8	F	684	0	692	14	0
9	H	1058	0	1018	36	0
10	I	945	0	890	27	0
11	J	532	0	543	31	0
12	K	919	0	929	24	0
13	L	337	0	352	12	0
14	A	2	0	0	0	0
14	B	1	0	0	0	0
14	C	1	0	0	0	0
14	I	2	0	0	0	0
14	J	1	0	0	0	0
14	L	1	0	0	0	0
15	A	1	0	0	0	0
16	B	31	0	14	2	0
All	All	29013	0	28503	779	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 779 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)
16:B:1301:APC:O4'	16:B:1301:APC:C1'	1.63	1.25
9:H:108:SER:HB2	9:H:111:LEU:HD13	1.21	1.15
4:A:451:HIS:CE1	4:A:1074:GLU:HG3	1.88	1.07
4:A:443:LEU:HD12	5:B:1146:PHE:CZ	2.01	0.95
9:H:109:LYS:O	9:H:111:LEU:HD12	1.77	0.85

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 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	
4	A	1368/1733 (79%)	1298 (95%)	69 (5%)	1 (0%)	51	82
5	B	1089/1224 (89%)	1043 (96%)	46 (4%)	0	100	100
6	C	265/318 (83%)	259 (98%)	6 (2%)	0	100	100
7	E	211/215 (98%)	204 (97%)	7 (3%)	0	100	100
8	F	84/155 (54%)	82 (98%)	2 (2%)	0	100	100
9	H	129/146 (88%)	123 (95%)	6 (5%)	0	100	100
10	I	115/122 (94%)	110 (96%)	5 (4%)	0	100	100
11	J	63/70 (90%)	63 (100%)	0	0	100	100
12	K	112/120 (93%)	108 (96%)	4 (4%)	0	100	100
13	L	41/70 (59%)	36 (88%)	5 (12%)	0	100	100
All	All	3477/4173 (83%)	3326 (96%)	150 (4%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	286	HIS

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all 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	
4	A	1200/1520 (79%)	1144 (95%)	56 (5%)	26	57
5	B	955/1061 (90%)	925 (97%)	30 (3%)	40	68
6	C	235/274 (86%)	231 (98%)	4 (2%)	60	80
7	E	194/197 (98%)	182 (94%)	12 (6%)	18	48
8	F	73/137 (53%)	71 (97%)	2 (3%)	44	70
9	H	115/128 (90%)	113 (98%)	2 (2%)	60	80
10	I	109/116 (94%)	102 (94%)	7 (6%)	17	47
11	J	60/65 (92%)	60 (100%)	0	100	100
12	K	99/102 (97%)	94 (95%)	5 (5%)	24	54
13	L	37/57 (65%)	35 (95%)	2 (5%)	22	52
All	All	3077/3657 (84%)	2957 (96%)	120 (4%)	32	61

5 of 120 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	B	106	ASP
10	I	51	ASN
5	B	552	MET
10	I	46	HIS
13	L	31	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
5	B	1015	HIS
5	B	1025	HIS
7	E	174	GLN
4	A	877	HIS
4	A	1004	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	R	8/9 (88%)	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 (i)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

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	$\#\|Z\| > 2$
2	G35	T	19	2	18,23,24	4.73	14 (77%)	20,33,36	2.07	4 (20%)

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
2	G35	T	19	2	-	3/10/41/42	0/2/2/2

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	19	G35	C2-N3	9.37	1.49	1.33
2	T	19	G35	O4'-C4'	7.89	1.62	1.45
2	T	19	G35	C5-N7	6.58	1.46	1.37
2	T	19	G35	C3'-C4'	-6.45	1.35	1.53
2	T	19	G35	C8-N9	6.13	1.46	1.37

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	T	19	G35	O4'-C1'-N9	6.07	115.85	108.65
2	T	19	G35	C5-C4-N9	4.05	107.65	102.28
2	T	19	G35	C4'-O4'-C1'	-3.74	100.40	109.45
2	T	19	G35	O4'-C1'-C2'	-3.15	100.30	106.25

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	T	19	G35	C3'-C4'-C5'-O5'
2	T	19	G35	C4'-C5'-O5'-P
2	T	19	G35	O4'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	T	19	G35	2	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 10 ligands modelled in this entry, 9 are monoatomic - leaving 1 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	$\#\|Z\| > 2$
16	APC	B	1301	-	27,33,33	5.05	11 (40%)	31,52,52	2.30	6 (19%)

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
16	APC	B	1301	-	-	6/15/38/38	0/3/3/3

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
16	B	1301	APC	O4'-C1'	16.29	1.63	1.41
16	B	1301	APC	C2'-C1'	-14.74	1.31	1.53
16	B	1301	APC	PB-O3B	8.26	1.67	1.58
16	B	1301	APC	O4'-C4'	-6.45	1.30	1.45
16	B	1301	APC	PA-O5'	4.72	1.64	1.57

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
16	B	1301	APC	C5-C6-N6	8.19	132.80	120.35
16	B	1301	APC	N3-C2-N1	-5.54	120.02	128.68
16	B	1301	APC	N6-C6-N1	-5.44	107.28	118.57
16	B	1301	APC	PB-O3B-PG	-3.21	121.32	132.62
16	B	1301	APC	C3'-C2'-C1'	2.92	105.38	100.98

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
16	B	1301	APC	PB-C3A-PA-O1A
16	B	1301	APC	PB-C3A-PA-O2A
16	B	1301	APC	PB-C3A-PA-O5'
16	B	1301	APC	O4'-C4'-C5'-O5'
16	B	1301	APC	C3'-C4'-C5'-O5'

There are no ring outliers.

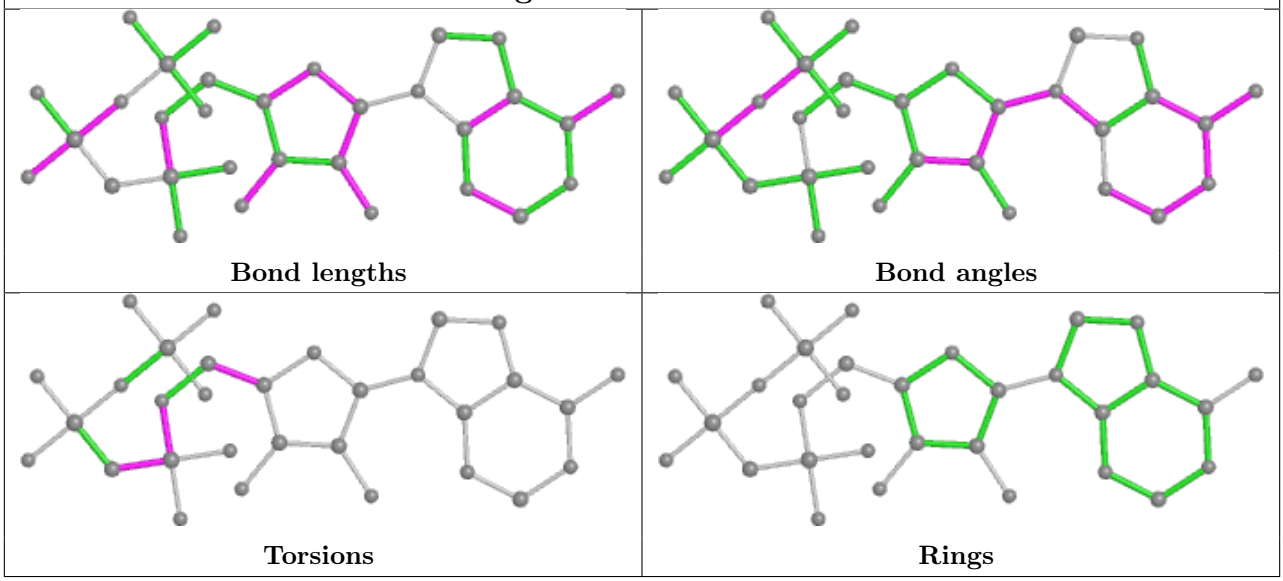
1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
16	B	1301	APC	2	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 APC B 1301



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 (i)

6.1 Protein, DNA and RNA chains (i)

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(Å^2)	Q<0.9
1	R	9/9 (100%)	-0.59	0	100	100	89, 97, 145, 188	0
2	T	24/29 (82%)	-0.44	0	100	100	86, 171, 219, 235	0
3	N	15/18 (83%)	0.03	0	100	100	183, 203, 251, 254	0
4	A	1384/1733 (79%)	0.01	34 (2%)	57	55	51, 104, 184, 265	0
5	B	1109/1224 (90%)	-0.08	11 (0%)	82	81	32, 85, 153, 231	0
6	C	267/318 (83%)	-0.24	0	100	100	56, 92, 138, 180	0
7	E	213/215 (99%)	0.14	19 (8%)	9	11	85, 140, 236, 292	0
8	F	86/155 (55%)	-0.25	1 (1%)	79	77	66, 107, 149, 209	0
9	H	133/146 (91%)	0.16	6 (4%)	33	33	82, 125, 183, 224	0
10	I	117/122 (95%)	-0.29	0	100	100	73, 106, 144, 167	0
11	J	65/70 (92%)	-0.33	0	100	100	42, 73, 127, 145	0
12	K	114/120 (95%)	-0.18	0	100	100	62, 93, 131, 180	0
13	L	43/70 (61%)	0.51	3 (6%)	16	18	58, 146, 220, 268	0
All	All	3579/4229 (84%)	-0.05	74 (2%)	63	62	32, 100, 186, 292	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	69	THR	7.2
7	E	93	MET	5.7
7	E	110	PHE	5.6
4	A	258	GLY	5.0
4	A	66	LYS	5.0

6.2 Non-standard residues in protein, DNA, RNA chains (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
2	G35	T	19	22/23	0.84	0.29	138,155,172,184	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

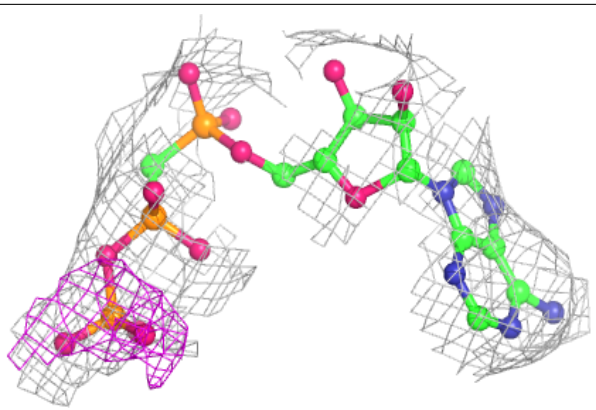
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

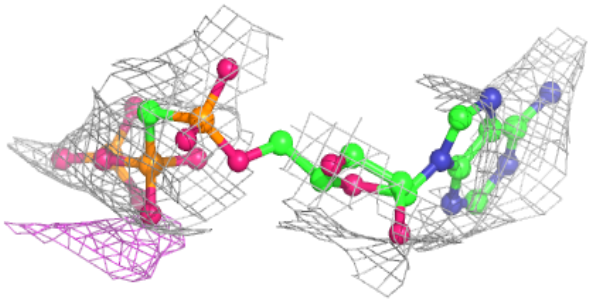
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
14	ZN	A	1801	1/1	0.66	0.12	393,393,393,393	0
16	APC	B	1301	31/31	0.83	0.29	131,160,243,251	0
14	ZN	L	101	1/1	0.89	0.06	231,231,231,231	0
14	ZN	A	1802	1/1	0.92	0.08	102,102,102,102	0
14	ZN	B	1302	1/1	0.94	0.10	209,209,209,209	0
14	ZN	C	401	1/1	0.97	0.10	131,131,131,131	0
14	ZN	I	201	1/1	0.98	0.14	110,110,110,110	0
14	ZN	J	101	1/1	0.98	0.22	79,79,79,79	0
15	MG	A	1803	1/1	0.99	0.14	39,39,39,39	0
14	ZN	I	202	1/1	0.99	0.10	80,80,80,80	0

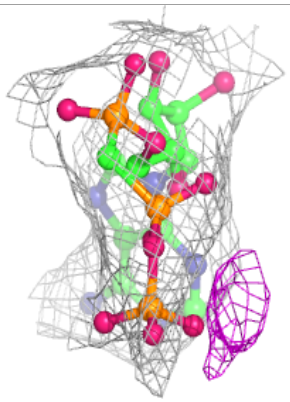
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around APC B 1301:

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







6.5 Other polymers (i)

There are no such residues in this entry.