



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 11, 2019 – 04:51 AM EST

PDB ID : 4A3L  
Title : RNA Polymerase II initial transcribing complex with a 7nt DNA-RNA hybrid and soaked with AMPCPP  
Authors : Cheung, A.C.M.; Sainsbury, S.; Cramer, P.  
Deposited on : 2011-09-30  
Resolution : 3.50 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 : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : rb-20031633  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20031633

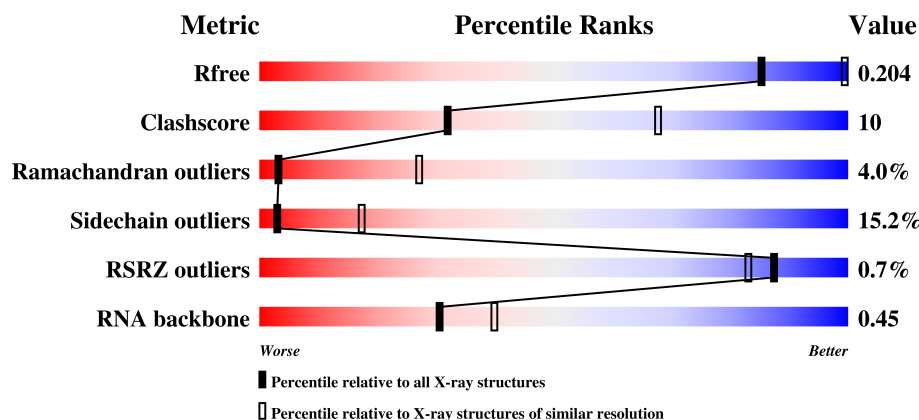
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.50 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))
$R_{free}$	111664	1391 (3.60-3.40)
Clashscore	122126	1485 (3.60-3.40)
Ramachandran outliers	120053	1446 (3.60-3.40)
Sidechain outliers	120020	1447 (3.60-3.40)
RSRZ outliers	108989	1303 (3.60-3.40)
RNA backbone	2636	1052 (4.10-2.90)

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. 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	<div> <div>55%</div> <div>22%</div> <div>5%</div> <div>•</div> <div>18%</div> </div>
2	B	1224	<div> <div>59%</div> <div>28%</div> <div>•</div> <div>9%</div> </div>
3	C	318	<div> <div>56%</div> <div>25%</div> <div>•</div> <div>16%</div> </div>
4	D	221	<div> <div>50%</div> <div>27%</div> <div>•</div> <div>19%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	E	215	 70% 26% .
6	F	155	 39% 14% . 46%
7	G	171	 61% 31% 8%
8	H	146	 2% 58% 27% 6% 9%
9	I	122	 2% 70% 24% . .
10	J	70	 57% 27% 9% 7%
11	K	120	 56% 35% 5% .
12	L	70	 6% 31% 24% 10% 34%
13	N	14	 21% 43% 36%
14	P	7	 57% 43%
15	T	26	 46% 23% 31%

## 2 Entry composition [i](#)

There are 18 unique types of molecules in this entry. The entry contains 31950 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
1	A	1425	Total	C	N	O	S	0	0	0
			11197	7051	1958	2126	62			

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

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

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

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

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

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

- Molecule 5 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC1.

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

- Molecule 6 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC2.

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

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

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

- Molecule 8 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC3.

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

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

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

- Molecule 10 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC5.

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

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

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

- Molecule 12 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC4.

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	N	9	Total	C	N	O	P	0	0	0
			183	89	34	52	8			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
14	P	7	Total	C	N	O	P	0	0	0
			152	68	31	46	7			

- Molecule 15 is a DNA chain called 5'-D(\*AP\*GP\*CP\*TP\*CP\*AP\*AP\*GP\*TP\*AP\*CP\*T  
P\*TP\*TP \*TP\*TP\*CP\*CP\*BRU\*GP\*GP\*TP\*CP\*AP\*TP\*T)-3'.

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

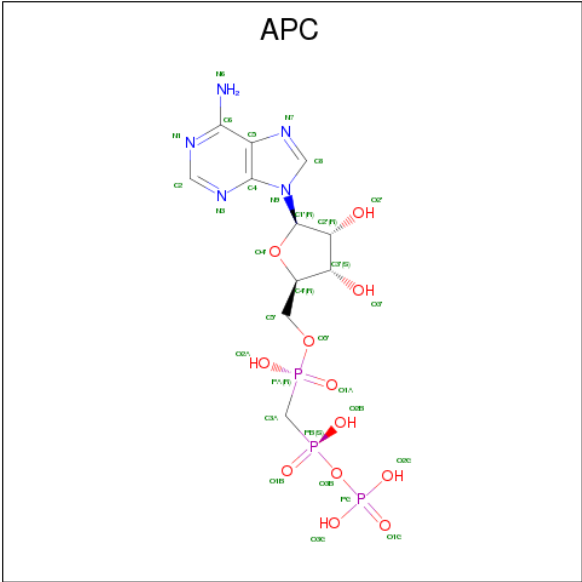
- Molecule 16 is ZINC ION (three-letter code: ZN) (formula: Zn).

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>).

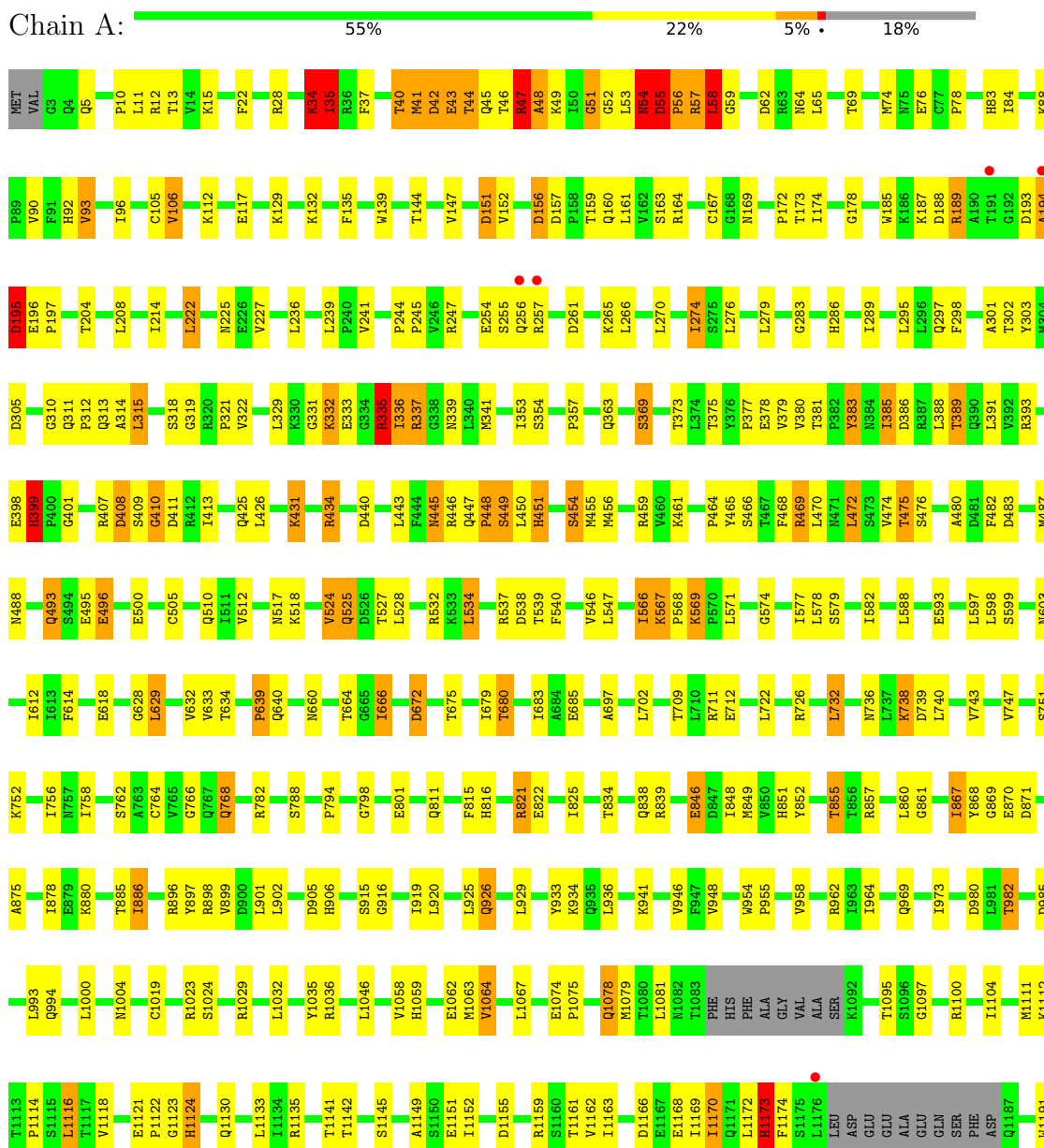


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
18	P	1	Total	C	N	O	P	0	0
			31	11	5	12	3		

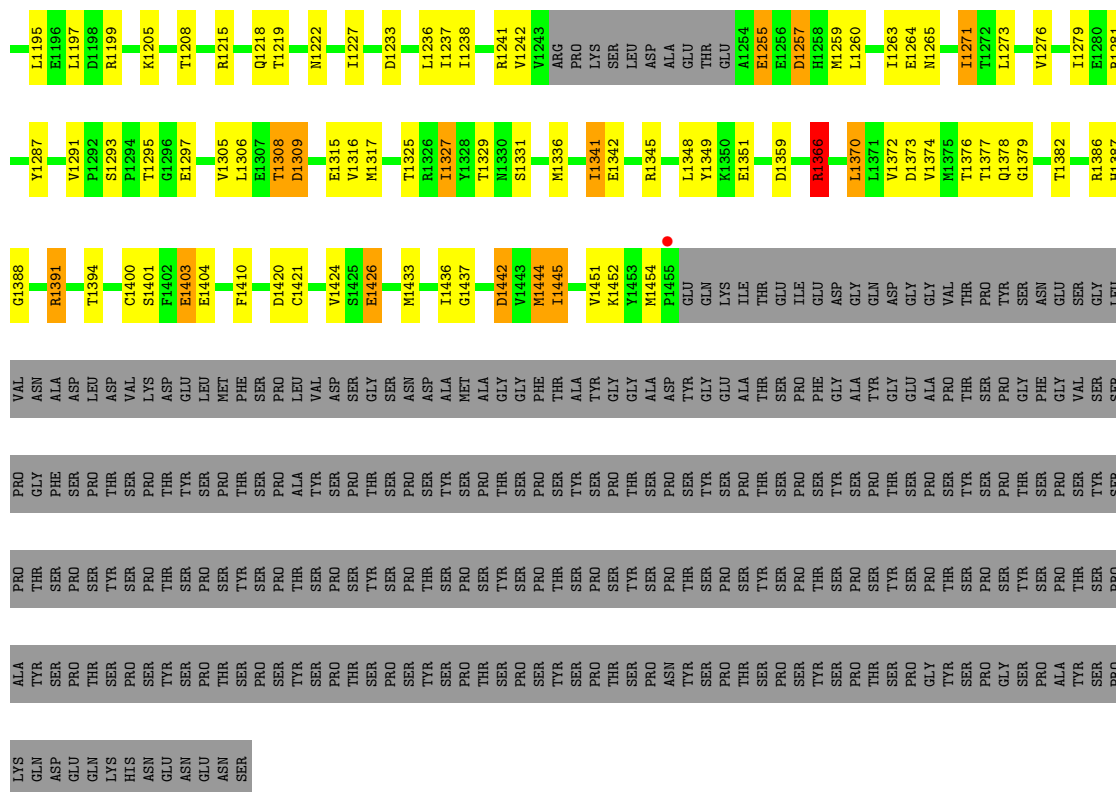
### 3 Residue-property plots

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 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.

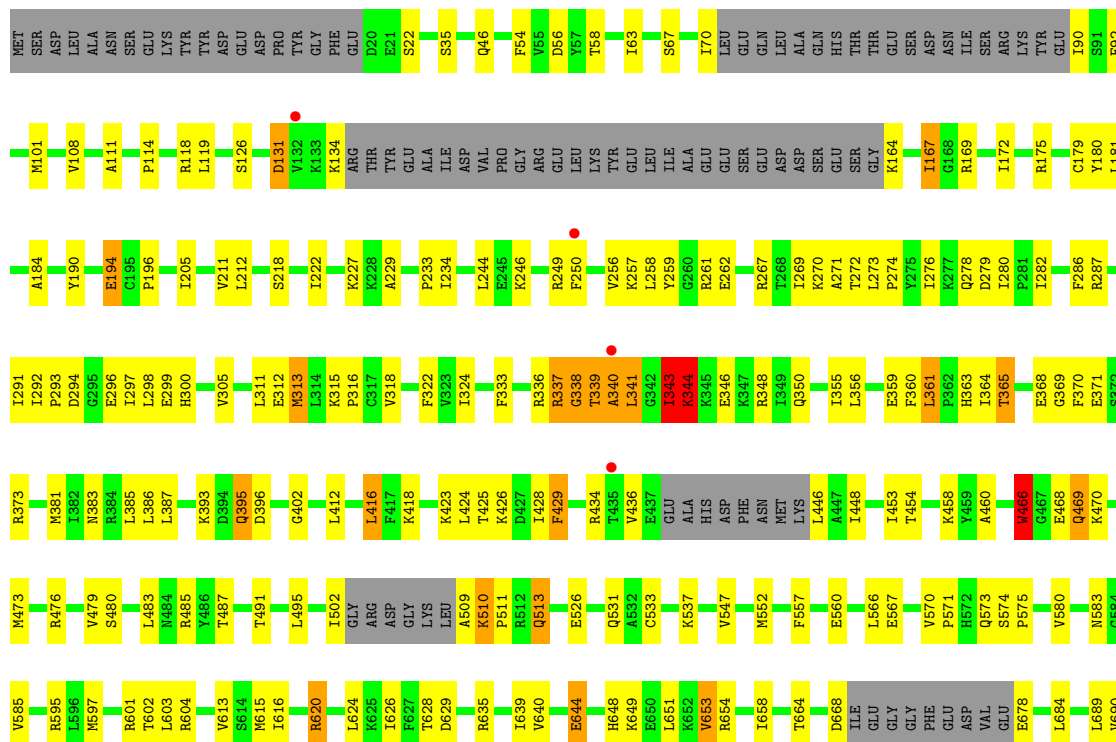
#### • Molecule 1: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB1

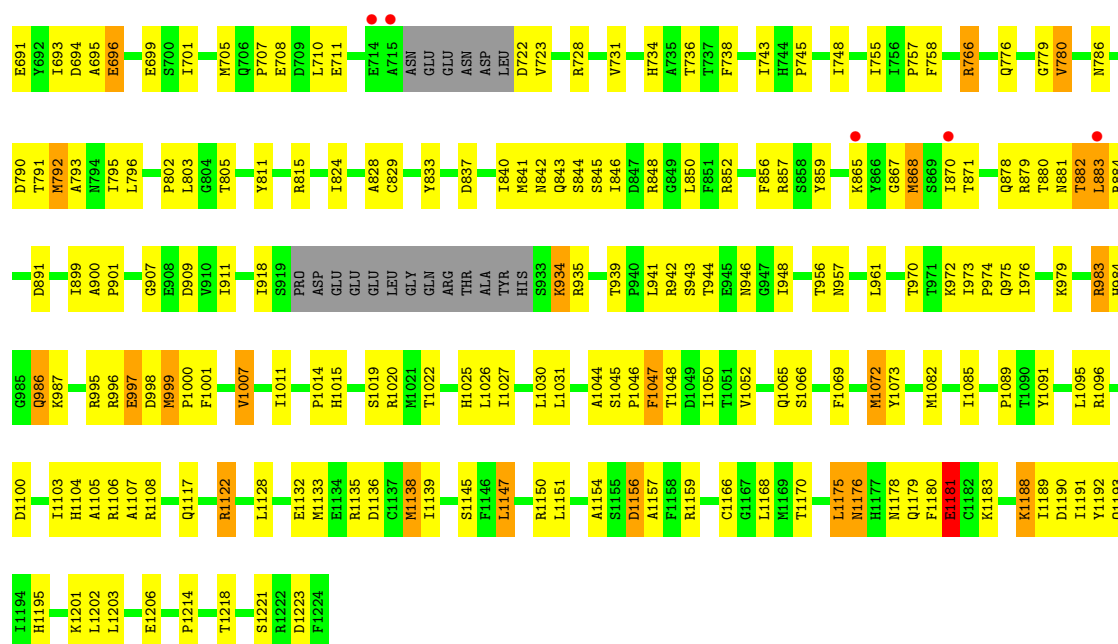






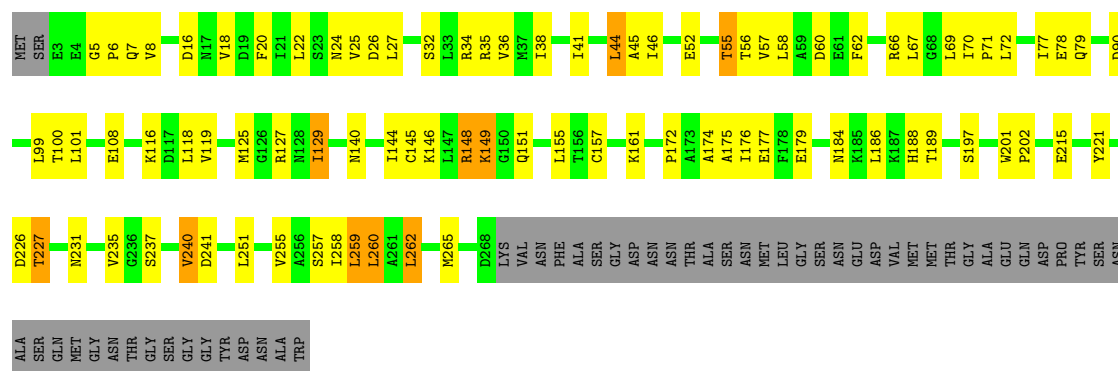
### ● Molecule 2: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB2





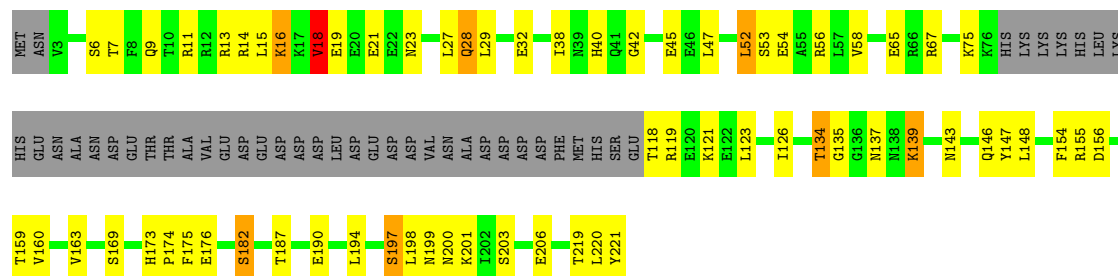
• Molecule 3: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB3

Chain C: 56% 25% 16%



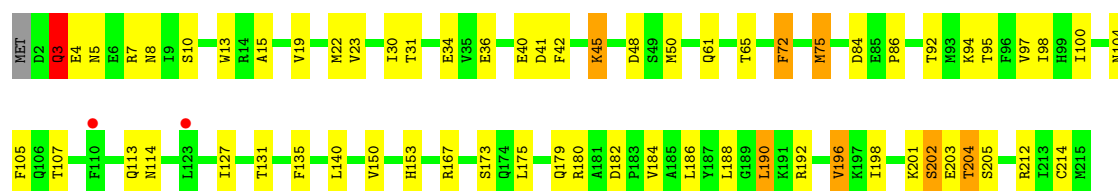
• Molecule 4: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB4

Chain D: 50% 27% 19%



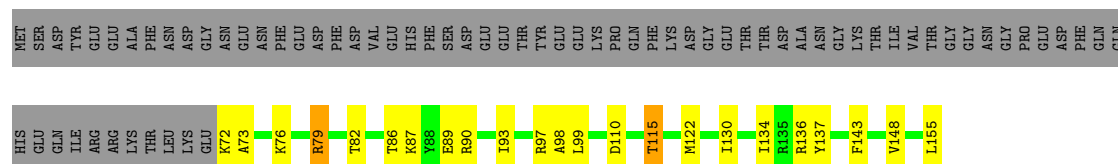
• Molecule 5: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC1

Chain E: 70% 26% 4%



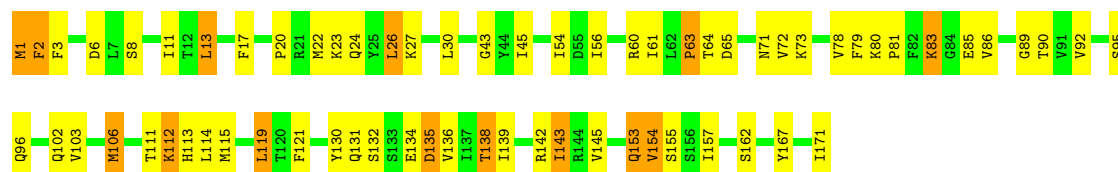
- Molecule 6: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC2

Chain F: 39% 14% 46%



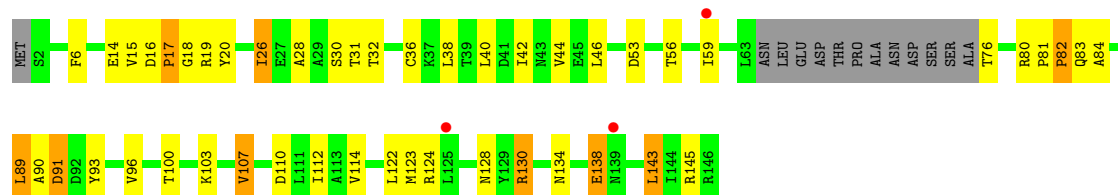
- Molecule 7: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB7

Chain G: 61% 31% 8%



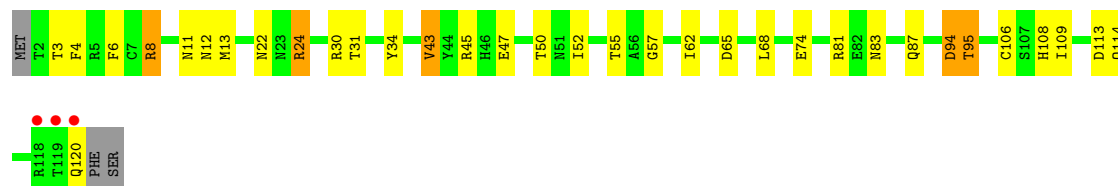
- Molecule 8: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC3

Chain H: 2% 58% 27% 6% 9%



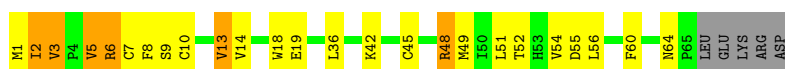
- Molecule 9: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB9

Chain I: 2% 70% 24% 4% 1%

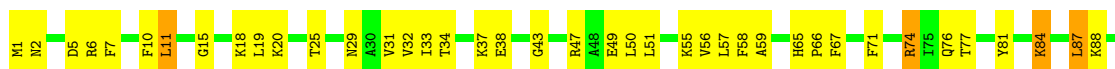


- Molecule 10: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC5

Chain J: 57% 27% 9% 7%



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



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	222.81Å 392.44Å 279.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.82 – 3.50 47.82 – 3.50	Depositor EDS
% Data completeness (in resolution range)	95.4 (47.82-3.50) 95.4 (47.82-3.50)	Depositor EDS
$R_{merge}$	0.77	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.83 (at 3.48Å)	Xtriage
Refinement program	BUSTER 2.11.2	Depositor
R, $R_{free}$	0.159 , 0.187 0.182 , 0.204	Depositor DCC
$R_{free}$ test set	2876 reflections (1.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	107.7	Xtriage
Anisotropy	0.359	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 99.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	0.039 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.048 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	31950	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	123.0	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.*

<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	1/11397 (0.0%)	0.85	8/15415 (0.1%)
2	B	0.51	0/9029	0.81	1/12171 (0.0%)
3	C	0.50	0/2133	0.80	0/2891
4	D	0.54	0/1444	0.88	0/1935
5	E	0.48	0/1788	0.73	0/2406
6	F	0.57	0/691	0.78	0/933
7	G	0.52	0/1368	0.83	0/1844
8	H	0.51	0/1086	0.85	0/1470
9	I	0.46	0/989	0.77	0/1331
10	J	0.52	0/541	0.85	1/727 (0.1%)
11	K	0.51	0/938	0.75	0/1267
12	L	0.53	0/365	0.96	0/485
13	N	1.13	0/205	1.07	0/315
14	P	1.19	0/170	0.86	0/263
15	T	1.33	0/383	1.10	0/586
All	All	0.55	1/32527 (0.0%)	0.83	10/44039 (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 (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	867	ILE	CG1-CD1	5.18	1.86	1.50

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	34	LYS	C-N-CA	7.75	141.08	121.70
1	A	399	HIS	N-CA-CB	6.86	122.95	110.60
1	A	56	PRO	C-N-CA	6.39	137.68	121.70
1	A	35	ILE	N-CA-CB	6.21	125.09	110.80
1	A	54	ASN	C-N-CA	5.83	136.28	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	11197	0	11257	246	0
2	B	8859	0	8901	170	0
3	C	2095	0	2051	53	0
4	D	1434	0	1460	27	0
5	E	1752	0	1776	35	0
6	F	679	0	701	15	0
7	G	1340	0	1357	45	0
8	H	1068	0	1040	20	0
9	I	971	0	927	15	0
10	J	532	0	542	18	0
11	K	920	0	929	35	0
12	L	363	0	386	14	0
13	N	183	0	104	6	0
14	P	152	0	77	1	0
15	T	365	0	204	6	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
16	L	1	0	0	0	0
17	A	1	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
18	P	31	0	14	1	0
All	All	31950	0	31726	611	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 10.

The worst 5 of 611 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	1.86	1.50
6:F:76:LYS:HA	6:F:79:ARG:HD3	1.49	0.94
2:B:766:ARG:HE	2:B:1020:ARG:HG2	1.38	0.88
1:A:53:LEU:HD23	1:A:54:ASN:H	1.38	0.87
1:A:868:TYR:CE1	1:A:1064:VAL:HG11	2.10	0.86

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	1417/1732 (82%)	1225 (86%)	132 (9%)	60 (4%)	3	26
2	B	1095/1224 (90%)	960 (88%)	86 (8%)	49 (4%)	3	24
3	C	264/318 (83%)	234 (89%)	26 (10%)	4 (2%)	11	48
4	D	174/221 (79%)	151 (87%)	14 (8%)	9 (5%)	2	20
5	E	212/215 (99%)	195 (92%)	12 (6%)	5 (2%)	6	38
6	F	82/155 (53%)	72 (88%)	10 (12%)	0	100	100
7	G	169/171 (99%)	153 (90%)	13 (8%)	3 (2%)	9	44
8	H	129/146 (88%)	103 (80%)	16 (12%)	10 (8%)	1	11

*Continued on next page...*



Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
9	I	117/122 (96%)	93 (80%)	20 (17%)	4 (3%)	4	31
10	J	63/70 (90%)	54 (86%)	6 (10%)	3 (5%)	2	22
11	K	113/120 (94%)	103 (91%)	9 (8%)	1 (1%)	19	60
12	L	44/70 (63%)	22 (50%)	15 (34%)	7 (16%)	0	2
All	All	3879/4564 (85%)	3365 (87%)	359 (9%)	155 (4%)	3	27

5 of 155 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	57	ARG
1	A	58	LEU
1	A	76	GLU
1	A	169	ASN
1	A	257	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	1243/1519 (82%)	1046 (84%)	197 (16%)	3	16
2	B	966/1061 (91%)	824 (85%)	142 (15%)	3	18
3	C	234/274 (85%)	203 (87%)	31 (13%)	4	22
4	D	160/200 (80%)	134 (84%)	26 (16%)	2	14
5	E	196/197 (100%)	170 (87%)	26 (13%)	4	22
6	F	74/137 (54%)	66 (89%)	8 (11%)	7	31
7	G	152/152 (100%)	125 (82%)	27 (18%)	2	10
8	H	117/128 (91%)	100 (86%)	17 (14%)	3	19
9	I	113/116 (97%)	98 (87%)	15 (13%)	4	22
10	J	60/65 (92%)	52 (87%)	8 (13%)	4	22
11	K	99/102 (97%)	81 (82%)	18 (18%)	2	10
12	L	40/57 (70%)	31 (78%)	9 (22%)	1	5

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	3454/4008 (86%)	2930 (85%)	524 (15%)	3 17

5 of 524 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	416	LEU
2	B	878	GLN
9	I	94	ASP
2	B	446	LEU
2	B	620	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	843	GLN
2	B	1084	GLN
9	I	11	ASN
2	B	862	GLN
2	B	975	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
14	P	6/7 (85%)	1 (16%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
14	P	8	G

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	14,21,22	1.61	4 (28%)	17,30,33	3.28	5 (29%)

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	-	0/3/21/22	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
15	T	22	BRU	O5'-C5'	-3.51	1.39	1.44
15	T	22	BRU	BR-C5	-2.61	1.83	1.90
15	T	22	BRU	C2-N3	-2.61	1.33	1.38
15	T	22	BRU	C6-C5	-2.23	1.34	1.39

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
15	T	22	BRU	C5-C4-N3	-6.66	115.67	123.64
15	T	22	BRU	BR-C5-C4	-3.78	115.77	121.50
15	T	22	BRU	C2'-C1'-N1	-3.11	107.01	114.27
15	T	22	BRU	O4'-C1'-N1	4.12	114.73	107.78
15	T	22	BRU	C4-N3-C2	9.58	123.30	115.14

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

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

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	-	28,33,33	3.68	11 (39%)	27,52,52	2.23	8 (29%)

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 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
18	P	1011	APC	C4-N3	-2.85	1.31	1.35
18	P	1011	APC	C8-N7	-2.64	1.29	1.34
18	P	1011	APC	C5'-C4'	2.00	1.58	1.51
18	P	1011	APC	PG-O2G	2.82	1.66	1.54
18	P	1011	APC	O4'-C1'	3.62	1.46	1.41

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
18	P	1011	APC	N3-C2-N1	-6.73	123.10	128.86
18	P	1011	APC	C4-C5-N7	-4.58	104.99	109.41
18	P	1011	APC	O1A-PA-C3A	-2.36	103.06	109.15
18	P	1011	APC	PG-O3B-PB	-2.14	124.89	132.40
18	P	1011	APC	O3'-C3'-C2'	2.72	120.55	111.83

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
18	P	1011	APC	1	0

## 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	5.76
1	B	351:TYR	C	352:ALA	N	3.03

## 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	1425/1732 (82%)	-0.30	6 (0%) 92 90	58, 109, 170, 229	0
2	B	1115/1224 (91%)	-0.20	9 (0%) 86 80	60, 125, 185, 220	0
3	C	266/318 (83%)	-0.34	0 100 100	79, 108, 151, 176	0
4	D	178/221 (80%)	-0.30	0 100 100	80, 120, 168, 185	0
5	E	214/215 (99%)	-0.27	2 (0%) 84 78	84, 146, 186, 199	0
6	F	84/155 (54%)	-0.51	0 100 100	66, 90, 120, 142	0
7	G	171/171 (100%)	-0.37	0 100 100	84, 106, 146, 169	0
8	H	133/146 (91%)	0.21	3 (2%) 60 53	125, 153, 187, 208	0
9	I	119/122 (97%)	-0.09	3 (2%) 57 50	116, 161, 191, 203	0
10	J	65/70 (92%)	-0.40	0 100 100	85, 104, 151, 156	0
11	K	115/120 (95%)	-0.36	1 (0%) 84 78	74, 107, 143, 161	0
12	L	46/70 (65%)	0.34	4 (8%) 10 11	88, 166, 186, 202	0
13	N	9/14 (64%)	0.10	0 100 100	202, 213, 240, 250	0
14	P	7/7 (100%)	-0.52	0 100 100	100, 112, 162, 182	0
15	T	17/26 (65%)	-0.25	0 100 100	92, 153, 246, 250	0
All	All	3964/4611 (85%)	-0.25	28 (0%) 87 83	58, 117, 181, 250	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1176	LEU	5.9
12	L	26	THR	4.9
11	K	115	ALA	4.7
12	L	25	ALA	4.4
9	I	119	THR	4.2

## 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	0.96	0.13	100,112,119,124	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	0.82	0.27	148,151,173,174	0
17	MG	A	2458	1/1	0.96	0.11	95,95,95,95	0
16	ZN	I	1122	1/1	0.96	0.03	202,202,202,202	0
16	ZN	L	1071	1/1	0.96	0.04	203,203,203,203	0
16	ZN	C	1269	1/1	0.99	0.11	91,91,91,91	0
16	ZN	A	2456	1/1	0.99	0.10	146,146,146,146	0
16	ZN	A	2457	1/1	1.00	0.14	76,76,76,76	0
16	ZN	I	1121	1/1	1.00	0.13	137,137,137,137	0
16	ZN	B	2225	1/1	1.00	0.18	85,85,85,85	0
16	ZN	J	1066	1/1	1.00	0.24	98,98,98,98	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.