



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 23, 2020 – 01:20 am BST

PDB ID : 4A3E  
Title : RNA Polymerase II initial transcribing complex with a 5nt DNA-RNA hybrid and soaked with AMPCPP  
Authors : Cheung, A.C.M.; Sainsbury, S.; Cramer, P.  
Deposited on : 2011-09-30  
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](mailto: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 ⓘ symbol.

---

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 : 2.11  
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) : 2.11

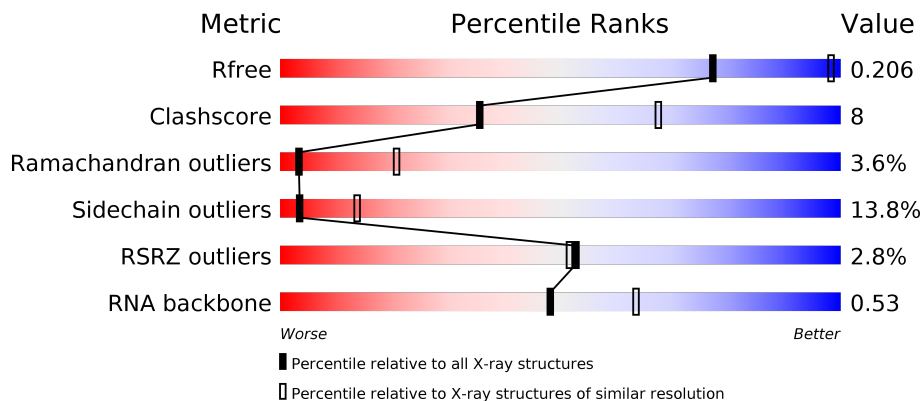
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	1732	 2% 56% 21% 18%
2	B	1224	 3% 65% 22% 9%
3	C	318	 60% 19% 5% 16%
4	D	221	 58% 19% 19%

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	E	215	
6	F	155	
7	G	171	
8	H	146	
9	I	122	
10	J	70	
11	K	120	
12	L	70	
13	N	14	
14	P	5	
15	T	26	

## 2 Entry composition [i](#)

There are 18 unique types of molecules in this entry. The entry contains 31855 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 II SUBUNIT RPB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1423	11182	7043	1955	2122	62	0	0	0

- Molecule 2 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1115	8859	5609	1554	1641	55	0	0	0

- Molecule 3 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	266	2095	1317	348	417	13	0	0	0

- Molecule 4 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	178	1434	887	257	288	2	0	0	0

- Molecule 5 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 1.

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

- Molecule 6 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	F	84	679	434	115	127	3	0	0	0

- Molecule 7 is a protein called RPB7, DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	G	171	1340	861	222	249	8	0	0	0

- Molecule 8 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	H	133	1068	673	180	211	4	0	0	0

- Molecule 9 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
9	I	119	971	596	179	186	10	0	0	0

- Molecule 10 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 5.

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

- Molecule 11 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
11	K	115	920	590	157	171	2	0	0	1

- Molecule 12 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
12	L	46	363	224	72	63	4	0	0	0

- Molecule 13 is a DNA chain called NON TEMPLATE DNA 5'-D(\*TP\*AP\*AP\*GP\*TP\*AP\*CP\*TP\*TP\*GP\*AP \*GP\*CP\*TP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
13	N	8	165	79	29	49	8	0	0	0

- Molecule 14 is a RNA chain called TRANSCRIPT RNA 5'-R(\*CP\*AP\*GP\*GP\*AP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
14	P	4	90	40	20	26	4	0	0	0

- Molecule 15 is a DNA chain called TEMPLATE DNA 5'-D(\*AP\*GP\*CP\*TP\*CP\*AP\*AP\*GP\*TP\*AP\*CP\*TP\*TP \*TP\*TP\*TP\*CP\*CP\*BRU\*GP\*GP\*TP\*CP\*AP\*TP\*TP)-3'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	Br	C	N	O	P			
15	T	18	365	1	175	58	113	18	0	0	0

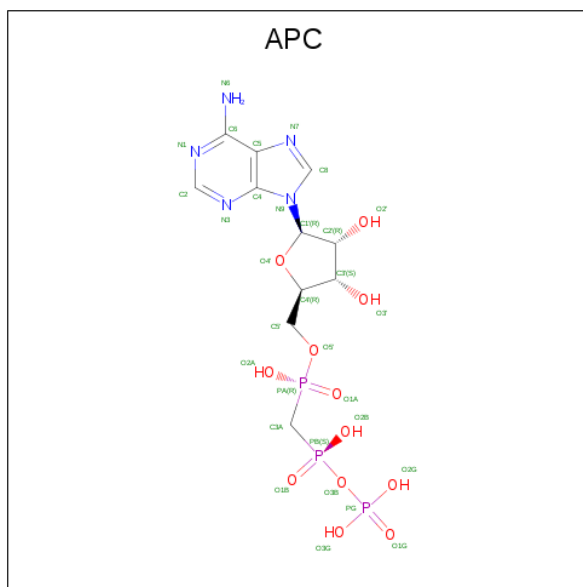
- Molecule 16 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 17 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 18 is DIPHOSPHOMETHYLPHOSPHONIC ACID ADENOSYL ESTER (three-letter code: APC) (formula: C<sub>11</sub>H<sub>18</sub>N<sub>5</sub>O<sub>12</sub>P<sub>3</sub>).

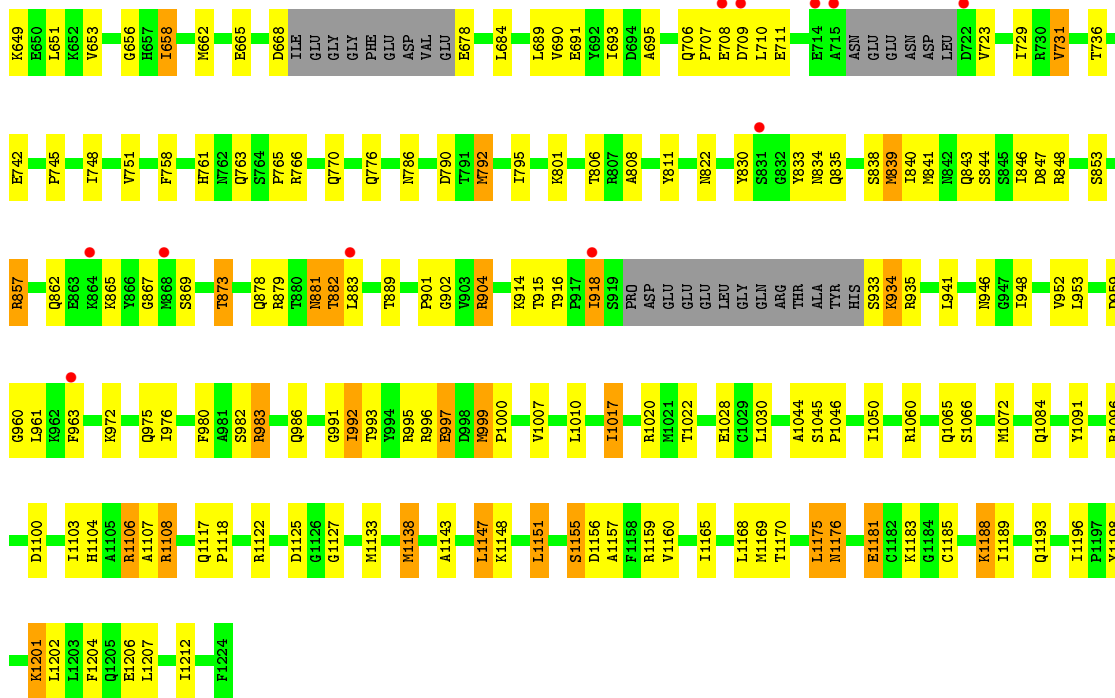


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
18	P	1	31	11	5	12	3	0	0

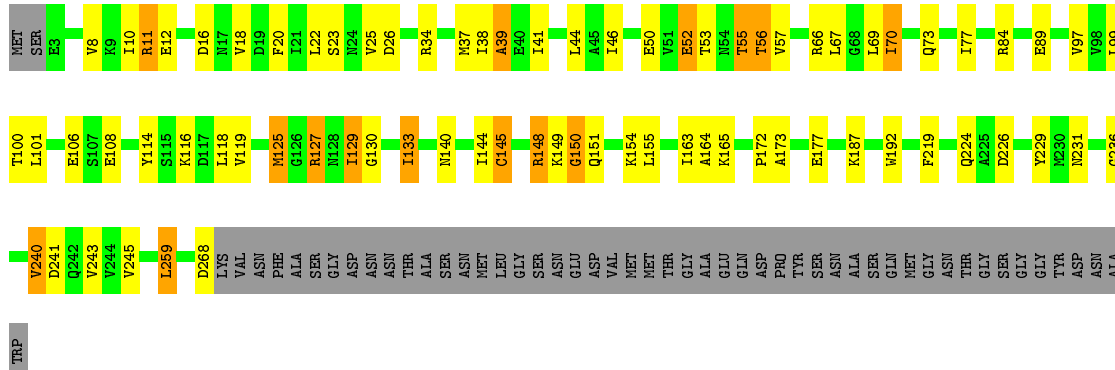




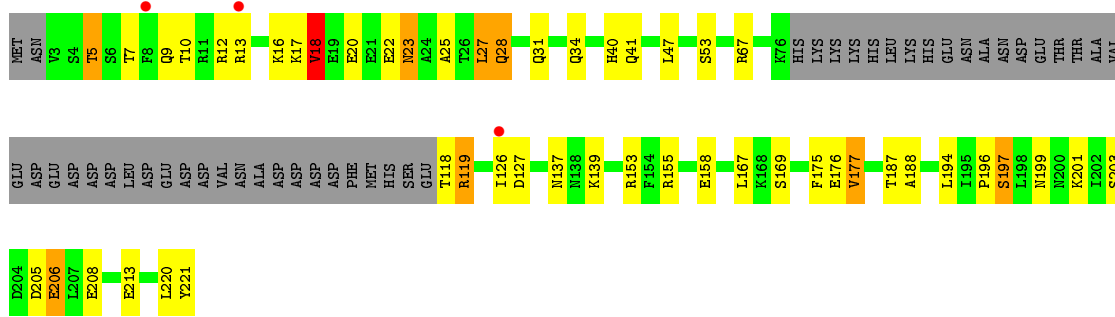




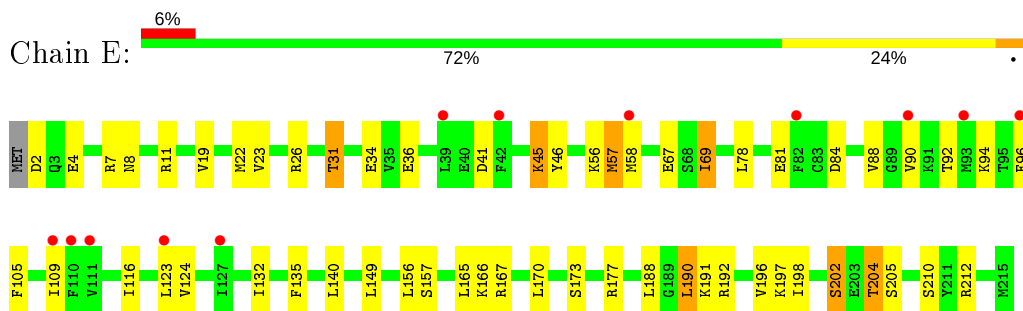
• Molecule 3: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB3



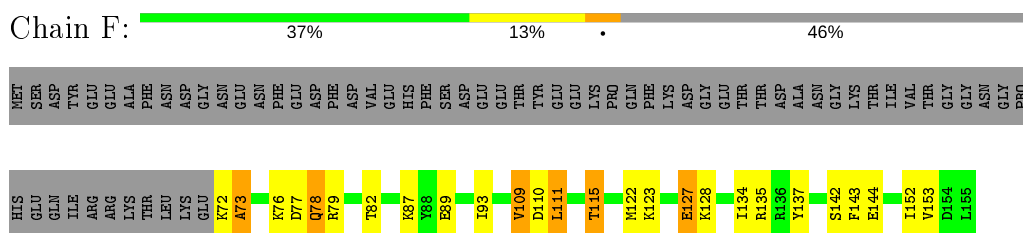
• Molecule 4: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB4



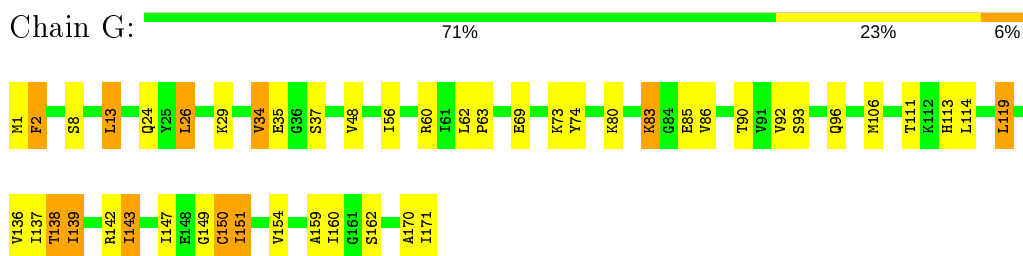
- Molecule 5: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 1



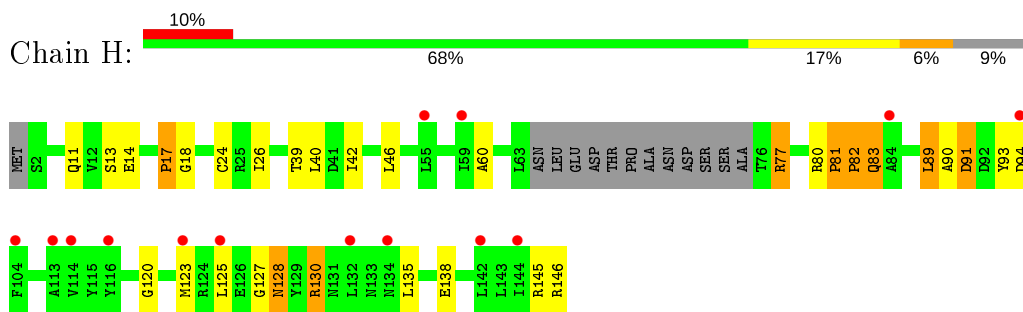
- Molecule 6: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 2



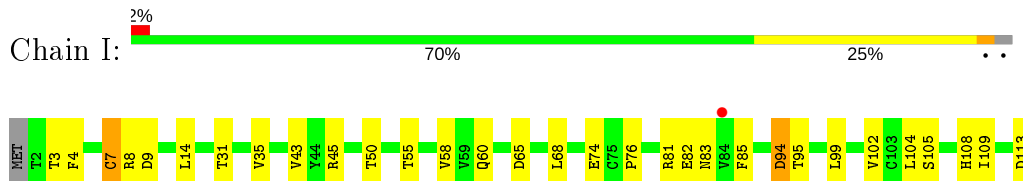
- Molecule 7: RPB7, DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB7



- Molecule 8: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 3



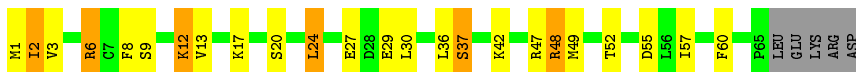
- Molecule 9: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB9





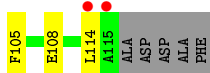
- Molecule 10: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 5

Chain J: 59% 26% 9% 7%



- Molecule 11: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB11

Chain K: 2% 65% 28%



- Molecule 12: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC 4

Chain L: 9% 31% 21% 11% 28%



- Molecule 13: NON TEMPLATE DNA 5'-D(\*TP\*AP\*AP\*GP\*TP\*AP\*CP\*TP\*TP\*GP\*AP\*GP\*CP\*TP)-3'

Chain N: 29% 29% 43%



- Molecule 14: TRANSCRIPT RNA 5'-R(\*CP\*AP\*GP\*GP\*AP)-3'

Chain P: 60% 20% 20%



- Molecule 15: TEMPLATE DNA 5'-D(\*AP\*GP\*CP\*TP\*CP\*AP\*AP\*GP\*TP\*AP\*CP\*TP\*TP\*TP\*TP\*CP\*CP\*BRU\*GP\*GP\*TP\*CP\*AP\*TP\*TP)-3'

Chain T: 4% 23% 35% 12% 26%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	223.68Å 394.17Å 283.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	54.61 – 3.40 54.61 – 3.40	Depositor EDS
% Data completeness (in resolution range)	97.8 (54.61-3.40) 97.8 (54.61-3.40)	Depositor EDS
$R_{merge}$	0.63	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.07 (at 3.40Å)	Xtriage
Refinement program	BUSTER 2.11.2	Depositor
R, $R_{free}$	0.160 , 0.188 0.181 , 0.206	Depositor DCC
$R_{free}$ test set	3319 reflections (1.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	111.9	Xtriage
Anisotropy	0.261	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 100.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	0.029 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.039 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	31855	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	125.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.53% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: MG, APC, ZN, BRU

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	0.54	0/11382	0.84	9/15394 (0.1%)
2	B	0.51	0/9029	0.79	3/12171 (0.0%)
3	C	0.49	0/2133	0.79	2/2891 (0.1%)
4	D	0.51	0/1444	0.82	0/1935
5	E	0.46	0/1788	0.71	0/2406
6	F	0.61	0/691	0.83	0/933
7	G	0.55	0/1368	0.82	0/1844
8	H	0.50	0/1086	0.80	0/1470
9	I	0.47	0/989	0.76	0/1331
10	J	0.55	0/541	0.86	0/727
11	K	0.47	0/938	0.72	0/1267
12	L	0.56	0/365	1.01	0/485
13	N	1.37	0/184	1.03	0/282
14	P	1.09	0/101	0.78	0/156
15	T	1.52	2/383 (0.5%)	1.16	2/586 (0.3%)
All	All	0.55	2/32422 (0.0%)	0.82	16/43878 (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
1	A	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
15	T	25	DT	C1'-N1	7.01	1.58	1.49
15	T	17	DT	C3'-O3'	-5.99	1.36	1.44

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	34	LYS	C-N-CA	7.41	140.22	121.70
1	A	399	HIS	N-CA-CB	7.12	123.41	110.60
1	A	56	PRO	C-N-CA	6.21	137.23	121.70
1	A	194	ALA	C-N-CA	5.82	136.26	121.70
1	A	54	ASN	C-N-CA	5.80	136.20	121.70

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	34	LYS	Mainchain,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	11182	0	11244	209	0
2	B	8859	0	8901	147	0
3	C	2095	0	2051	42	0
4	D	1434	0	1460	16	0
5	E	1752	0	1776	29	0
6	F	679	0	701	19	0
7	G	1340	0	1357	31	0
8	H	1068	0	1040	12	0
9	I	971	0	927	9	0
10	J	532	0	542	15	0
11	K	920	0	929	23	0
12	L	363	0	386	13	0
13	N	165	0	92	5	0
14	P	90	0	44	1	0
15	T	365	0	204	14	0
16	A	2	0	0	0	0
16	B	1	0	0	0	0
16	C	1	0	0	0	0
16	I	2	0	0	0	0
16	J	1	0	0	0	0

*Continued on next page...*



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
16	L	1	0	0	0	0
17	A	1	0	0	0	0
18	P	31	0	14	2	0
All	All	31855	0	31668	516	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 8.

The worst 5 of 516 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:867:ILE:CD1	1:A:867:ILE:CG1	1.78	1.60
4:D:40:HIS:HB3	7:G:73:LYS:HE3	1.26	1.13
1:A:53:LEU:HD23	1:A:54:ASN:H	1.22	1.01
5:E:4:GLU:HB3	5:E:7:ARG:HE	1.33	0.92
10:J:48:ARG:HE	10:J:49:MET:HE2	1.36	0.89

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
1	A	1415/1732 (82%)	1247 (88%)	113 (8%)	55 (4%)	<b>3</b> 19
2	B	1095/1224 (90%)	963 (88%)	94 (9%)	38 (4%)	<b>3</b> 21
3	C	264/318 (83%)	243 (92%)	19 (7%)	2 (1%)	19 <b>51</b>
4	D	174/221 (79%)	154 (88%)	13 (8%)	7 (4%)	<b>3</b> 18
5	E	212/215 (99%)	198 (93%)	12 (6%)	2 (1%)	17 49
6	F	82/155 (53%)	75 (92%)	5 (6%)	2 (2%)	<b>6</b> 28
7	G	169/171 (99%)	156 (92%)	10 (6%)	3 (2%)	<b>8</b> 32

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
8	H	129/146 (88%)	103 (80%)	17 (13%)	9 (7%)	1	7
9	I	117/122 (96%)	94 (80%)	15 (13%)	8 (7%)	1	8
10	J	63/70 (90%)	52 (82%)	7 (11%)	4 (6%)	1	9
11	K	113/120 (94%)	109 (96%)	4 (4%)	0	100	100
12	L	44/70 (63%)	28 (64%)	8 (18%)	8 (18%)	0	0
All	All	3877/4564 (85%)	3422 (88%)	317 (8%)	138 (4%)	3	21

5 of 138 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	42	ASP
1	A	57	ARG
1	A	74	MET
1	A	76	GLU
1	A	189	ARG

### 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	
1	A	1241/1519 (82%)	1053 (85%)	188 (15%)	3	12
2	B	966/1061 (91%)	850 (88%)	116 (12%)	5	19
3	C	234/274 (85%)	204 (87%)	30 (13%)	4	16
4	D	160/200 (80%)	129 (81%)	31 (19%)	1	4
5	E	196/197 (100%)	172 (88%)	24 (12%)	5	18
6	F	74/137 (54%)	66 (89%)	8 (11%)	6	24
7	G	152/152 (100%)	134 (88%)	18 (12%)	5	19
8	H	117/128 (91%)	103 (88%)	14 (12%)	5	19
9	I	113/116 (97%)	101 (89%)	12 (11%)	6	24
10	J	60/65 (92%)	50 (83%)	10 (17%)	2	8
11	K	99/102 (97%)	85 (86%)	14 (14%)	3	13

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
12	L	40/57 (70%)	30 (75%)	10 (25%)	0 2
All	All	3452/4008 (86%)	2977 (86%)	475 (14%)	3 14

5 of 475 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	347	LYS
2	B	857	ARG
9	I	94	ASP
2	B	393	LYS
2	B	574	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	357	GLN
2	B	975	GLN
10	J	23	ASN
2	B	587	HIS
2	B	957	ASN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
14	P	3/5 (60%)	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 ⓘ

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
15	BRU	T	22	15,14	15,21,22	1.33	2 (13%)	17,30,33	2.37	4 (23%)

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
15	BRU	T	22	15,14	-	0/4/21/22	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
15	T	22	BRU	C4-C5	4.33	1.44	1.38
15	T	22	BRU	C4-N3	2.34	1.37	1.33

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
15	T	22	BRU	C4-N3-C2	8.15	122.02	115.14
15	T	22	BRU	C5-C4-N3	-4.20	118.61	123.64
15	T	22	BRU	BR-C5-C4	-2.23	118.11	121.50
15	T	22	BRU	BR-C5-C6	2.14	122.17	117.31

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates 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
18	APC	P	1011	-	27,33,33	2.65	8 (29%)	31,52,52	1.47	5 (16%)

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
18	APC	P	1011	-	-	0/15/38/38	0/3/3/3

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
18	P	1011	APC	PB-O3B	8.97	1.68	1.58
18	P	1011	APC	PA-O5'	5.03	1.64	1.57
18	P	1011	APC	PA-O1A	4.25	1.61	1.51
18	P	1011	APC	PG-O1G	4.03	1.63	1.50
18	P	1011	APC	O4'-C1'	3.39	1.45	1.41

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
18	P	1011	APC	N3-C2-N1	-3.75	122.81	128.68
18	P	1011	APC	C4-C5-N7	-3.34	105.92	109.40
18	P	1011	APC	O2G-PG-O3B	2.93	114.48	104.64
18	P	1011	APC	PG-O3B-PB	-2.34	124.38	132.62
18	P	1011	APC	O2'-C2'-C3'	2.14	118.76	111.82

There are no chirality outliers.

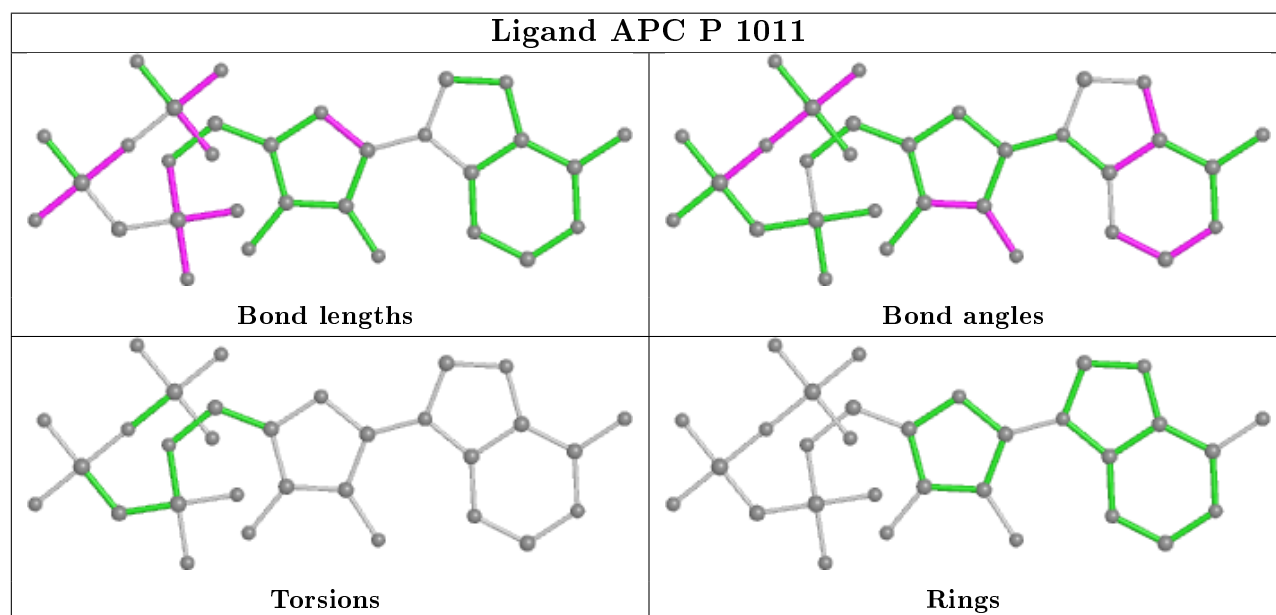
There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
18	P	1011	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 than 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.



## 5.7 Other polymers [i](#)

There are no such residues in this entry.

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

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	B	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	934:LYS	C	935:ARG	N	5.31
1	B	351:TYR	C	352:ALA	N	3.13

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	1423/1732 (82%)	0.11	35 (2%) 57 55	65, 110, 174, 230	0
2	B	1115/1224 (91%)	0.25	36 (3%) 47 46	66, 124, 185, 221	0
3	C	266/318 (83%)	-0.10	0 100 100	81, 107, 150, 180	0
4	D	178/221 (80%)	0.10	3 (1%) 70 68	86, 123, 174, 194	0
5	E	214/215 (99%)	0.16	12 (5%) 24 25	88, 150, 192, 202	0
6	F	84/155 (54%)	-0.26	0 100 100	64, 92, 119, 142	0
7	G	171/171 (100%)	0.08	0 100 100	84, 107, 145, 165	0
8	H	133/146 (91%)	0.67	14 (10%) 6 7	123, 155, 192, 207	0
9	I	119/122 (97%)	0.03	3 (2%) 57 55	114, 156, 190, 213	0
10	J	65/70 (92%)	0.05	0 100 100	88, 108, 145, 161	0
11	K	115/120 (95%)	-0.06	2 (1%) 70 68	80, 110, 147, 166	0
12	L	46/70 (65%)	0.77	6 (13%) 3 4	93, 184, 198, 207	0
13	N	8/14 (57%)	-0.17	0 100 100	214, 229, 237, 240	0
14	P	4/5 (80%)	0.07	0 100 100	146, 154, 170, 189	0
15	T	17/26 (65%)	0.46	1 (5%) 22 23	156, 210, 225, 227	0
All	All	3958/4609 (85%)	0.15	112 (2%) 53 51	64, 119, 185, 240	0

The worst 5 of 112 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
12	L	27	LEU	6.8
2	B	339	THR	5.4
1	A	194	ALA	5.2
2	B	250	PHE	5.0
2	B	340	ALA	4.6



## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
15	BRU	T	22	20/21	0.92	0.11	170,185,187,190	0

## 6.3 Carbohydrates [i](#)

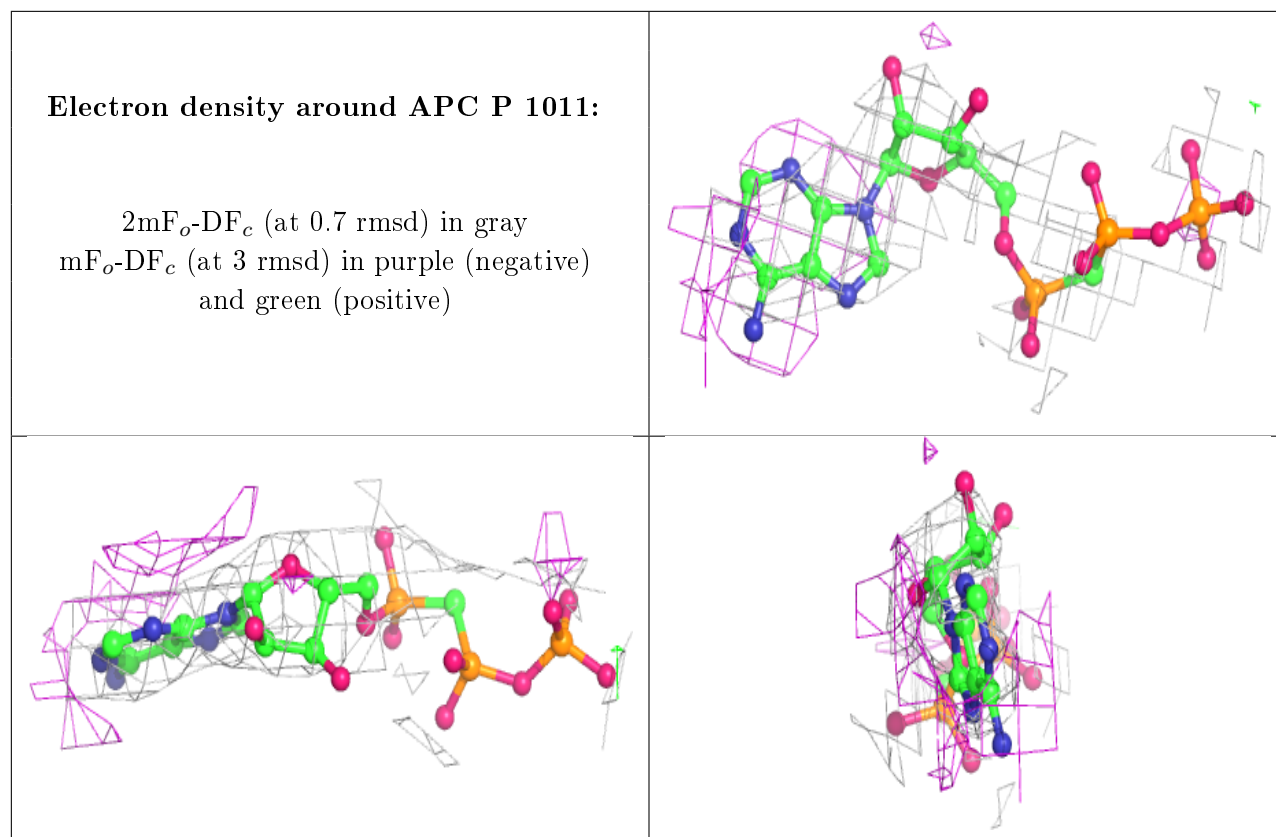
There are no carbohydrates in this entry.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
18	APC	P	1011	31/31	0.61	0.39	195,198,211,212	0
17	MG	A	2458	1/1	0.97	0.17	110,110,110,110	0
16	ZN	I	1122	1/1	0.98	0.05	191,191,191,191	0
16	ZN	A	2457	1/1	0.99	0.18	85,85,85,85	0
16	ZN	J	1066	1/1	0.99	0.23	98,98,98,98	0
16	ZN	A	2456	1/1	0.99	0.06	138,138,138,138	0
16	ZN	I	1121	1/1	0.99	0.15	130,130,130,130	0
16	ZN	L	1071	1/1	0.99	0.05	198,198,198,198	0
16	ZN	C	1269	1/1	1.00	0.16	96,96,96,96	0
16	ZN	B	2225	1/1	1.00	0.23	95,95,95,95	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.



## 6.5 Other polymers [i](#)

There are no such residues in this entry.