wwPDB X-ray Structure Validation Summary Report (i)

Apr 17, 2024 – 12:40 PM EDT

PDB ID	:	8UKR
Title	:	RNA polymerase II elongation complex with Fapy-dG lesion soaking with ATP before chemistry
Authors	:	Hou, P.; Oh, J.; Wang, D.
Deposited on	:	2023-10-15
Resolution	:	3.78 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:

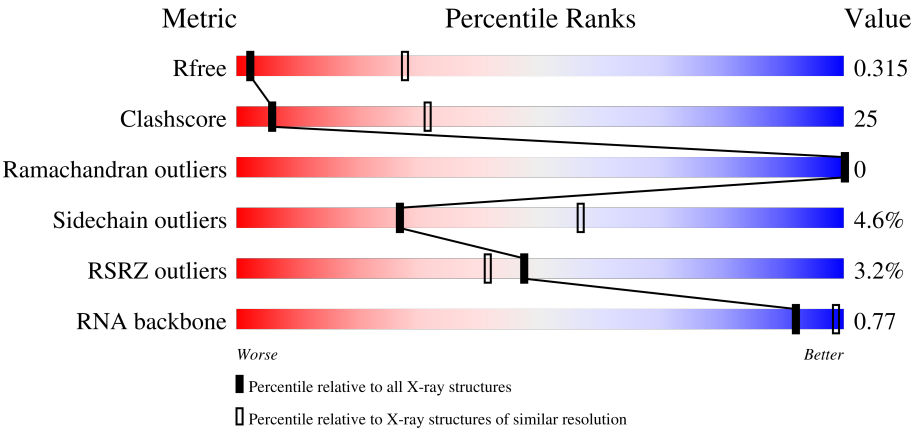
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.1

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 3.78 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	1038 (3.96-3.60)
Clashscore	141614	1100 (3.96-3.60)
Ramachandran outliers	138981	1062 (3.96-3.60)
Sidechain outliers	138945	1058 (3.96-3.60)
RSRZ outliers	127900	1009 (3.98-3.58)
RNA backbone	3102	1035 (4.52-3.00)

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	33% 33% 33%
2	T	29	10% 62% 10% 17%
3	N	18	6% 61% 6% 28%
4	A	1733	3% 40% 38% · 20%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	B	1224	2% 46% 44% 8%
6	C	318	2% 45% 37% 16%
7	E	215	7% 47% 47%
8	F	155	% 28% 26% 45%
9	H	146	6% 42% 45% 9%
10	I	122	5% 40% 51% 6%
11	J	70	43% 43% 7% 7%
12	K	120	% 50% 42% 5%
13	L	70	9% 24% 29% 6% 39%

2 Entry composition (i)

There are 16 unique types of molecules in this entry. The entry contains 28985 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
1	R	9	Total	C	N	O	P	0	0	0
			194	88	40	58	8			

- Molecule 2 is a DNA chain called tsDNA with Fapy-dG lesion.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	24	Total	C	N	O	P	0	0	0
			481	230	76	151	24			

- Molecule 3 is a DNA chain called ntsDNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	N	13	Total	C	N	O	P	0	0	0
			275	128	61	73	13			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	1384	Total	C	N	O	S	0	0	0
			10828	6831	1896	2041	60			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	B	1123	Total	C	N	O	S	0	0	0
			8859	5607	1552	1647	53			

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	E	212	Total	C	N	O	S	0	0	0
			1724	1094	305	314	11			

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	H	133	Total	C	N	O	S	0	0	0
			1064	670	179	211	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	I	118	Total	C	N	O	S	0	0	0
			952	585	173	184	10			

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

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

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	L	43	Total	C	N	O	S	0	0	0
			332	205	64	59	4			

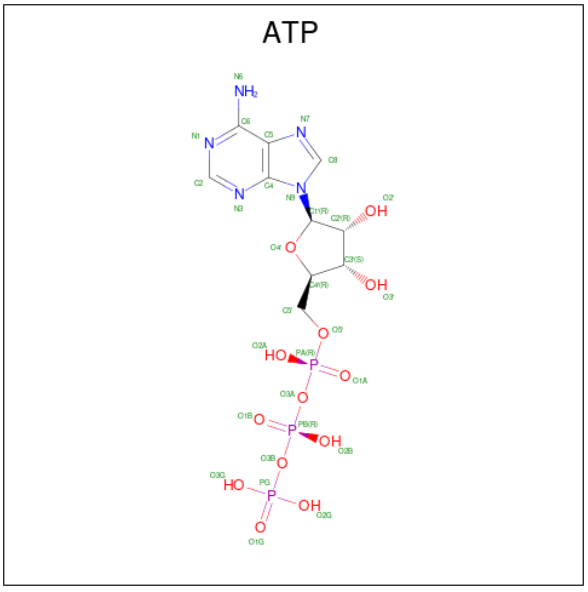
- Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

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

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

- Molecule 16 is ADENOSINE-5'-TRIPHOSPHATE (three-letter code: ATP) (formula: $C_{10}H_{16}N_5O_{13}P_3$) (labeled as "Ligand of Interest" by depositor).



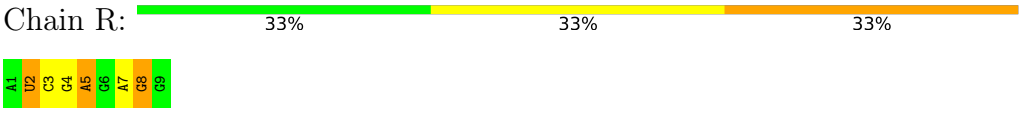
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
16	B	1	Total C N O P 31 10 5 13 3	0	0

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.

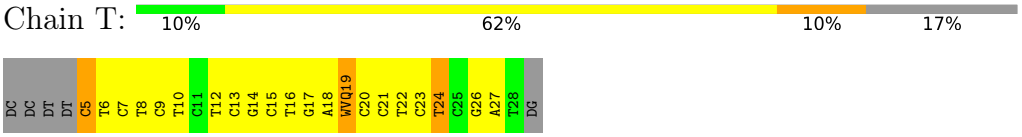
• Molecule 1: RNA

Chain R:



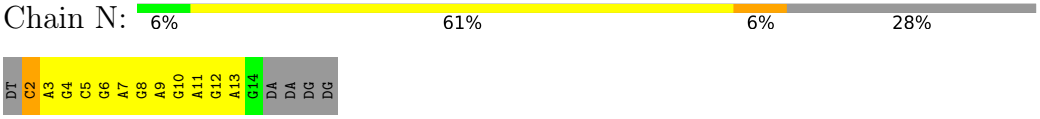
• Molecule 2: tsDNA with Fapy-dG lesion

Chain T:



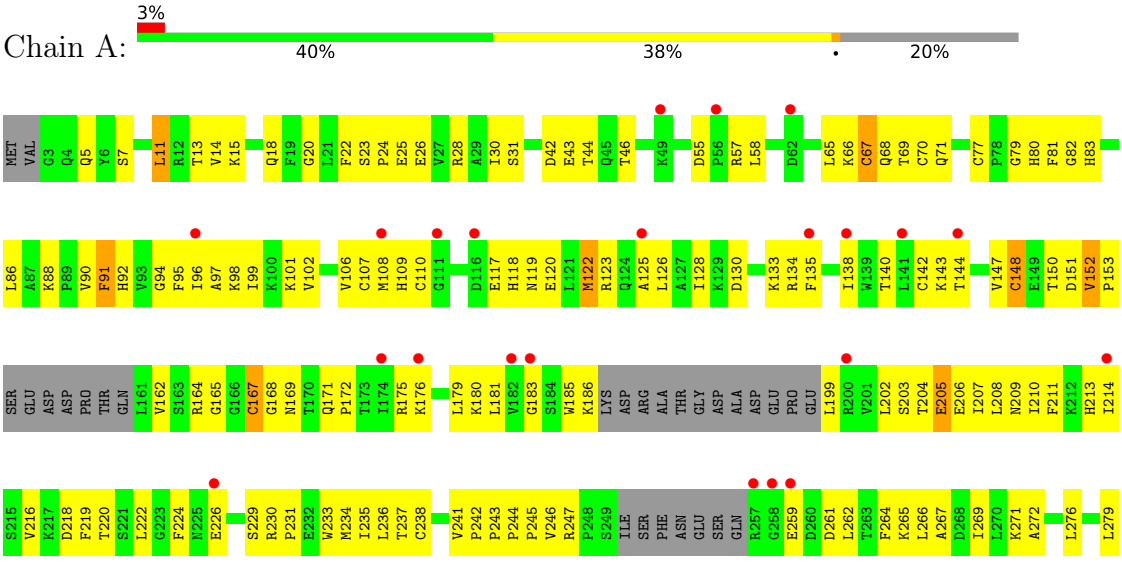
• Molecule 3: ntsDNA

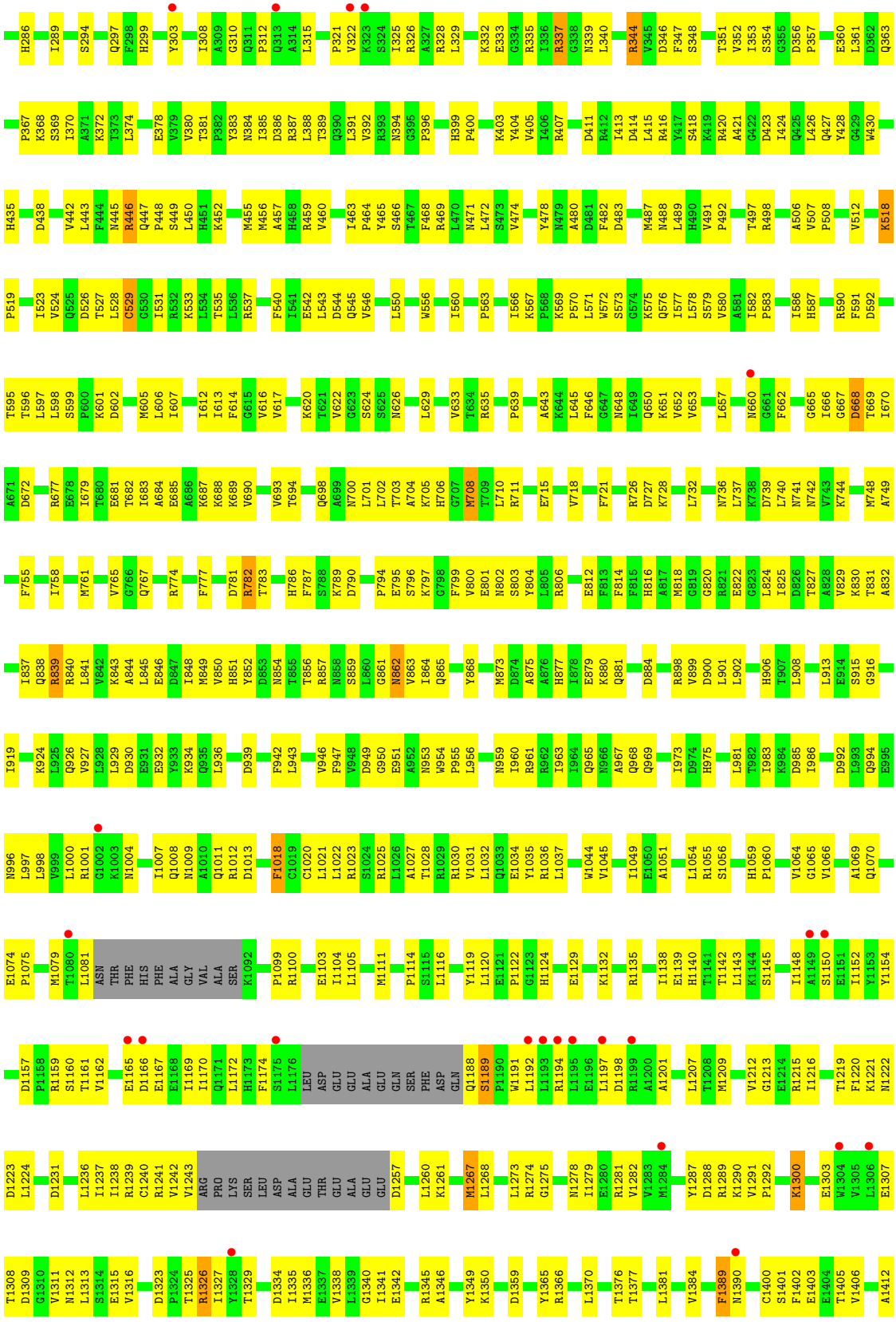
Chain N:

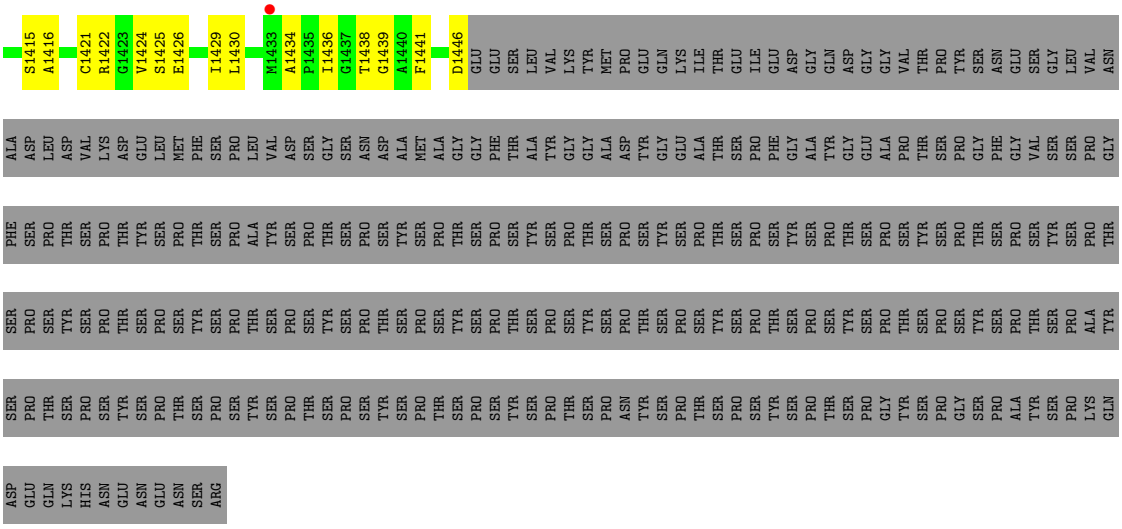


• Molecule 4: DNA-directed RNA polymerase II subunit RPB1

Chain A:

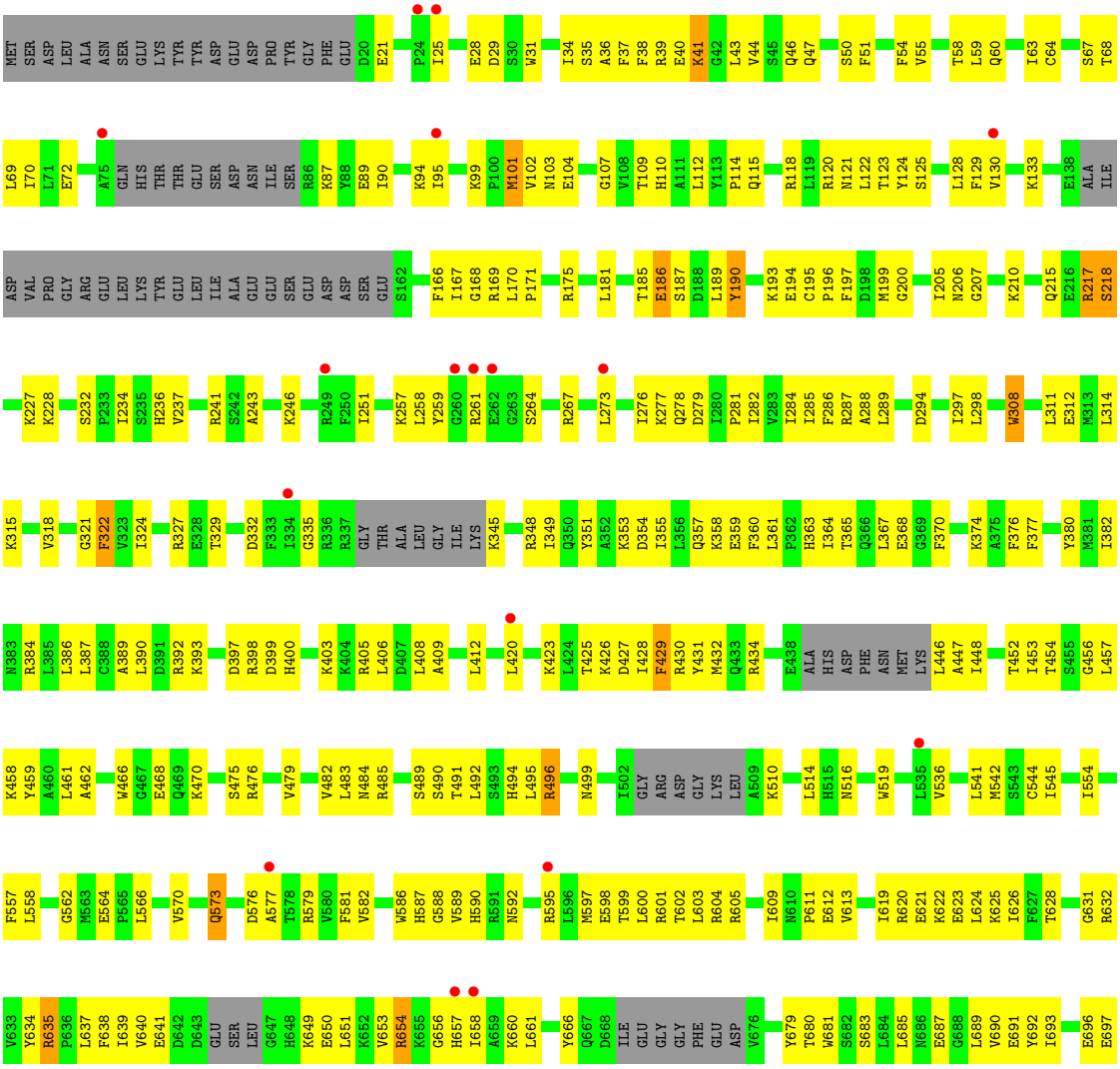


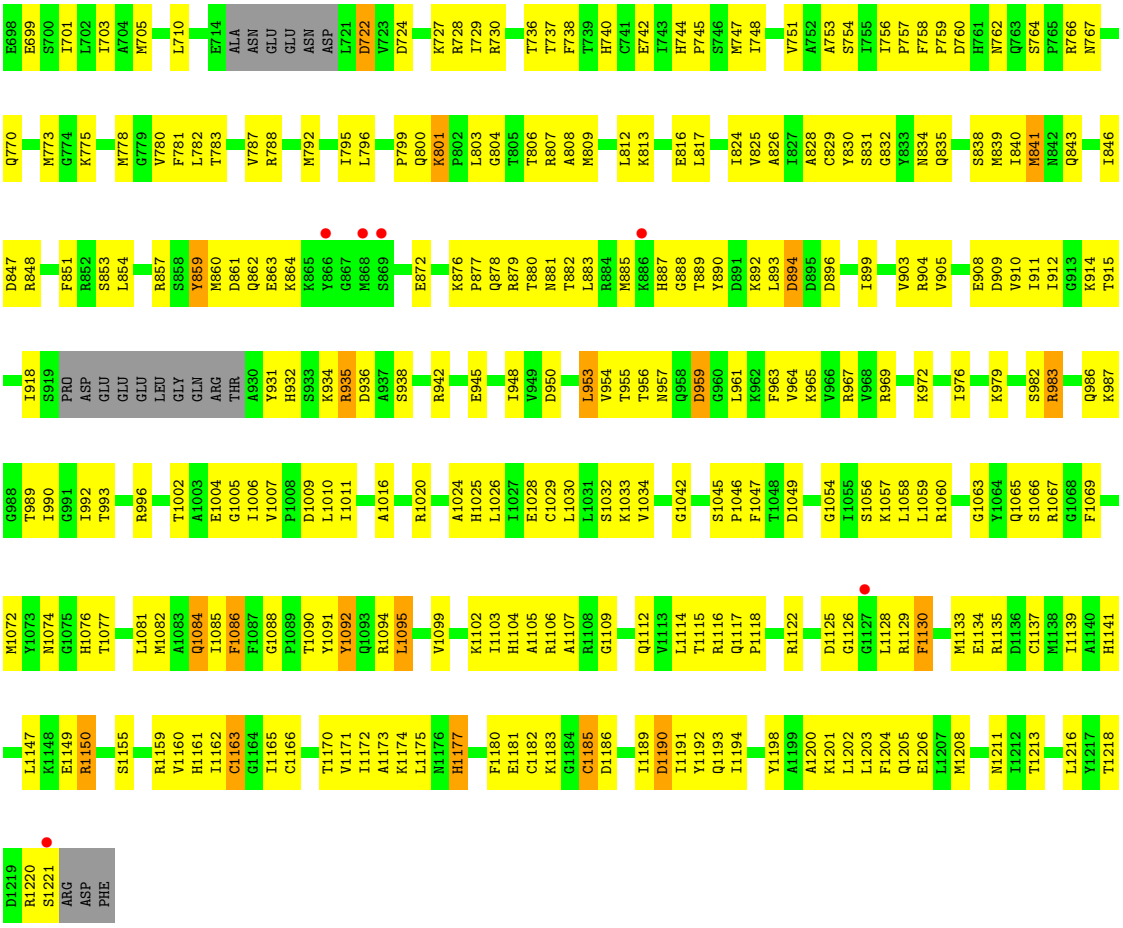

H286 I289 S294 Q297 F298 H299 Y303 I308 A309 G310 Q311 P312 Q313 A314 L315 P321 V322 K323 S324 I325 R326 A327 R328 L329 K332 E333 G334 R335 I336 R337 G338 N339 L340 R344 V345 D346 F347 S348 T351 V352 I353 S354 G355 D356 P357 E360 L361 D362 Q363
P367 K368 S369 I370 A371 K372 T373 L374 E378 V379 V380 T381 P382 Y383 N384 I385 D386 R387 L388 T389 Q390 L391 V392 R393 N394 G395 P396 H399 P400 K403 Y404 V405 I406 R407 D411 R412 I413 D414 L415 R416 Y417 S418 K419 R420 A421 G422 D423 I424 Q425 L426 Q427 Y428 G429 W430
H435 D438 V442 L443 F444 N445 R446 Q447 P448 S449 L450 H451 K452 M455 M456 A457 H458 R459 V460 I463 P464 Y465 S466 T467 F468 R469 L470 N471 L472 S473 V474 Y478 N479 A480 D481 F482 D483 M487 N488 L489 H490 V491 P492 T497 R498 A506 V507 P508 V512 K518
P519 I523 V524 Q525 D526 T527 L528 C529 G530 I531 R532 K533 L534 T535 L536 R537 F540 I541 E542 L543 D544 Q545 V546 L550 W556 I560 N471 P563 I566 K567 P568 K569 P570 L571 W572 S573 G574 K575 Q576 I577 L578 S579 V580 A581 I582 P583 I586 H587 R590 F591 D592
T595 T596 L597 L598 S599 P600 K601 D602 M605 L606 I607 I612 I613 F614 G615 V616 V617 K620 T621 V622 G623 S624 S625 N626 L629 V633 T634 R635 P639 A643 K644 L645 F646 G647 N648 I649 Q650 K651 V652 V653 L657 N660 G661 F662 G665 I666 G667 D668 T669 I670
A671 D672 R677 E678 I679 T680 E681 T682 I683 A684 E685 A686 K687 K688 K689 V690 V693 T694 Q698 A699 N700 L701 L702 T703 A704 K705 H706 G707 M708 T709 L710 R711 E715 V718 F721 R726 D727 K728 L732 N736 L737 K738 D739 L740 N741 N742 V743 K744 M748 A749
F755 I758 M761 V765 G766 Q767 R774 F777 D781 R782 T783 H786 F787 S788 K789 D790 P794 E795 S796 K797 G798 F799 V800 E801 N802 S803 Y804 L805 R806 E812 F813 F814 F815 H816 A817 M818 G819 G820 R821 E822 G823 L824 I825 D826 T827 A828 V829 K830 T831 A832
I837 Q838 R839 R840 L841 V842 K843 A844 L845 E846 D847 I848 M849 V850 H851 Y852 D853 N854 T855 T856 R857 N858 S859 L860 G861 N862 V863 I864 Q865 Y868 M873 D874 A875 A876 H877 I878 E879 K880 Q881 D884 R898 V899 D900 L901 L902 H906 T907 L908 L913 E914 S915 G916
I919 K924 L925 Q926 Y927 L928 L929 D930 E931 E932 Y933 K934 Q935 L936 D939 F942 L943 V946 F947 Y948 D949 G950 E951 A952 N953 W954 P955 L956 N959 I960 R961 R962 I963 I964 Q965 N966 A967 Q968 Q969 I973 D974 H975 L981 T982 I983 K984 D985 I986 D992 L993 Q994 E995
N996 L997 L998 V999 L1000 R1001 G1002 K1003 N1004 I1007 Q1008 N1009 A1010 Q1011 R1012 D1013 F1018 C1019 C1020 L1021 L1022 R1023 S1024 R1025 L1026 A1027 T1028 R1029 R1030 V1031 L1032 Q1033 E1034 Y1035 R1036 L1037 W1044 V1045 I1049 E1050 A1051 L1054 R1055 S1056 H1059 P1060 V1064 G1065 V1066 A1069 Q1070
E1074 P1075 M1079 T1080 L1081 ASN THR PHE HIS PHE ALA GLY VAL ALA SER K1092 P1099 R1100 E1103 I1104 L1105 M1111 P1114 S1115 L1116 Y1119 L1120 E1121 P1122 G1123 H1124 E1129 K1132 R1135 I1138 E1139 H1140 T1141 T1142 L1143 K1144 S1145 I1148 A1149 S1150 E1151 T1152 Y1153 Y1154
D1157 P1158 R1159 S1160 T1161 V1162 E1165 D1166 E1167 E1168 I1169 I1170 Q1171 L1172 H1173 F1174 S1175 L1176 LEU ASP GLU GLU ALA GLU GLN SER PHE ASP GLN Q1188 S1189 P1190 W1191 L1192 L1193 R1194 L1195 E1196 L1197 D1198 R1199 A1200 A1201 L1207 T1208 M1209 V1212 G1213 E1214 R1215 I1216 T1219 F1220 K1221 N1222
D1223 L1224 D1231 L1236 I1237 I1238 R1239 C1240 R1241 V1242 V1243 ARG PRO LYS SER LEU ASP ALA GLU THR GLU ALA GLU GLU D1257 L1260 K1261 M1267 L1268 L1273 R1274 G1275 N1278 I1279 E1280 R1281 V1282 V1283 M1284 Y1287 D1288 R1289 K1290 V1291 P1292 K1300 E1303 W1304 Y1305 L1306 E1307
T1308 D1309 G1310 V1311 N1312 L1313 S1314 E1315 V1316 D1323 P1324 T1325 R1326 I1327 Y1328 T1329 D1334 I1335 M1336 E1337 V1338 L1339 G1340 I1341 E1342 R1345 A1346 Y1349 K1350 D1359 Y1365 R1366 L1370 T1376 T1377 L1381 V1384 F1389 N1390 C1400 S1401 F1402 E1403 E1404 T1405 V1406 A1412



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

Chain B:



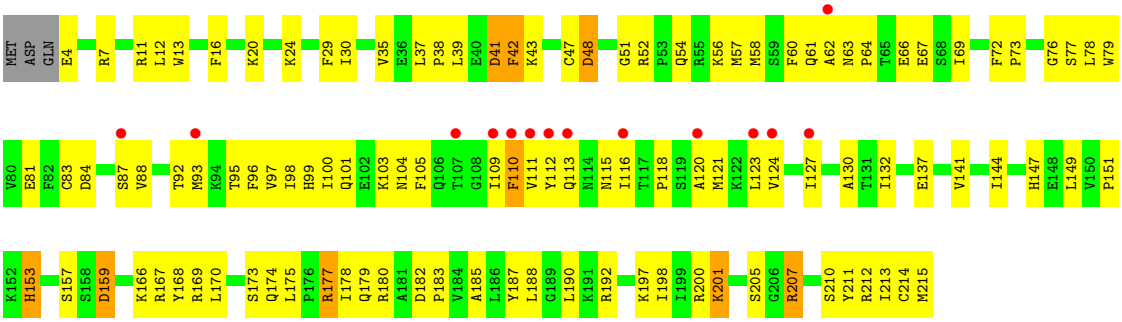


• Molecule 6: DNA-directed RNA polymerase II subunit RPB3

Chain C:

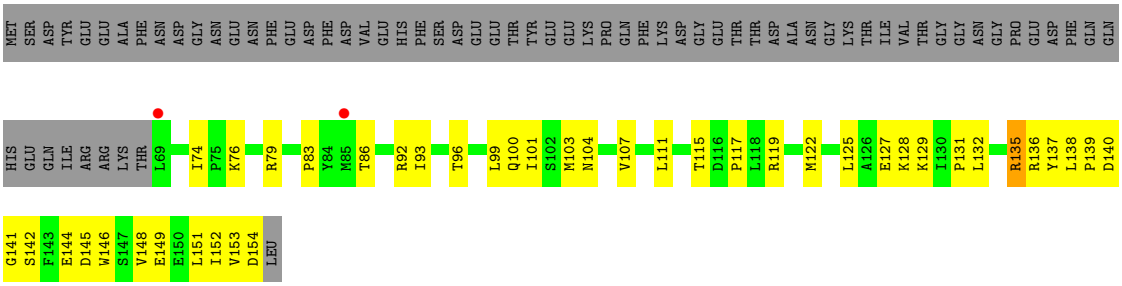
• Molecule 7: DNA-directed RNA polymerases I, II, and III subunit RPABC1

Chain E:



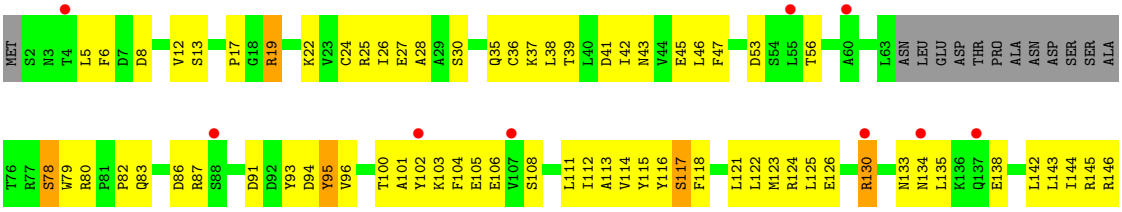
• Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC2

Chain F:



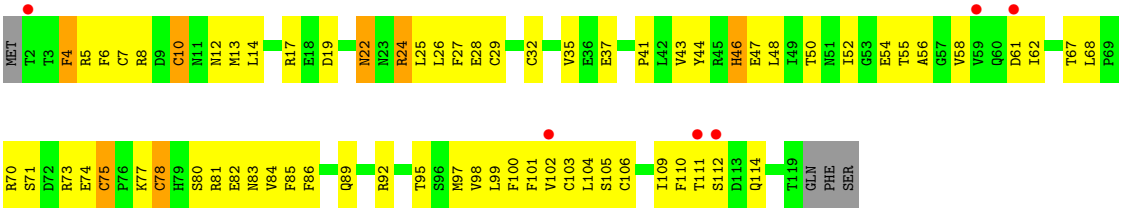
• Molecule 9: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H:



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

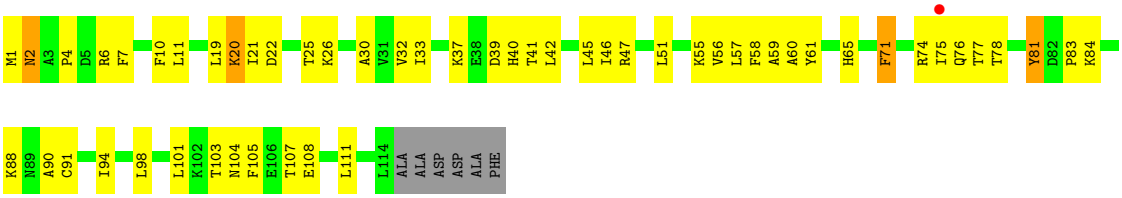
Chain I:



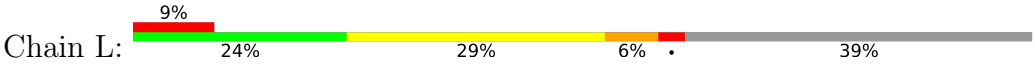
• Molecule 11: DNA-directed RNA polymerases I, II, and III subunit RPABC5

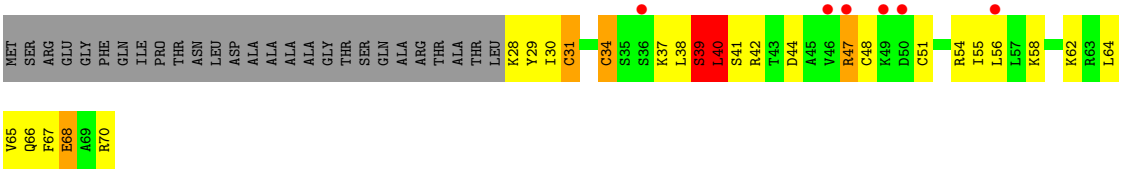
Chain J:

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



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





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	159.62Å 222.63Å 190.70Å 90.00° 97.69° 90.00°	Depositor
Resolution (Å)	47.96 – 3.78 47.96 – 3.78	Depositor EDS
% Data completeness (in resolution range)	97.6 (47.96-3.78) 97.6 (47.96-3.78)	Depositor EDS
R_{merge}	0.57	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.18 (at 3.77Å)	Xtriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
R, R_{free}	0.261 , 0.310 0.264 , 0.315	Depositor DCC
R_{free} test set	1996 reflections (3.10%)	wwPDB-VP
Wilson B-factor (Å^2)	112.7	Xtriage
Anisotropy	0.584	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 109.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.39$, $< L^2 > = 0.22$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	28985	wwPDB-VP
Average B, all atoms (Å^2)	148.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.62% 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: ATP, ZN, MG, WVQ

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	0.41	0/218	1.10	0/339
2	T	0.70	0/507	1.09	2/775 (0.3%)
3	N	0.75	0/311	0.94	1/479 (0.2%)
4	A	0.31	0/11020	0.58	1/14907 (0.0%)
5	B	0.31	0/9030	0.57	0/12186
6	C	0.31	0/2139	0.56	0/2899
7	E	0.33	0/1759	0.61	1/2367 (0.0%)
8	F	0.29	0/696	0.56	0/943
9	H	0.31	0/1082	0.64	0/1466
10	I	0.35	0/970	0.65	0/1308
11	J	0.32	0/541	0.59	0/727
12	K	0.33	0/937	0.62	0/1265
13	L	0.82	2/333 (0.6%)	1.19	4/442 (0.9%)
All	All	0.34	2/29543 (0.0%)	0.62	9/40103 (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
13	L	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
13	L	40	LEU	CG-CD2	-7.21	1.25	1.51
13	L	39	SER	C-O	-5.50	1.12	1.23

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
13	L	40	LEU	CB-CG-CD1	13.94	134.70	111.00
13	L	40	LEU	CB-CG-CD2	-7.81	97.72	111.00
3	N	2	DC	O4'-C4'-C3'	-7.47	101.51	104.50
13	L	39	SER	CB-CA-C	6.52	122.48	110.10
7	E	48	ASP	CB-CG-OD1	6.43	124.08	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
13	L	39	SER	Peptide
13	L	40	LEU	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	194	0	98	7	0
2	T	481	0	262	27	0
3	N	275	0	144	20	0
4	A	10828	0	10875	594	0
5	B	8859	0	8816	462	0
6	C	2101	0	2056	112	0
7	E	1724	0	1751	92	0
8	F	684	0	692	36	0
9	H	1064	0	1029	68	0
10	I	952	0	897	68	0
11	J	532	0	542	31	0
12	K	919	0	929	57	0
13	L	332	0	347	40	1
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	12	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	28985	0	28450	1441	1

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 25.

The worst 5 of 1441 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)
4:A:70:CYS:SG	4:A:80:HIS:CE1	2.58	0.96
4:A:117:GLU:HG2	4:A:122:MET:HE1	1.49	0.94
4:A:613:ILE:HG21	9:H:102:TYR:HB3	1.49	0.93
4:A:881:GLN:HB2	4:A:956:LEU:HD12	1.52	0.92
5:B:892:LYS:NZ	5:B:904:ARG:O	2.07	0.88

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
13:L:39:SER:O	13:L:39:SER:O[2_555]	1.95	0.25

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	1370/1733 (79%)	1337 (98%)	33 (2%)	0	100	100
5	B	1103/1224 (90%)	1085 (98%)	18 (2%)	0	100	100
6	C	265/318 (83%)	261 (98%)	4 (2%)	0	100	100
7	E	210/215 (98%)	207 (99%)	3 (1%)	0	100	100
8	F	84/155 (54%)	81 (96%)	3 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
9	H	129/146 (88%)	128 (99%)	1 (1%)	0	100	100
10	I	116/122 (95%)	112 (97%)	4 (3%)	0	100	100
11	J	63/70 (90%)	63 (100%)	0	0	100	100
12	K	112/120 (93%)	110 (98%)	2 (2%)	0	100	100
13	L	41/70 (59%)	37 (90%)	4 (10%)	0	100	100
All	All	3493/4173 (84%)	3421 (98%)	72 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	1194/1520 (79%)	1155 (97%)	39 (3%)	38	64
5	B	955/1061 (90%)	904 (95%)	51 (5%)	22	54
6	C	235/274 (86%)	226 (96%)	9 (4%)	33	61
7	E	192/197 (98%)	182 (95%)	10 (5%)	23	55
8	F	73/137 (53%)	72 (99%)	1 (1%)	67	82
9	H	116/128 (91%)	109 (94%)	7 (6%)	19	51
10	I	110/116 (95%)	101 (92%)	9 (8%)	11	40
11	J	60/65 (92%)	55 (92%)	5 (8%)	11	40
12	K	99/102 (97%)	93 (94%)	6 (6%)	18	50
13	L	36/57 (63%)	31 (86%)	5 (14%)	3	21
All	All	3070/3657 (84%)	2928 (95%)	142 (5%)	27	57

5 of 142 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
9	H	117	SER
10	I	22	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
11	J	43	ARG
5	B	308	TRP
5	B	218	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
5	B	1025	HIS
5	B	1176	ASN
12	K	2	ASN
6	C	135	GLN
9	H	134	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	R	7/9 (77%)	3 (42%)	0

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	R	2	U
1	R	5	A
1	R	8	G

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	WVQ	T	19	2	19,24,25	3.47	6 (31%)	20,33,36	1.60	5 (25%)

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	WVQ	T	19	2	-	4/6/40/41	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	19	WVQ	C5-N7	12.56	1.46	1.28
2	T	19	WVQ	C4-N9	4.65	1.45	1.35
2	T	19	WVQ	C2-N2	4.05	1.45	1.34
2	T	19	WVQ	O6-C6	-3.06	1.18	1.23
2	T	19	WVQ	C6-N1	-2.99	1.32	1.38

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	T	19	WVQ	N3-C2-N1	-4.34	119.39	126.43
2	T	19	WVQ	C2'-C1'-N9	-2.61	108.91	113.56
2	T	19	WVQ	N2-C2-N3	2.30	120.30	116.57
2	T	19	WVQ	N9-C4-N3	-2.08	116.60	119.70
2	T	19	WVQ	N2-C2-N1	2.07	120.31	117.06

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	T	19	WVQ	O4'-C4'-C5'-O5'
2	T	19	WVQ	C3'-C4'-C5'-O5'
2	T	19	WVQ	O4'-C1'-N9-C4
2	T	19	WVQ	C4'-C5'-O5'-P

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	T	19	WVQ	1	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	ATP	B	1302	-	26,33,33	0.60	0	31,52,52	0.77	1 (3%)

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	ATP	B	1302	-	-	6/18/38/38	0/3/3/3

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
16	B	1302	ATP	C5-C6-N6	2.30	123.85	120.35

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
16	B	1302	ATP	C5'-O5'-PA-O1A

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
16	B	1302	ATP	C5'-O5'-PA-O2A
16	B	1302	ATP	C4'-C5'-O5'-PA
16	B	1302	ATP	O4'-C4'-C5'-O5'
16	B	1302	ATP	C3'-C4'-C5'-O5'

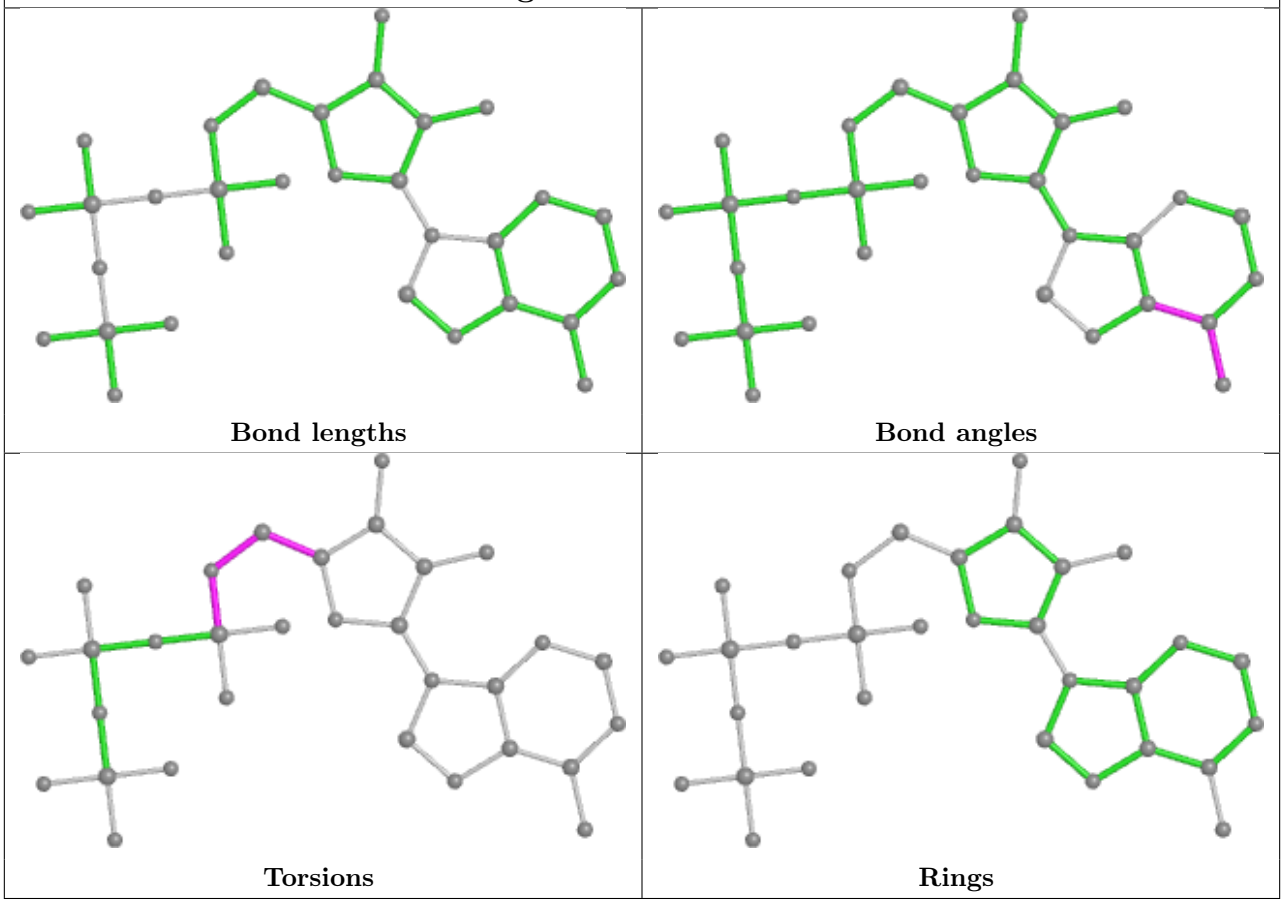
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
16	B	1302	ATP	1	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 ATP B 1302



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, 95^{th} 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%)	-0.03	0 100 100	111, 134, 207, 245	0
2	T	23/29 (79%)	-0.23	0 100 100	113, 216, 300, 320	0
3	N	13/18 (72%)	-0.25	0 100 100	215, 252, 297, 312	0
4	A	1384/1733 (79%)	0.00	46 (3%) 46 41	81, 142, 221, 332	0
5	B	1123/1224 (91%)	-0.05	23 (2%) 65 61	61, 131, 197, 288	0
6	C	267/318 (83%)	-0.05	7 (2%) 56 50	58, 119, 186, 246	0
7	E	212/215 (98%)	-0.01	14 (6%) 18 15	126, 174, 259, 319	0
8	F	86/155 (55%)	-0.21	2 (2%) 60 56	90, 134, 191, 258	0
9	H	133/146 (91%)	0.42	9 (6%) 17 14	96, 162, 229, 346	0
10	I	118/122 (96%)	0.06	6 (5%) 28 26	114, 171, 225, 258	0
11	J	65/70 (92%)	-0.12	0 100 100	68, 118, 167, 225	0
12	K	114/120 (95%)	-0.19	1 (0%) 84 82	72, 122, 173, 199	0
13	L	43/70 (61%)	0.56	6 (13%) 2 3	112, 255, 330, 400	0
All	All	3590/4229 (84%)	-0.01	114 (3%) 47 41	58, 139, 227, 400	0

The worst 5 of 114 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
7	E	112	TYR	6.4
7	E	93	MET	6.1
4	A	1192	LEU	5.9
4	A	1197	LEU	4.9
5	B	869	SER	4.8

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^{th} 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	WVQ	T	19	23/24	0.81	0.29	171,184,201,215	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

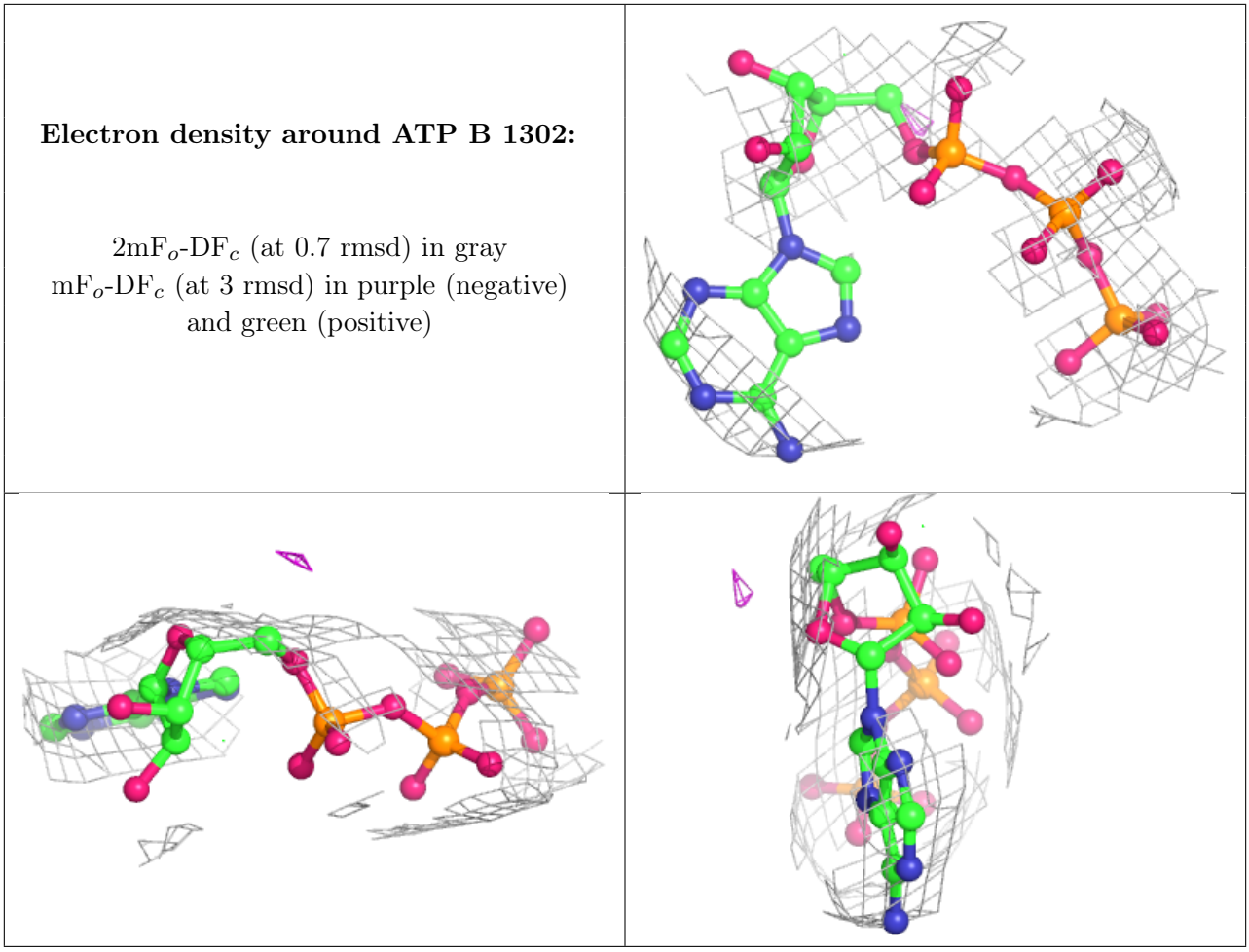
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	0.59	0.20	342,342,342,342	0
14	ZN	B	1301	1/1	0.66	0.08	216,216,216,216	0
16	ATP	B	1302	31/31	0.84	0.31	140,168,215,232	0
15	MG	A	1803	1/1	0.88	0.21	107,107,107,107	0
14	ZN	L	101	1/1	0.88	0.35	369,369,369,369	0
14	ZN	I	201	1/1	0.89	0.10	216,216,216,216	0
14	ZN	A	1802	1/1	0.91	0.07	208,208,208,208	0
14	ZN	I	202	1/1	0.93	0.10	287,287,287,287	0
14	ZN	C	401	1/1	0.93	0.13	158,158,158,158	0
14	ZN	J	101	1/1	0.95	0.22	115,115,115,115	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 ATP B 1302:

$2mF_o$-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

There are no such residues in this entry.