



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 22, 2020 – 12:41 PM BST

PDB ID : 2YU9  
Title : RNA polymerase II elongation complex in 150 mM MG+2 with UTP  
Authors : Wang, D.; Bushnell, D.A.; Westover, K.D.; Kaplan, C.D.; Kornberg, R.D.  
Deposited on : 2007-04-06  
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 : 2.13.1  
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) : 2.13.1

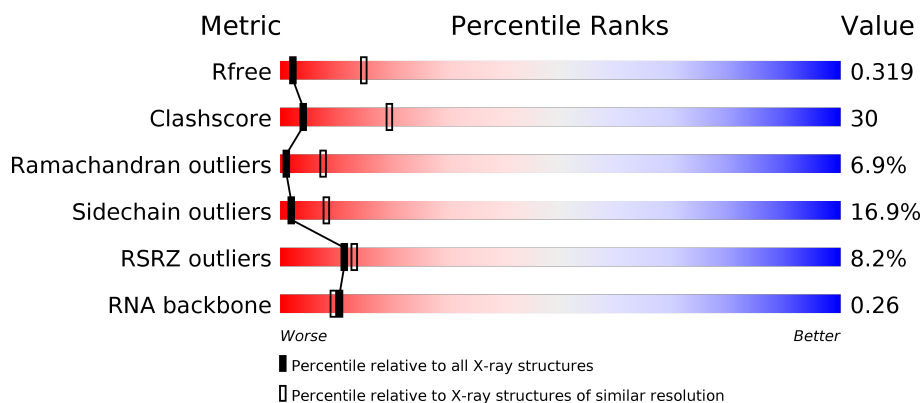
# 1 Overall quality at a glance

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	R	10	<div> <div>70%</div> <div>30%</div> </div>
2	T	28	<div> <div>43%</div> <div>29%</div> <div>43%</div> <div>29%</div> </div>
3	N	14	<div> <div>64%</div> <div>43%</div> <div>43%</div> <div>14%</div> </div>
4	A	1733	<div> <div>7%</div> <div>41%</div> <div>30%</div> <div>9%</div> <div>19%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	B	1224	
6	C	318	
7	E	215	
8	F	155	
9	H	146	
10	I	122	
11	J	70	
12	K	120	
13	L	70	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
16	UTP	A	2003[A]	-	-	X	-

## 2 Entry composition [i](#)

There are 16 unique types of molecules in this entry. The entry contains 29530 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 5'-R(\*AP\*UP\*CP\*GP\*AP\*GP\*AP\*GP\*GP\*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	R	10	Total	C	N	O	P	0	0	0
			216	98	45	64	9			

- Molecule 2 is a DNA chain called 28-MER DNA template strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	28	Total	C	N	O	P	0	0	0
			566	271	104	164	27			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	N	14	Total	C	N	O	P	0	0	0
			284	137	49	85	13			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	1411	Total	C	N	O	S	0	0	0
			11094	6994	1945	2094	61			

- Molecule 5 is a protein called DNA-directed RNA polymerase II 140 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	B	1121	Total	C	N	O	S	0	0	0
			8914	5643	1563	1653	55			

- Molecule 6 is a protein called DNA-directed RNA polymerase II 45 kDa polypeptide.

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

- Molecule 7 is a protein called DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide.

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

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	F	85	Total	C	N	O	S	0	0	0
			688	439	116	130	3			

- Molecule 9 is a protein called DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide.

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

- Molecule 10 is a protein called DNA-directed RNA polymerase II subunit 9.

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

- Molecule 11 is a protein called DNA-directed RNA polymerases I/II/III subunit 10.

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 13.6 kDa polypeptide.

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 7.7 kDa polypep-

tide.

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

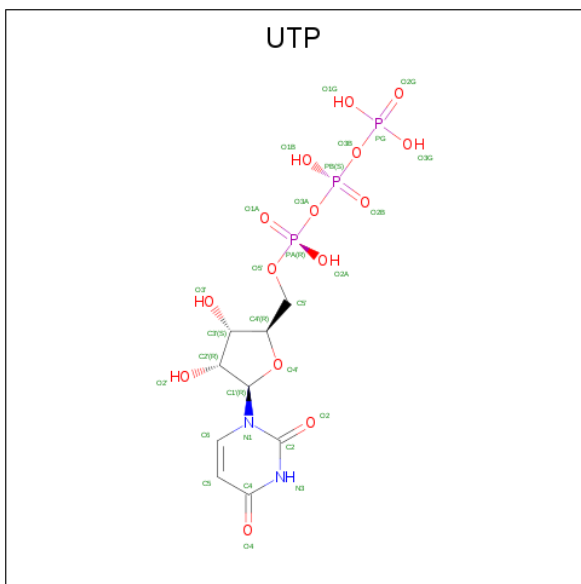
- Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
15	A	2	Total	Mg	0	0
			2	2		

- Molecule 16 is URIDINE 5'-TRIPHOSPHATE (three-letter code: UTP) (formula: C<sub>9</sub>H<sub>15</sub>N<sub>2</sub>O<sub>15</sub>P<sub>3</sub>).

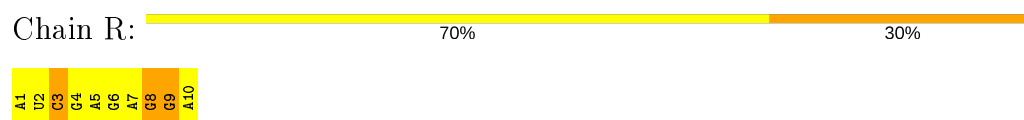


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
16	A	1	Total	C	N	O	P	0	1
			58	18	4	30	6		

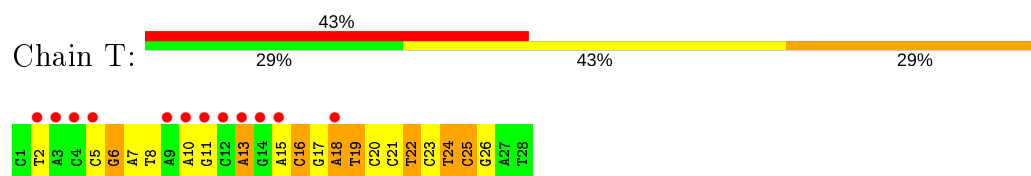
### 3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

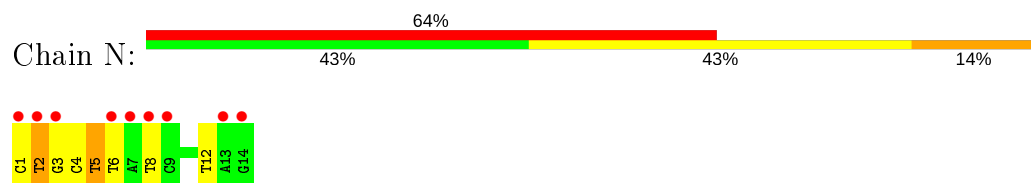
- Molecule 1: 5'-R(\*AP\*UP\*CP\*GP\*AP\*GP\*AP\*GP\*GP\*A)-3'



- Molecule 2: 28-MER DNA template strand



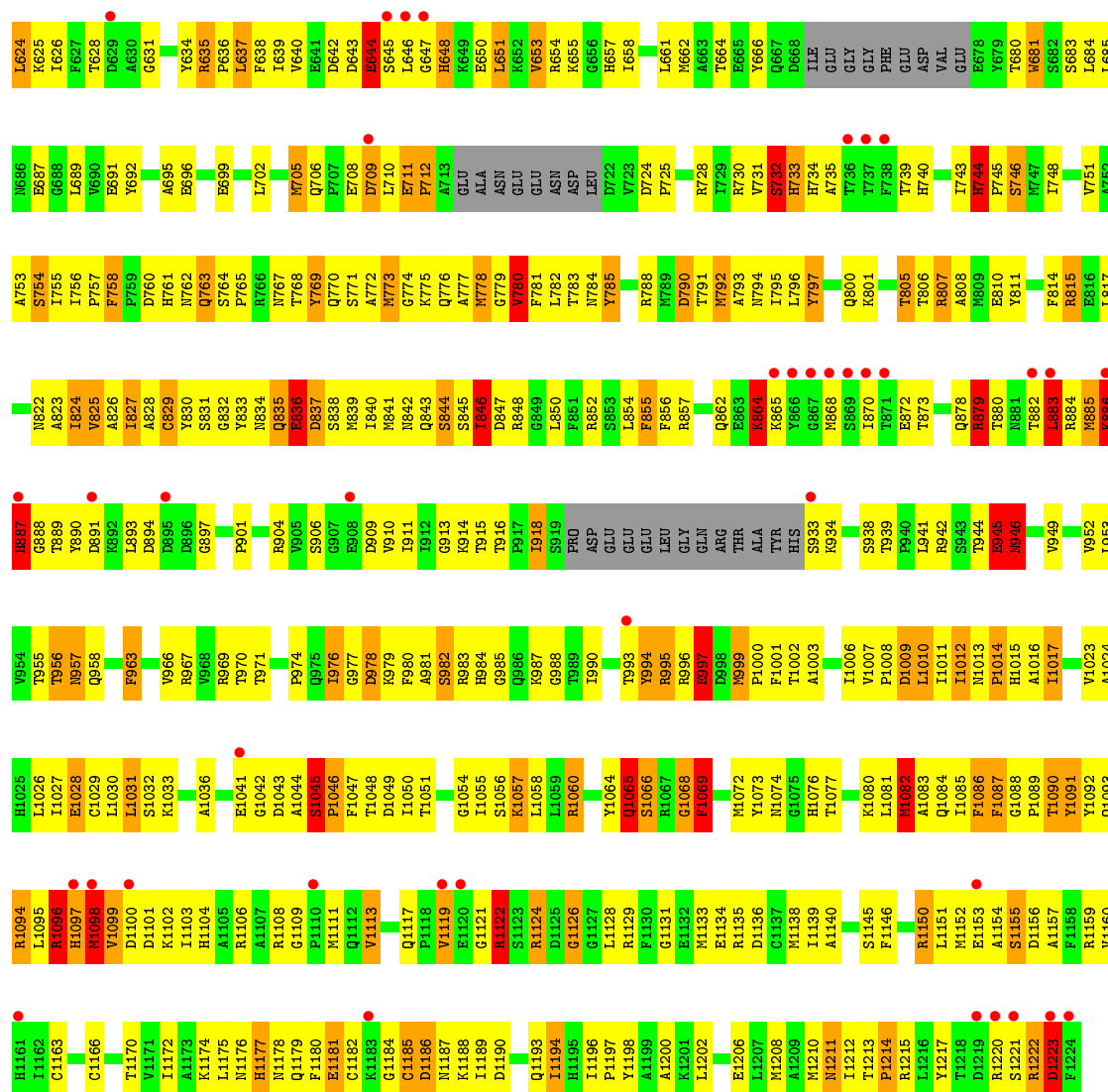
- Molecule 3: 5'-D(\*CP\*TP\*GP\*CP\*TP\*TP\*AP\*TP\*CP\*GP\*GP\*TP\*AP\*G)-3'



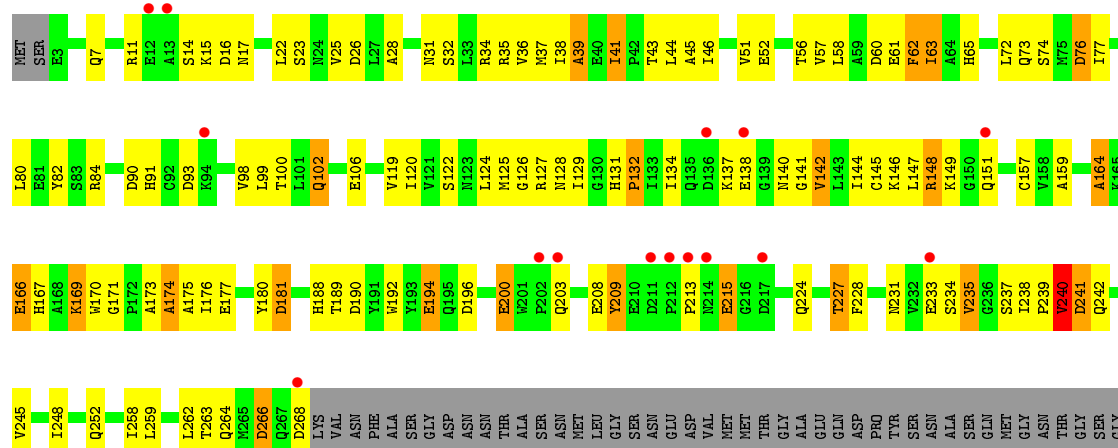


S1331	F1332	I1333	M1336	G1340	I1341	E1342	A1343	G1344	R1345	L1348	Y1349	K1350	Y1353	M1354	V1355	I1356	D1359	G1360	S1361	Y1362	V1363	M1364	D1365	R1366	H1367	M1368	A1369	L1370	V1372	T1376	T1377	Q1378	L1381	T1382	V1383	T1385	R1386	F1389	N1390	R1391	S1392	N1393	T1394	M1398	R1399	C1400	T1401	F1402						
ASP	ALA	GLU	THR	GLU	A1254	D1257	H1258	F1259	L1261	E1264	L1268	E1269	N1270	I1271	T1272	L1273	L1274	E1280	R1281	K1286	I1287	M1288	D1289	R1290	K1291	E1297	Y1298	E1301	P1302	E1303	W1304	V1305	E1307	T1308	D1309	M1312	L1313	S1314	E1315	V1319	G1321	I1322	D1323	T1324	V1243	R1244	PRO	LYS	SER	LEU				
F1084	H1085	F1086	A1087	G1088	V1089	A1090	S1091	K1092	K1093	V1094	T1095	S1096	G1097	V1098	P1099	R1100	L1105	H1106	V1107	M1110	M1111	K1112	T1113	S1115	P1116	L1117	T1118	Y1119	L1120	D1127	Q1128	E1129	Q1130	A1131	I1134	H1140	T1141	K1144	S1145	V1146	T1147	I1148	Y1153	Y1154	P1075	A1076	T1077	T1080	T1081	M1082	T1083			
P1164	E1165	E1168	Q1171	L1172	H1173	F1174	S1175	L1176	LEU	ASP	GLU	GLU	ALA	GLN	SER	PHE	ASP	Q1187	Q1188	L1192	L1193	L1197	D1198	R1199	A1200	A1201	D1204	K1205	D1206	M1209	T1219	K1220	K1221	N1222	D1223	L1224	F1225	D1231	N1232	D1233	E1234	G1240	V1243	R1244	PRO	LYS	SER	LEU						
R1012	D1013	A1014	C1020	L1021	S1024	R1025	L1026	A1027	T1028	R1029	R1030	V1031	L1032	Y1035	H1036	L1037	T1038	K1039	Q1040	A1041	F1042	W1044	V1045	R1046	S1047	N1048	H1049	E1050	R1055	S1056	V1057	V1058	H1059	P1060	M1063	V1064	L1067	A1068	A1069	D1069	P1070	S1071	L1072	G1073	E1074	S1075	A1076	T1077	T1080	T1081	M1082	T1083		
K924	L925	Q926	L929	D930	E931	K938	K941	L943	R944	E945	V946	F947	G950	W954	P957	V958	N959	I960	R961	R962	I963	G964	Q965	N966	Q969	T970	F971	H972	H975	S979	D980	L981	T982	K991	Q994	E995	N996	V999	L1000	M1004	E1005	I1006	I1007	Q1011										
H851	T852	D853	T855	T856	R857	R858	S859	L860	G861	T862	V863	T864	T867	T868	D870	D871	T873	H874	A875	T878	E879	R880	Y881	S882	L883	D884	T886	R886	R895	R896	T907	L908	D909	P910	S911	L912	L913	E914	S915	Q916	T919	L920	G921	M848	L849	R923								
K773	R774	T775	A776	F777	G778	F779	V780	D781	R782	T783	L784	P785	H786	F787	S788	K789	D790	S796	K797	G798	F799	V800	E801	H802	Y804	G807	L808	T809	P810	Q811	E812	F813	F814	F815	H816	R821	L825	D826	T827	A828	Y829	Y836	R839	R840	L845	E846	D847	T848	M849	V850				
L679	L691	D692	K695	E696	A697	I608	D609	G610	I612	V617	E618	K619	R620	T621	V622	G623	S624	G628	L629	I630	H631	L632	Q633	F634	E635	L636	R637	F640	F646	I649	Q650	K651	V652	V653	N654	L657	L658	H659	N660	G661	F662	T664	G665	I666	G667	D668	T669	I670	A671	Q674	V675	M676	R677	E678
P514	N517	K518	D519	C520	M521	Q525	L528	R532	K533	L534	T535	G536	R537	F540	I541	E542	L543	D544	Q545	V546	L547	N548	D557	P563	I565	I566	K567	P568	K569	P570	N572	S573	G574	K575	L578	S579	V580	H587	L588	Q589	A506	V507	P508	D592	E593	T595								
Q447	P448	L450	R451	K452	M453	N454	M455	D456	A457	H458	R459	V460	L463	T464	Y465	T467	F469	I470	E542	L470	H471	L472	S473	V474	Y475	Y478	N479	A480	D481	F482	D483	G484	D485	E486	N487	N488	L489	H490	V491	P492	T497	R498	A499	E500	L501	S502	Q503	A506	V507	P508	D592	E593	T595	
Q313	A314	L315	Q316	R317	S318	G319	P321	K322	V323	L324	R325	R326	A327	L329	K330	G331	K332	E333	G334	R335	I336	R337	G338	M341	G342	K343	R344	V345	D346	A349	R350	T351	V352	I353	D356	P357	E360	L361	D362	Q363	V364	G365	V366	P367	K368	S369	I370	A371	K372	T373	L374	T375	R446	
P377	T381	Y383	T386	R387	Q390	L391	V392	R393	K394	G395	P396	N397	E398	H399	F400	G401	A402	L403	I336	Y404	L406	S409	G410	D411	R412	L415	R416	Y417	S418	D423	I424	Q425	Y428	G429	V430	K431	V432	E433	R434	H435	T436	D437	M438	D439	R440	Y441	P442	L443	F444	R446				



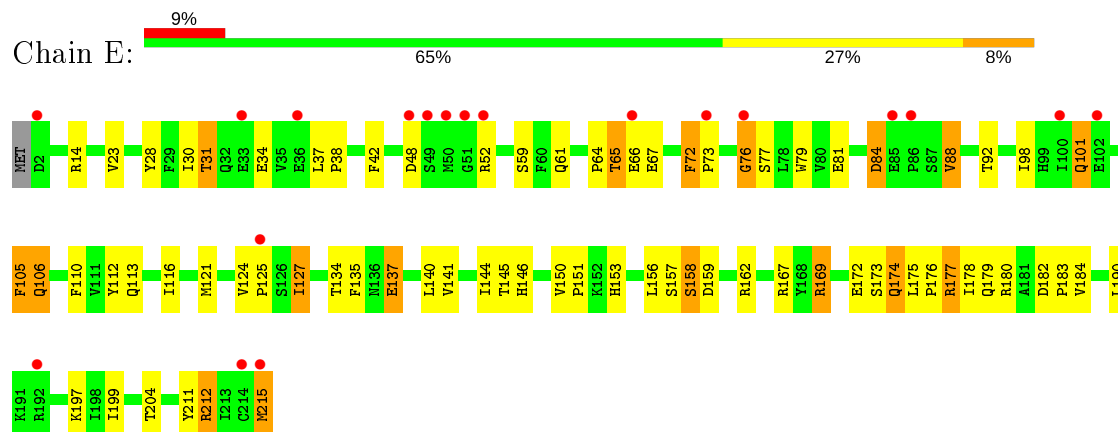


● Molecule 6: DNA-directed RNA polymerase II 45 kDa polypeptide

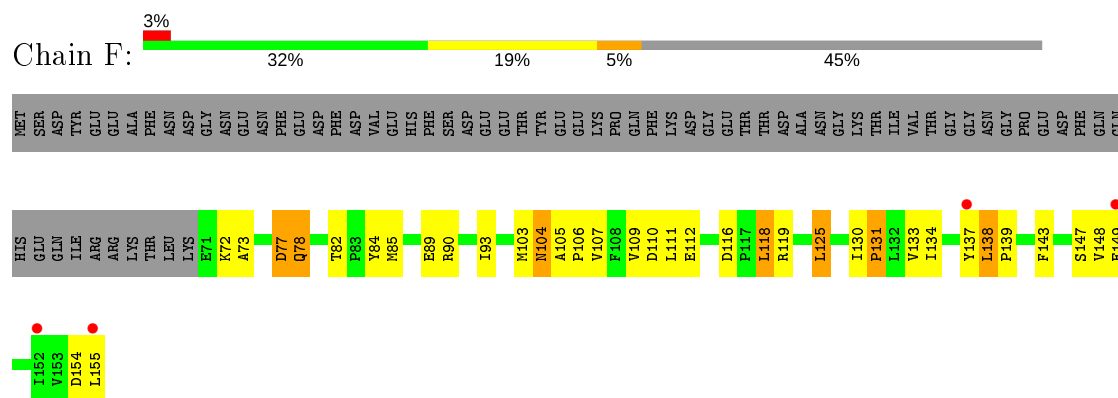


GLY  
TYR  
ASP  
ASN  
ALA  
TRP

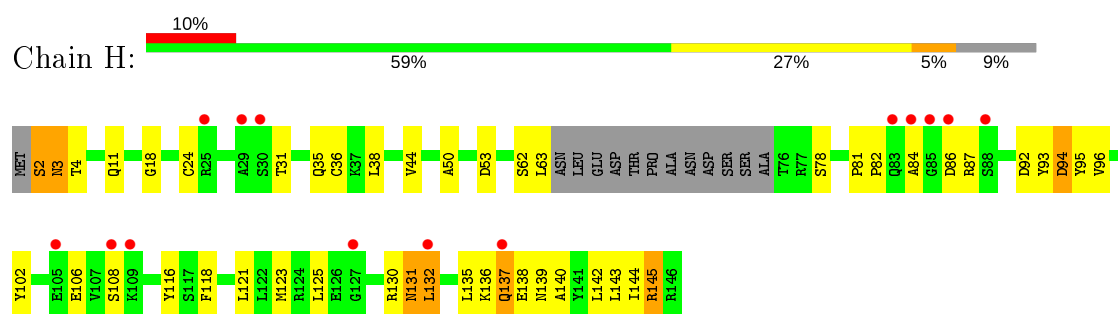
- Molecule 7: DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide



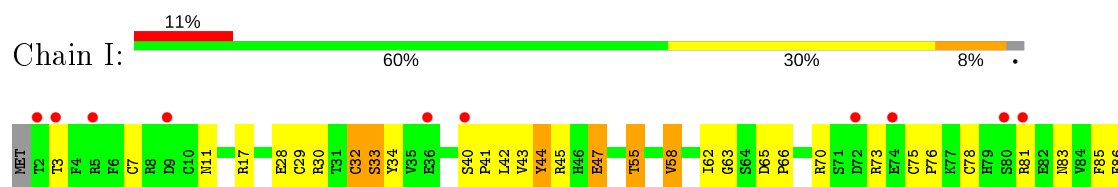
- Molecule 8: DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide

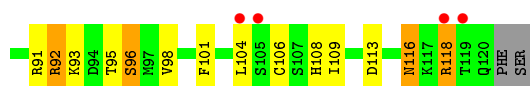


- Molecule 9: DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide

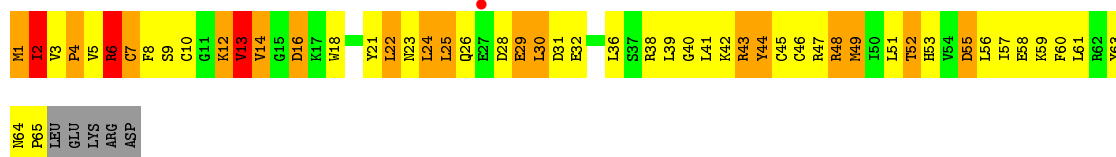
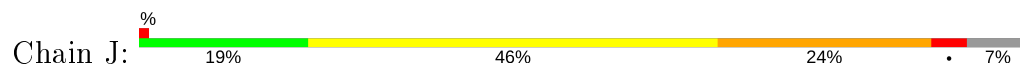


- Molecule 10: DNA-directed RNA polymerase II subunit 9

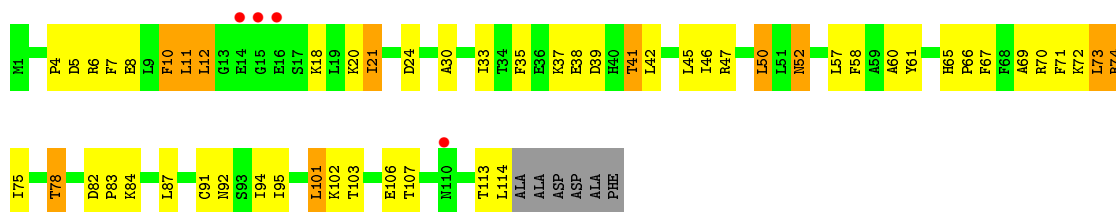




- Molecule 11: DNA-directed RNA polymerases I/II/III subunit 10



- Molecule 12: DNA-directed RNA polymerase II 13.6 kDa polypeptide



- Molecule 13: DNA-directed RNA polymerases I, II, and III 7.7 kDa polypeptide



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	170.61Å 222.73Å 196.16Å 90.00° 101.87° 90.00°	Depositor
Resolution (Å)	19.98 – 3.40 19.97 – 3.40	Depositor EDS
% Data completeness (in resolution range)	90.0 (19.98-3.40) 90.0 (19.97-3.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.77 (at 3.44Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.283 , 0.344 0.271 , 0.319	Depositor DCC
$R_{free}$ test set	2609 reflections (2.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	101.5	Xtriage
Anisotropy	0.350	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.20 , 121.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.39$ , $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	29530	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	146.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.80% 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: UTP, MG, ZN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	R	0.58	0/243	1.05	1/378 (0.3%)
2	T	0.91	1/634 (0.2%)	1.72	18/975 (1.8%)
3	N	1.12	1/317 (0.3%)	1.59	6/488 (1.2%)
4	A	0.49	0/11292	0.66	0/15267
5	B	0.54	0/9087	0.68	0/12253
6	C	0.56	0/2133	0.68	0/2891
7	E	0.46	0/1788	0.61	0/2406
8	F	0.46	0/700	0.65	0/945
9	H	0.44	0/1086	0.66	0/1470
10	I	0.46	0/989	0.65	0/1331
11	J	0.50	0/541	0.65	0/727
12	K	0.53	0/937	0.64	0/1265
13	L	0.59	0/365	0.83	0/485
All	All	0.53	2/30112 (0.0%)	0.73	25/40881 (0.1%)

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	18
5	B	0	20
6	C	0	1
9	H	0	1
11	J	0	4
All	All	0	44

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	N	12	DT	C5-C7	10.67	1.56	1.50
2	T	18	DA	O3'-P	7.41	1.70	1.61

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	T	25	DC	O4'-C1'-N1	13.75	117.62	108.00
3	N	1	DC	O4'-C1'-N1	10.87	115.61	108.00
2	T	16	DC	O4'-C1'-N1	10.29	115.21	108.00
2	T	25	DC	C1'-O4'-C4'	-10.21	99.89	110.10
2	T	25	DC	O4'-C4'-C3'	-9.01	100.60	106.00

There are no chirality outliers.

5 of 44 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	A	117	GLU	Peptide
4	A	218	ASP	Peptide
4	A	222	LEU	Peptide
4	A	53	LEU	Peptide
4	A	54	ASN	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	R	216	0	109	59	0
2	T	566	0	316	44	0
3	N	284	0	161	3	0
4	A	11094	0	11178	663	0
5	B	8914	0	8953	727	0
6	C	2095	0	2051	120	0
7	E	1752	0	1776	52	0
8	F	688	0	707	23	0
9	H	1068	0	1040	38	0
10	I	971	0	927	30	0
11	J	532	0	542	121	0
12	K	919	0	929	41	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
13	L	363	0	386	10	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	2	0	0	0	0
16	A	58	0	22	12	0
All	All	29530	0	29097	1746	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 30.

The worst 5 of 1746 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:A:2003[A]:UTP:H5'2	16:A:2003[A]:UTP:C6	1.44	1.52
9:H:2:SER:CB	9:H:3:ASN:HB2	1.35	1.50
4:A:1444:MET:CG	4:A:1445:ILE:HG13	1.49	1.42
9:H:2:SER:HB2	9:H:3:ASN:CB	1.52	1.40
4:A:1040:GLN:N	4:A:1041:ALA:HB3	1.43	1.31

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	1401/1733 (81%)	1052 (75%)	250 (18%)	99 (7%)	<b>1</b> <b>7</b>
5	B	1105/1224 (90%)	836 (76%)	177 (16%)	92 (8%)	<b>1</b> <b>5</b>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
6	C	264/318 (83%)	197 (75%)	54 (20%)	13 (5%)	2	14
7	E	212/215 (99%)	184 (87%)	22 (10%)	6 (3%)	5	24
8	F	83/155 (54%)	70 (84%)	7 (8%)	6 (7%)	1	7
9	H	129/146 (88%)	99 (77%)	22 (17%)	8 (6%)	1	10
10	I	117/122 (96%)	88 (75%)	22 (19%)	7 (6%)	1	10
11	J	63/70 (90%)	49 (78%)	7 (11%)	7 (11%)	0	3
12	K	112/120 (93%)	96 (86%)	14 (12%)	2 (2%)	8	32
13	L	44/70 (63%)	26 (59%)	14 (32%)	4 (9%)	1	4
All	All	3530/4173 (85%)	2697 (76%)	589 (17%)	244 (7%)	1	8

5 of 244 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	55	ASP
4	A	68	GLN
4	A	93	VAL
4	A	117	GLU
4	A	118	HIS

### 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	
4	A	1231/1520 (81%)	1016 (82%)	215 (18%)	2	7
5	B	972/1061 (92%)	799 (82%)	173 (18%)	2	6
6	C	234/274 (85%)	204 (87%)	30 (13%)	4	16
7	E	196/197 (100%)	173 (88%)	23 (12%)	5	20
8	F	75/137 (55%)	68 (91%)	7 (9%)	9	31
9	H	117/128 (91%)	102 (87%)	15 (13%)	4	16
10	I	113/116 (97%)	93 (82%)	20 (18%)	2	6
11	J	60/65 (92%)	40 (67%)	20 (33%)	0	1

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
12	K	99/102 (97%)	80 (81%)	19 (19%)	1	4
13	L	40/57 (70%)	32 (80%)	8 (20%)	1	3
All	All	3137/3657 (86%)	2607 (83%)	530 (17%)	2	8

5 of 530 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	B	245	GLU
5	B	662	MET
11	J	16	ASP
5	B	344	LYS
5	B	487	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 79 such sidechains are listed below:

Mol	Chain	Res	Type
5	B	178	ASN
5	B	657	HIS
11	J	53	HIS
5	B	278	GLN
5	B	499	ASN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	R	8/10 (80%)	2 (25%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	R	8	G
1	R	9	G

There are no RNA pucker outliers to report.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 5.5 Carbohydrates

There are no monosaccharides in this entry.

## 5.6 Ligand geometry

Of 12 ligands modelled in this entry, 10 are monoatomic - leaving 2 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	UTP	A	2003[B]	-	26,30,30	1.95	5 (19%)	34,47,47	2.03	9 (26%)
16	UTP	A	2003[A]	-	26,30,30	1.95	5 (19%)	34,47,47	2.03	9 (26%)

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	UTP	A	2003[B]	-	-	1/22/38/38	0/2/2/2
16	UTP	A	2003[A]	-	-	6/22/38/38	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
16	A	2003[B]	UTP	C6-C5	-4.88	1.39	1.52
16	A	2003[A]	UTP	C6-C5	-4.85	1.39	1.52
16	A	2003[B]	UTP	C5-C4	-4.56	1.39	1.50
16	A	2003[A]	UTP	C5-C4	-4.55	1.39	1.50
16	A	2003[B]	UTP	C6-N1	-3.68	1.40	1.47

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
16	A	2003[B]	UTP	C4-N3-C2	-7.13	119.87	125.79

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
16	A	2003[A]	UTP	C4-N3-C2	-7.11	119.89	125.79
16	A	2003[A]	UTP	PB-O3B-PG	-3.75	119.95	132.83
16	A	2003[B]	UTP	PB-O3B-PG	-3.73	120.01	132.83
16	A	2003[A]	UTP	PB-O3A-PA	-3.73	120.03	132.83

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

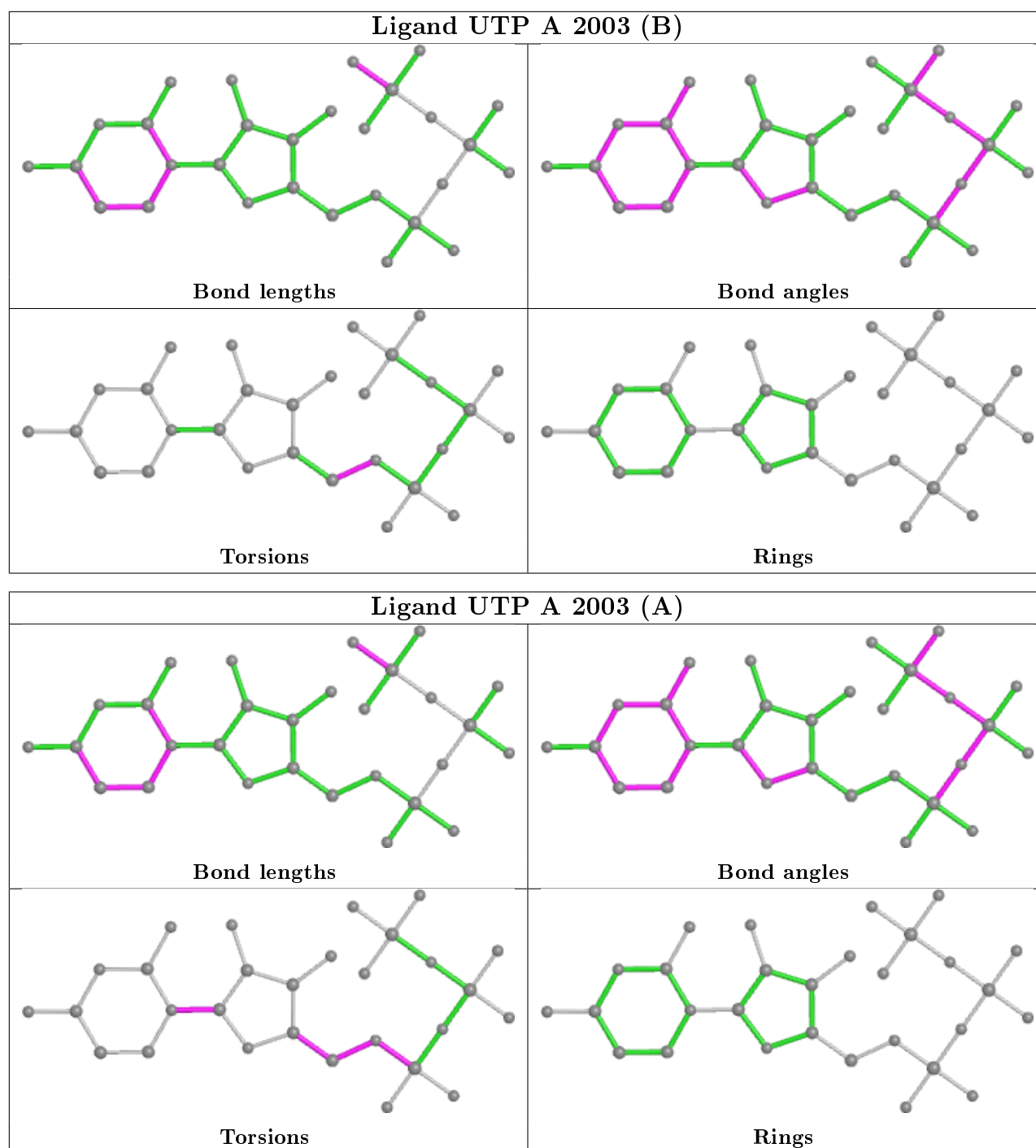
Mol	Chain	Res	Type	Atoms
16	A	2003[A]	UTP	C4'-C5'-O5'-PA
16	A	2003[A]	UTP	C3'-C4'-C5'-O5'
16	A	2003[A]	UTP	O4'-C4'-C5'-O5'
16	A	2003[B]	UTP	C4'-C5'-O5'-PA
16	A	2003[A]	UTP	C2'-C1'-N1-C6

There are no ring outliers.

2 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
16	A	2003[B]	UTP	3	0
16	A	2003[A]	UTP	9	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](#)

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<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	R	10/10 (100%)	0.66	0 100 100	118, 142, 150, 153	0
2	T	28/28 (100%)	1.72	12 (42%) 0 0	138, 167, 200, 200	0
3	N	14/14 (100%)	2.74	9 (64%) 0 0	155, 184, 200, 200	0
4	A	1411/1733 (81%)	0.40	114 (8%) 12 13	101, 149, 166, 193	0
5	B	1121/1224 (91%)	0.34	81 (7%) 15 17	84, 141, 159, 176	0
6	C	266/318 (83%)	0.27	15 (5%) 24 25	109, 139, 155, 178	0
7	E	214/215 (99%)	0.57	19 (8%) 9 11	147, 163, 172, 177	0
8	F	85/155 (54%)	0.30	4 (4%) 31 31	144, 160, 173, 178	0
9	H	133/146 (91%)	0.54	14 (10%) 6 7	141, 154, 173, 182	0
10	I	119/122 (97%)	0.75	14 (11%) 4 5	134, 153, 169, 171	0
11	J	65/70 (92%)	0.04	1 (1%) 73 72	99, 120, 137, 144	0
12	K	114/120 (95%)	0.26	4 (3%) 44 43	130, 147, 166, 171	0
13	L	46/70 (65%)	1.20	11 (23%) 0 0	148, 174, 178, 179	0
All	All	3626/4225 (85%)	0.42	298 (8%) 11 13	84, 147, 169, 200	0

The worst 5 of 298 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	B	250	PHE	10.2
5	B	1221	SER	10.0
5	B	339	THR	8.9
4	A	1090	ALA	7.7
5	B	1224	PHE	7.5

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

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

### 6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

### 6.4 Ligands ⓘ

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.

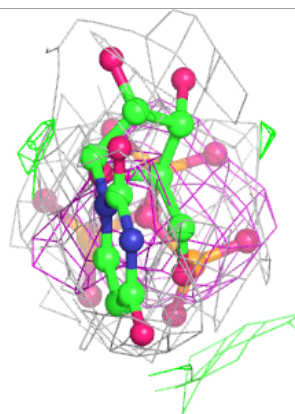
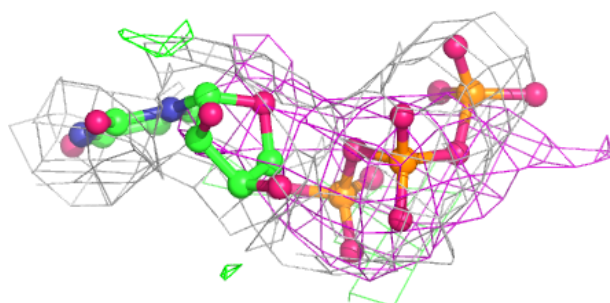
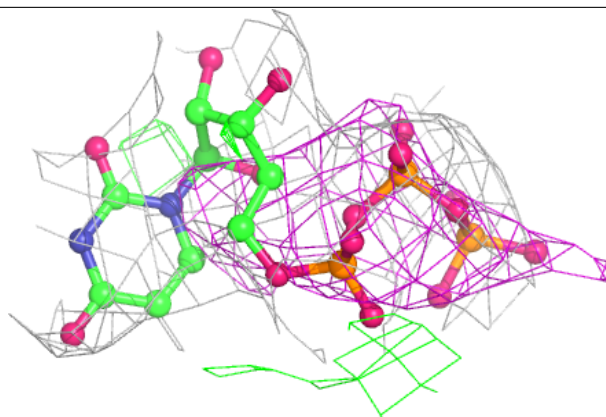
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
14	ZN	I	204	1/1	0.81	0.12	154,154,154,154	0
16	UTP	A	2003[A]	29/29	0.82	0.38	132,133,138,138	29
16	UTP	A	2003[B]	29/29	0.82	0.38	137,137,137,137	29
14	ZN	A	1734	1/1	0.87	0.25	157,157,157,157	0
14	ZN	I	203	1/1	0.90	0.10	141,141,141,141	0
15	MG	A	2001	1/1	0.91	0.15	124,124,124,124	0
14	ZN	A	1735	1/1	0.93	0.08	149,149,149,149	0
14	ZN	C	319	1/1	0.95	0.09	149,149,149,149	0
14	ZN	L	105	1/1	0.95	0.27	175,175,175,175	0
14	ZN	B	1307	1/1	0.96	0.19	142,142,142,142	0
15	MG	A	2002	1/1	0.96	0.33	59,59,59,59	0
14	ZN	J	101	1/1	0.98	0.07	109,109,109,109	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.

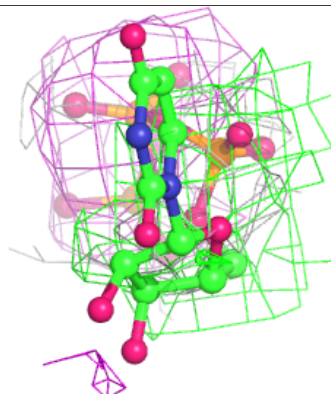
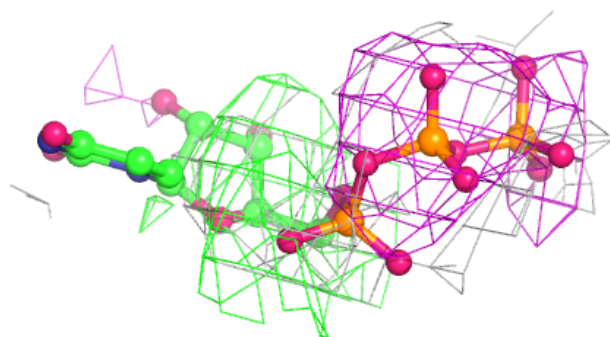
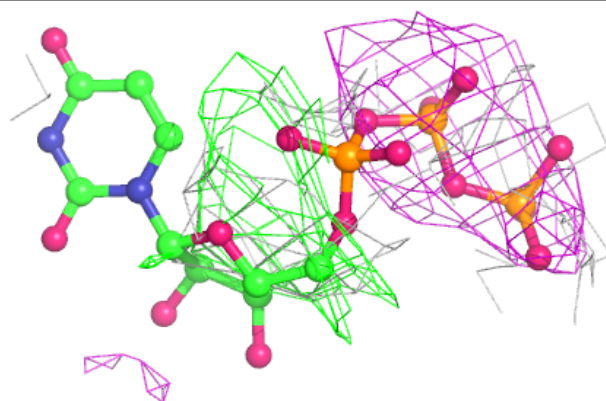


**Electron density around UTP A 2003 (A):**

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

**Electron density around UTP A 2003 (B):**

$2mF_o-DF_c$  (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.