



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 14, 2023 – 08:16 PM EDT

PDB ID : 1R9S  
Title : RNA POLYMERASE II STRAND SEPARATED ELONGATION COM-  
PLEX, MATCHED NUCLEOTIDE  
Authors : Westover, K.D.; Bushnell, D.A.; Kornberg, R.D.  
Deposited on : 2003-10-30  
Resolution : 4.25 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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)  
Xtrriage (Phenix) : 1.13  
EDS : 2.35  
buster-report : 1.1.7 (2018)  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.35

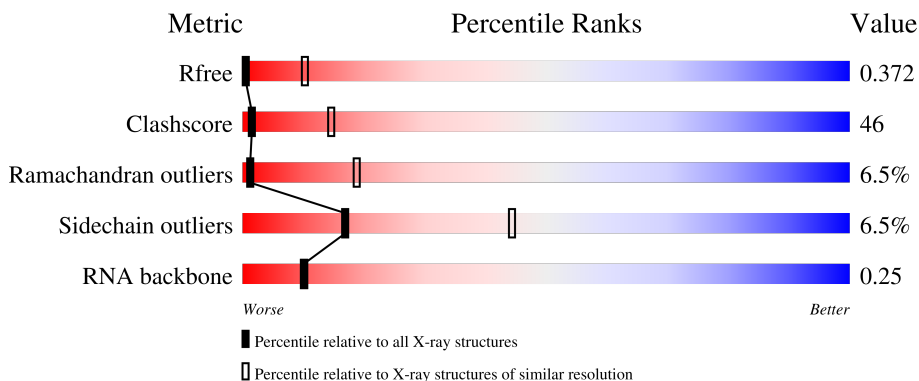
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.25 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.



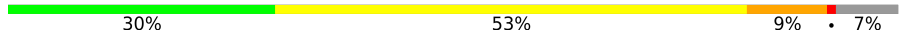
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1017 (4.72-3.78)
Clashscore	141614	1059 (4.72-3.80)
Ramachandran outliers	138981	1014 (4.72-3.80)
Sidechain outliers	138945	1018 (4.72-3.78)
RNA backbone	3102	1057 (5.50-3.00)

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  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	R	10	50% (red), 50% (orange)
2	T	14	21% (yellow), 64% (orange), 14% (red)
3	A	1733	31% (green), 41% (yellow), 7% (orange), 20% (grey)
4	B	1224	32% (green), 51% (yellow), 7% (orange), 10% (grey)
5	C	318	37% (green), 40% (yellow), 7% (orange), 16% (grey)
6	E	215	40% (green), 56% (yellow), 7% (orange), 1% (grey)

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	F	155	
8	H	146	
9	I	122	
10	J	70	
11	K	120	
12	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
14	UTP	R	3000	X	-	-	-
15	ZN	I	204	-	-	X	-

## 2 Entry composition [i](#)

There are 15 unique types of molecules in this entry. The entry contains 28491 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 strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	R	10	217	98	45	65	9	0	0	0

- Molecule 2 is a DNA chain called DNA strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	T	14	279	135	48	83	13	0	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II largest subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	1381	10857	6851	1899	2046	61	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerase II 140 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	B	1097	8720	5526	1523	1617	54	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerase II 45 kDa polypeptide.

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

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide.

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

- Molecule 7 is a protein called DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide.

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

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide.

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

- Molecule 9 is a protein called DNA-directed RNA polymerase II 14.2 kDa polypeptide.

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

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III 8.3 kDa polypeptide.

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

- Molecule 11 is a protein called DNA-directed RNA polymerase II 13.6 kDa polypeptide.

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

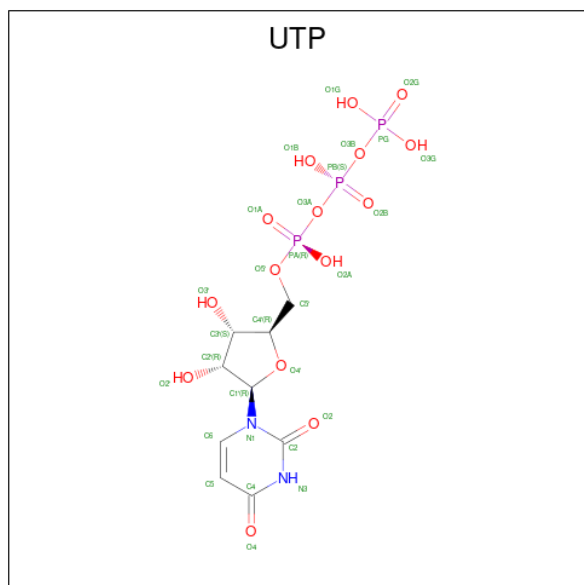
- Molecule 12 is a protein called DNA-directed RNA polymerases I, II, and III 7.7 kDa polypeptide.

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

- Molecule 13 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
13	R	1	Total Mg 1 1	0	0
13	A	1	Total Mg 1 1	0	0

- Molecule 14 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
14	R	1	Total C N O P 29 9 2 15 3	0	0

- Molecule 15 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

### 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. 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. 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.

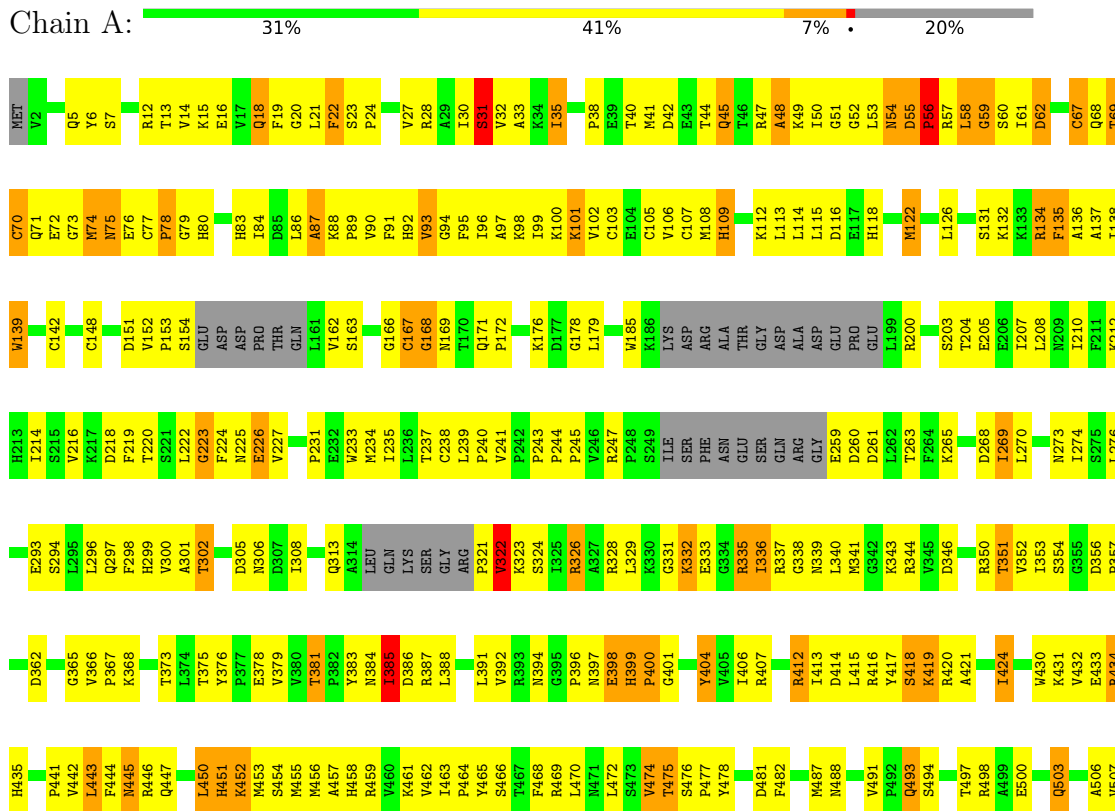
- Molecule 1: RNA strand



- Molecule 2: DNA strand



- Molecule 3: DNA-directed RNA polymerase II largest subunit

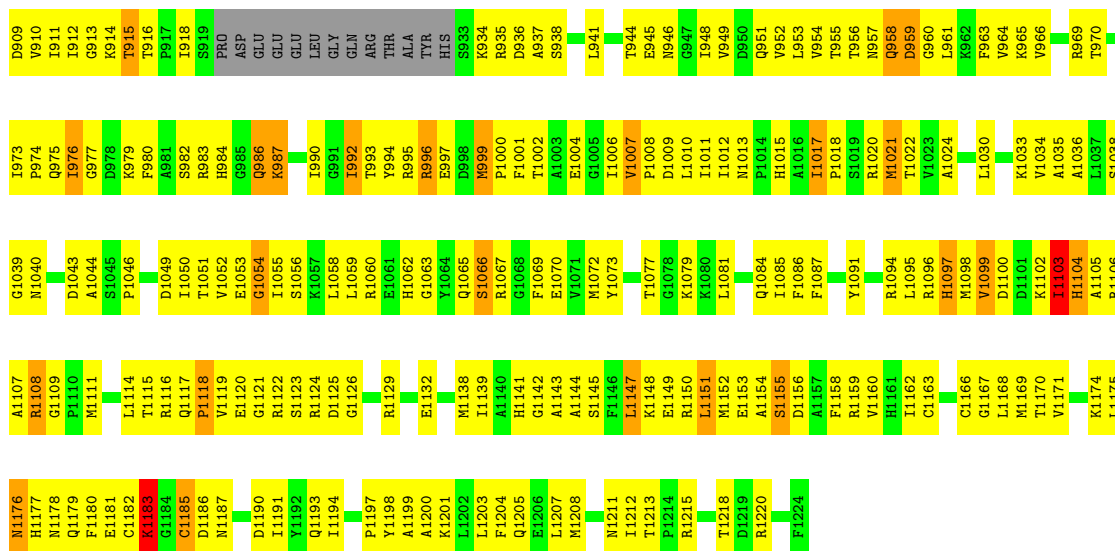


THR	P508	S773	K651	F814	R1030	V1088	E1168	I1237	G1310	T1376	G1439	GLY
SER	I511	G574	V652	F815	Y1035	P1089	I1169	I1238	V1311	T1377	A1440	SER
PRO	K575	K575	G654	H816	R1036	P1100	I1170	R1239	M1312	G1378	F1441	PRO
ASP	V512	K576	N654	A817	L1036	L1101	Q1171	C1240	L1037	A1379	D1442	ASP
TYR	S513	L577	F655	M818	L1037	E1102	L1172	V1241	E1315	L1380	V1443	TYR
ALA	P514	L578	M819	G819	T1038	K1103	H1173	V1242	V1316	M1444	M1444	ALA
PRO	Q515	S579	M656	C820	L1039	I1104	H1174	V1243	M1317	T1382	T1445	PRO
THR	S516	V580	L657	K820	K1039	I1105	E1176	ARG	L1318	S1383	ASP	THR
GLY	M517	V580	L658	R821	Q1040	L1106	LEU	PRO	V1319	V1384	GLU	GLY
SER	N584	N584	S663	G823	A1041	M1106	LEU	THR	P1320	P1385	GLU	SER
PRO	K518	G888	T664	G823	F1042	N1107	ASP	LYS	SER	R1386	THR	PRO
THR	P519	S888	T664	L824	D1043	V1107	GLU	LEU	GLY	H1387	VAL	THR
ALA	C520	H567	G665	H825	V1044	M1111	ALA	ASP	ALA	G1388	LEU	ALA
SER	M521	L568	I666	T826	V1045	K1112	GLY	LEU	GLY	F1389	THR	SER
PRO	G522	Q589	G667	T827	L1046	T1113	GLU	ALA	GLU	F1390	LYS	PRO
THR	I523	R590	I670	A828	S1047	P1114	GLN	GLU	GLN	N1390	TYR	THR
SER	V524	F591	I671	R829	M1048	K895	SER	THR	THR	R1391	ALA	SER
ASP	Q525	D592	S751	K830	L1116	L1116	ASP	GLU	GLU	S1392	PRO	ASP
TYR	L528	T595	D672	G833	E1050	T1117	PHE	A1254	A1254	M1393	TYR	TYR
SER	C529	L596	G673	R834	Q1052	V1118	ASP	E1255	T1329	T1394	GLN	TYR
PRO	G530	L598	T674	T834	F1053	E1121	G1188	H1256	M1330	G1395	LYS	SER
THR	I531	L598	T675	G835	L1054	P1122	S1189	M1258	M1331	A1396	ILE	PRO
SER	R532	L599	M676	G836	L1054	A1126	P1190	M1259	F1332	L1397	THR	THR
ASP	K533	S599	M676	L901	L1054	D1127	M1191	L1260	I1333	M1398	GLU	SER
PRO	K533	S599	E678	L901	L1054	Q1128	W1191	K1261	I1333	R1399	ILE	PRO
VAL	L534	N603	E679	R939	L1054	Q1129	L1193	K1262	D1334	C1400	GLY	SER
SER	T535	G694	I679	R840	V1057	A1126	L1193	I283	M1335	S1401	ASP	TYR
PRO	L536	M605	T680	R841	V1058	Q1128	L1194	E1284	M1336	F1402	ALA	SER
THR	L537	M606	I683	L841	H1058	Q1129	L1195	M1267	I1340	E1403	GLN	THR
GLY	D538	I607	C764	L842	P1060	E1129	L1196	ASP	I1341	A1404	GLY	GLY
PRO	T539	I608	A656	K843	E1062	Q1132	L1197	M1269	G1342	V1406	GLY	SER
ALA	F540	I608	A656	L844	M1063	K1132	L1197	K1269	A1343	E1407	VAL	PRO
SER	I612	I608	A656	L844	V1064	L1133	M1202	E1270	A1344	L1408	THR	PRO
TYR	L613	I613	K689	E846	G1065	I1134	K1205	I1271	G1344	L1409	PRO	TYR
PRO	L543	L543	V690	E847	V1066	R1135	D1206	T1272	R1345	L1410	THR	PRO
THR	Q545	D544	L691	I848	L1067	L998	L1207	L1273	A1346	F1410	THR	PRO
SER	M546	D615	I775	M849	A1068	L998	T1208	V1276	A1347	E1411	SER	THR
GLY	L547	V616	I775	V850	A1069	L1001	M1209	E1277	L1348	A1412	ASN	SER
VAL	M548	E618	G778	K852	G1073	R1001	G1210	E1278	G1350	G1413	GLY	PRO
TYR	M549	K619	F779	D853	E1074	M1004	Q1211	M1278	E1351	A1414	SER	VAL
SER	W552	G623	R782	L923	P1075	E1005	V1212	G1212	V1352	A1416	LEU	SER
THR	V553	G623	T783	K924	A1076	I1006	E1213	E1213	I1356	L1418	VAL	PRO
PRO	W556	G623	L784	L928	T1077	Q1008	I1216	V1282	I1356	L1419	ASN	THR
SER	D557	G628	L702	L929	L1081	A1014	A1148	V1284	M1284	D1420	ALA	PRO
THR	G558	L629	T703	L929	THR	V1015	A1149	M1284	M1284	C1421	ASP	SER
PRO	V559	V633	A704	D950	THR	T1016	S1150	G1289	G1360	G1421	LEU	PRO
SER	V559	V633	A704	E931	PHE	T1016	S1151	P1294	S1361	S1425	ASP	THR
PRO	T562	K637	T709	H864	HIS	L1017	I1152	T1295	Y1362	V1425	VAL	PRO
THR	P563	K637	L710	L864	PHE	F1018	Y1153	G1296	V1363	S1425	LYS	PRO
SER	A564	P639	R711	Q885	ALA	C1019	Y1154	E1297	M1222	V1428	THR	THR
PRO	I566	Q640	E712	F866	GLY	C1020	D1185	E1299	D1223	I1428	GLY	SER
THR	I566	Q640	E712	I867	VAL	L1021	D1185	Y1299	L1224	I1429	LEU	PRO
SER	K644	K644	E715	I868	SER	L1022	P1158	F1225	F1225	L1430	LEU	PRO
PRO	L645	K644	S803	G869	ALA	R1023	P1158	V1226	V1226	M1367	MET	SER
THR	P568	L645	Y804	R940	SER	R1023	P1158	K1300	K1300	H1368	THR	THR
SER	K569	F646	V719	E870	SER	K1092	T1161	V1228	V1228	A1369	SER	PRO
PRO	P570	G647	L808	D871	K1093	K1093	V1162	W1228	W1228	G1365	PRO	PRO
THR	L571	G647	L808	D871	V1094	K1093	V1162	V1228	V1228	R1366	ALA	THR
PRO	L571	N648	T809	G872	V1095	K1093	V1162	V1228	V1228	R1366	LEU	THR
SER	W572	I649	P810	H873	T1096	K1093	V1162	V1228	V1228	R1366	LEU	THR
			F813	D874	T1028	K1093	V1162	V1228	V1228	R1366	LEU	THR
			A725	A875	A952	K1093	V1162	V1228	V1228	R1366	LEU	THR



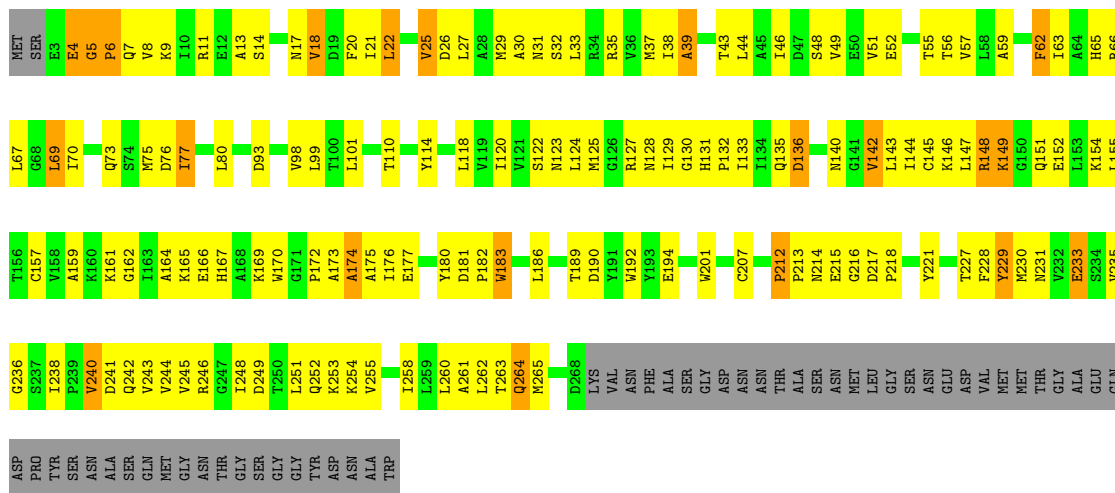






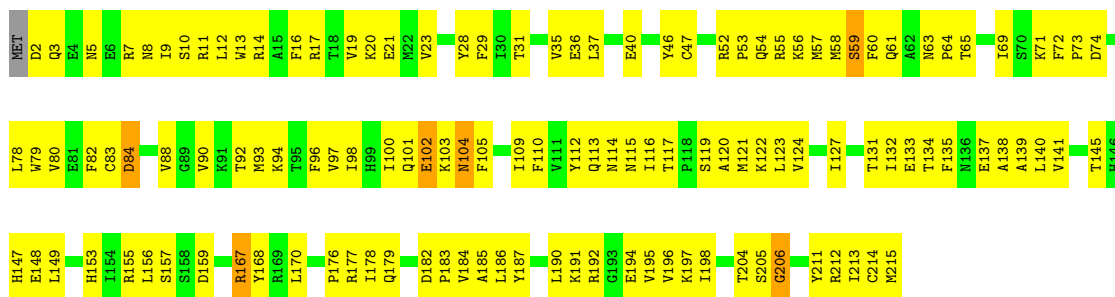
• Molecule 5: DNA-directed RNA polymerase II 45 kDa polypeptide

Chain C: 37% 40% 7% 16%




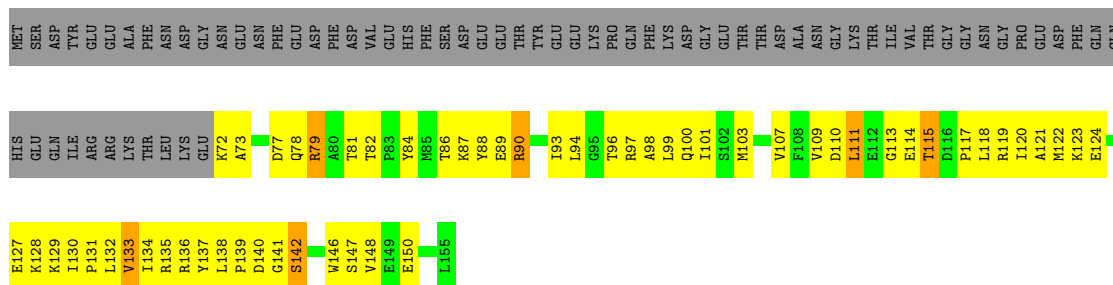
• Molecule 6: DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide

Chain E: 40% 56%



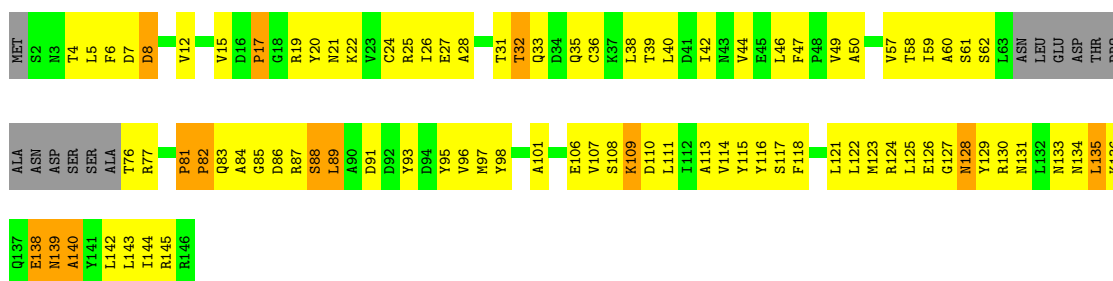
• Molecule 7: DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide

Chain F: 



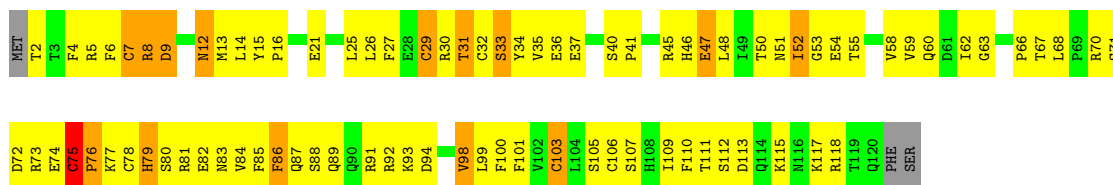
- Molecule 8: DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide

Chain H: 

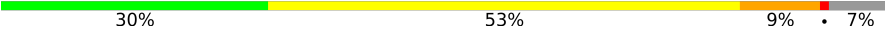


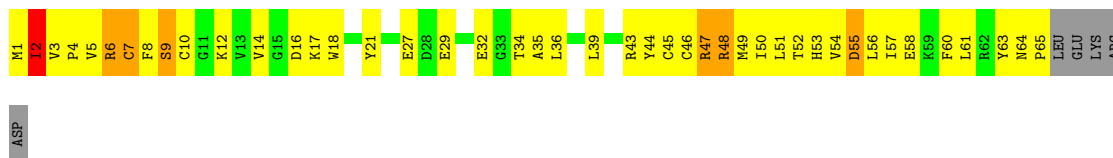
- Molecule 9: DNA-directed RNA polymerase II 14.2 kDa polypeptide

Chain I: 



- Molecule 10: DNA-directed RNA polymerases I, II, and III 8.3 kDa polypeptide

Chain J: 



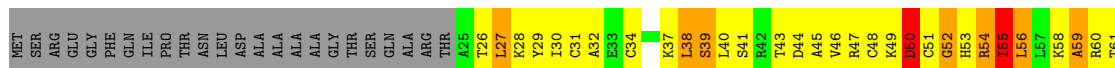
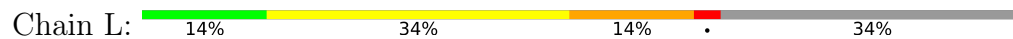
- Molecule 11: DNA-directed RNA polymerase II 13.6 kDa polypeptide

Chain K: 





- Molecule 12: 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, $\alpha$ , $\beta$ , $\gamma$	169.65Å 222.34Å 194.32Å 90.00° 101.67° 90.00°	Depositor
Resolution (Å)	40.00 – 4.25 39.94 – 4.20	Depositor EDS
% Data completeness (in resolution range)	95.8 (40.00-4.25) 84.6 (39.94-4.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.00 (at 4.13Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.349 , 0.398 0.319 , 0.372	Depositor DCC
$R_{free}$ test set	5207 reflections (10.13%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	96.8	Xtrriage
Anisotropy	0.402	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 104.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.71	EDS
Total number of atoms	28491	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	130.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.04% 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, UTP, 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	3.97	33/244 (13.5%)	3.92	44/380 (11.6%)
2	T	4.38	37/311 (11.9%)	3.99	55/477 (11.5%)
3	A	0.41	0/11048	0.71	5/14936 (0.0%)
4	B	0.46	0/8890	0.72	1/11990 (0.0%)
5	C	0.48	0/2133	0.76	2/2891 (0.1%)
6	E	0.36	0/1788	0.65	0/2406
7	F	0.40	0/691	0.64	0/933
8	H	0.40	0/1086	0.73	0/1470
9	I	0.48	0/989	0.76	1/1331 (0.1%)
10	J	0.53	0/541	0.78	0/727
11	K	0.46	0/937	0.68	0/1265
12	L	0.48	0/365	0.78	0/485
All	All	0.72	70/29023 (0.2%)	0.92	108/39291 (0.3%)

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	R	0	1
2	T	1	2
All	All	1	3

The worst 5 of 70 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	3	DG	O3'-P	-35.59	1.18	1.61
2	T	4	DA	O3'-P	-30.26	1.24	1.61
1	R	10	A	P-OP1	-23.82	1.08	1.49
1	R	5	A	O3'-P	-22.75	1.33	1.61

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	3	DG	C3'-O3'	-20.45	1.17	1.44

The worst 5 of 108 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	T	3	DG	O3'-P-O5'	-28.04	50.72	104.00
1	R	9	G	P-O3'-C3'	28.01	153.31	119.70
1	R	4	G	OP2-P-O3'	-25.63	48.80	105.20
2	T	3	DG	P-O3'-C3'	-25.44	89.17	119.70
2	T	7	DC	O5'-P-OP1	25.27	141.02	110.70

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	T	7	DC	C3'

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	R	10	A	Sidechain
2	T	7	DC	Sidechain
2	T	9	DC	Sidechain

## 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	217	0	110	48	0
2	T	279	0	160	69	0
3	A	10857	0	10959	1037	18
4	B	8720	0	8746	901	13
5	C	2095	0	2052	164	0
6	E	1752	0	1776	133	0
7	F	679	0	701	67	0
8	H	1068	0	1040	134	0
9	I	971	0	933	105	59
10	J	532	0	544	77	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
11	K	919	0	929	85	0
12	L	363	0	388	55	0
13	A	1	0	0	0	0
13	R	1	0	0	0	0
14	R	29	0	8	8	0
15	A	2	0	0	0	0
15	B	1	0	0	0	0
15	C	1	0	0	0	0
15	I	2	0	0	2	0
15	J	1	0	0	1	0
15	L	1	0	0	0	0
All	All	28491	0	28346	2619	59

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 46.

The worst 5 of 2619 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:R:10:A:P	1:R:10:A:OP1	1.08	1.47
2:T:6:DC:H2''	2:T:7:DC:C5'	1.54	1.36
2:T:7:DC:C5'	2:T:7:DC:O5'	1.78	1.30
3:A:567:LYS:HB2	3:A:568:PRO:HD2	1.18	1.17
2:T:6:DC:C2'	2:T:7:DC:H5'	1.75	1.15

The worst 5 of 59 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)
4:B:736:THR:CB	9:I:79:HIS:CD2[2_556]	0.78	1.42
4:B:736:THR:OG1	9:I:79:HIS:NE2[2_556]	0.83	1.37
9:I:81:ARG:O	9:I:81:ARG:CB[2_556]	1.01	1.19
4:B:736:THR:OG1	9:I:79:HIS:CD2[2_556]	1.02	1.18
3:A:923:LEU:CD1	9:I:35:VAL:C[4_546]	1.06	1.14



## 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	
3	A	1365/1733 (79%)	1023 (75%)	252 (18%)	90 (7%)	1	18
4	B	1077/1224 (88%)	839 (78%)	169 (16%)	69 (6%)	1	19
5	C	264/318 (83%)	208 (79%)	40 (15%)	16 (6%)	1	19
6	E	212/215 (99%)	170 (80%)	33 (16%)	9 (4%)	3	25
7	F	82/155 (53%)	64 (78%)	15 (18%)	3 (4%)	3	28
8	H	129/146 (88%)	93 (72%)	21 (16%)	15 (12%)	0	6
9	I	117/122 (96%)	93 (80%)	15 (13%)	9 (8%)	1	15
10	J	63/70 (90%)	48 (76%)	11 (18%)	4 (6%)	1	19
11	K	112/120 (93%)	96 (86%)	15 (13%)	1 (1%)	17	56
12	L	44/70 (63%)	22 (50%)	12 (27%)	10 (23%)	0	1
All	All	3465/4173 (83%)	2656 (77%)	583 (17%)	226 (6%)	1	19

5 of 226 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	31	SER
3	A	48	ALA
3	A	55	ASP
3	A	56	PRO
3	A	74	MET

### 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	
3	A	1206/1520 (79%)	1128 (94%)	78 (6%)	17	44
4	B	952/1061 (90%)	886 (93%)	66 (7%)	15	43
5	C	234/274 (85%)	222 (95%)	12 (5%)	24	51
6	E	196/197 (100%)	189 (96%)	7 (4%)	35	60
7	F	74/137 (54%)	68 (92%)	6 (8%)	11	38
8	H	117/128 (91%)	112 (96%)	5 (4%)	29	55
9	I	113/116 (97%)	104 (92%)	9 (8%)	12	38
10	J	60/65 (92%)	56 (93%)	4 (7%)	16	43
11	K	99/102 (97%)	90 (91%)	9 (9%)	9	32
12	L	40/57 (70%)	35 (88%)	5 (12%)	4	22
All	All	3091/3657 (84%)	2890 (94%)	201 (6%)	17	44

5 of 201 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	732	SER
4	B	1185	CYS
12	L	68	GLU
4	B	791	THR
4	B	996	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 76 such sidechains are listed below:

Mol	Chain	Res	Type
5	C	73	GLN
10	J	53	HIS
5	C	123	ASN
6	E	101	GLN
11	K	110	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	R	10/10 (100%)	3 (30%)	3 (30%)

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	R	2	U
1	R	5	A
1	R	6	G

All (3) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	R	1	A
1	R	4	G
1	R	5	A

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 11 ligands modelled in this entry, 10 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
14	UTP	R	3000	1,13	22,30,30	2.78	8 (36%)	27,47,47	3.59	9 (33%)

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
14	UTP	R	3000	1,13	1/1/7/7	4/20/38/38	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
14	R	3000	UTP	C3'-C4'	-5.50	1.38	1.53
14	R	3000	UTP	C5'-C4'	5.26	1.68	1.51
14	R	3000	UTP	O2'-C2'	5.17	1.55	1.43
14	R	3000	UTP	O3'-C3'	-4.34	1.32	1.43
14	R	3000	UTP	C2'-C1'	-4.33	1.47	1.53

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
14	R	3000	UTP	O4'-C4'-C5'	13.11	152.51	109.37
14	R	3000	UTP	O5'-C5'-C4'	6.78	132.33	108.99
14	R	3000	UTP	C5'-C4'-C3'	-6.59	90.48	115.18
14	R	3000	UTP	O3'-C3'-C2'	-4.56	97.06	111.82
14	R	3000	UTP	O4'-C4'-C3'	4.34	113.69	105.11

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
14	R	3000	UTP	C4'

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
14	R	3000	UTP	C5'-O5'-PA-O2A
14	R	3000	UTP	C5'-O5'-PA-O3A
14	R	3000	UTP	C4'-C5'-O5'-PA
14	R	3000	UTP	O4'-C4'-C5'-O5'

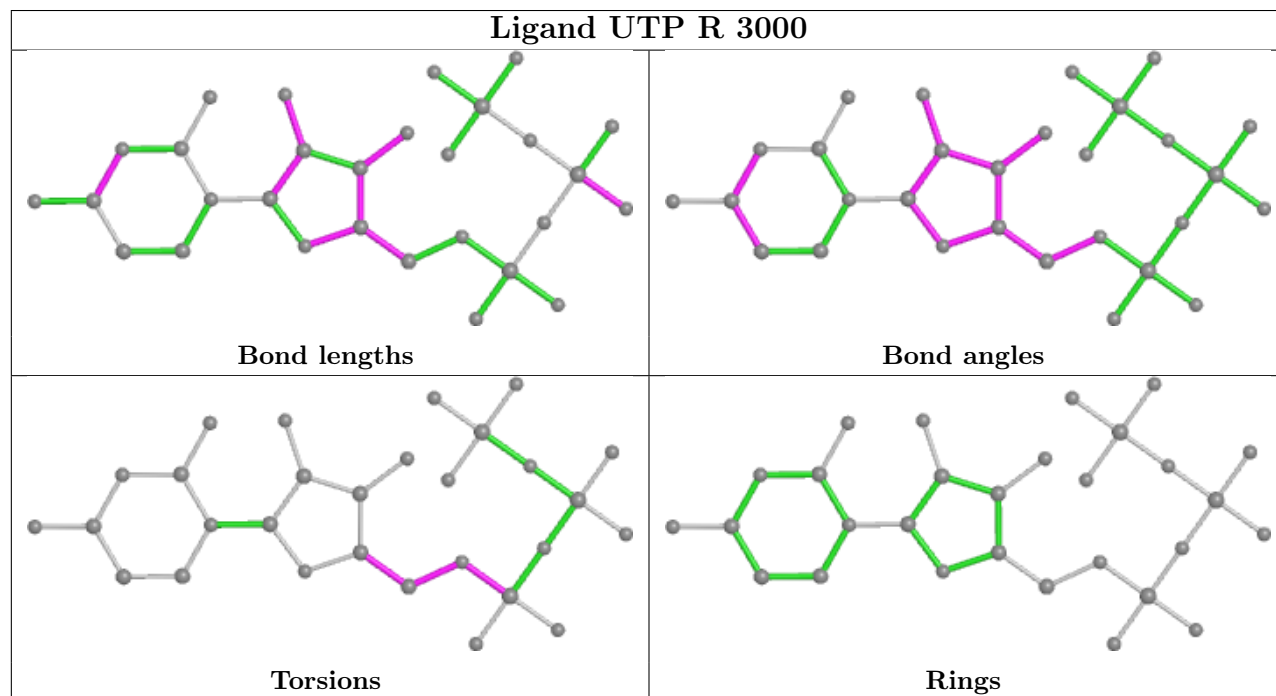
There are no ring outliers.

1 monomer is involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
14	R	3000	UTP	8	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will

also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



## 5.7 Other polymers [i](#)

There are no such residues in this entry.

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

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	T	3
1	R	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	T	7:DC	O3'	8:DT	P	1.39

*Continued on next page...*

*Continued from previous page...*

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	R	5:A	O3'	6:G	P	1.33
1	T	4:DA	O3'	5:DT	P	1.24
1	T	3:DG	O3'	4:DA	P	1.18

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.

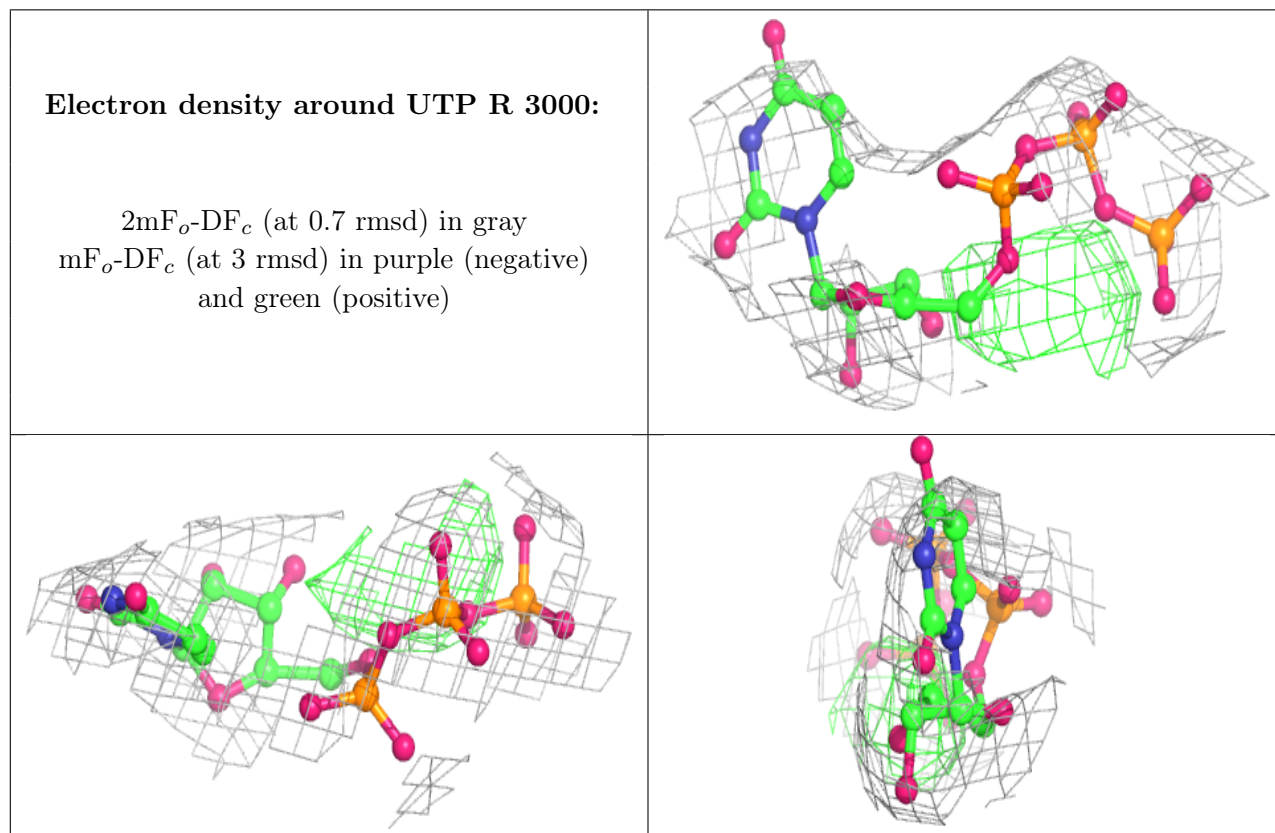
### 6.3 Carbohydrates [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.

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



## 6.5 Other polymers [i](#)

Unable to reproduce the depositors R factor - this section is therefore empty.