wwPDB X-ray Structure Validation Summary Report (i)

Nov 1, 2023 – 10:55 AM EDT

PDB ID	:	3HOY
Title	:	Complete RNA polymerase II elongation complex VI
Authors	:	Sydow, J.F.; Brueckner, F.; Cheung, A.C.M.; Damsma, G.E.; Dengl, S.; Lehmann, E.; Vassylyev, D.; Cramer, P.
Deposited on	:	2009-06-03
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*
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:

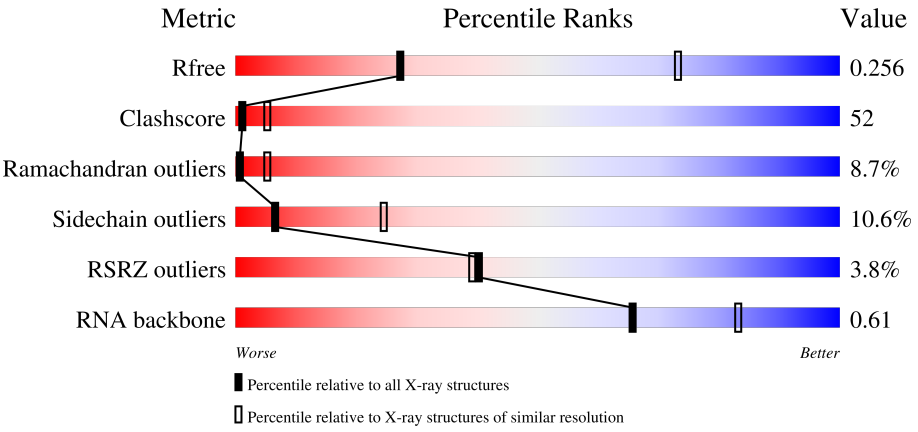
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 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 3.40 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.



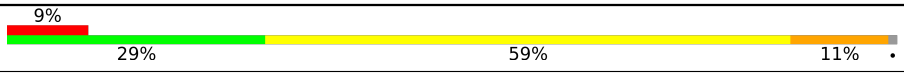
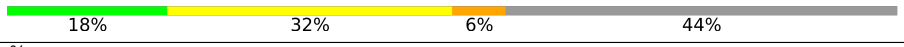
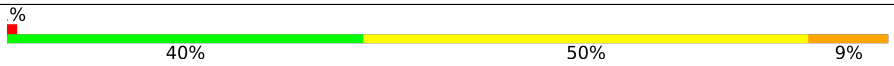
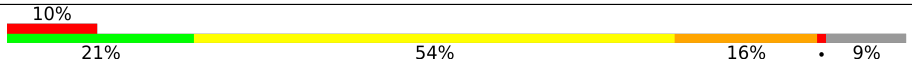
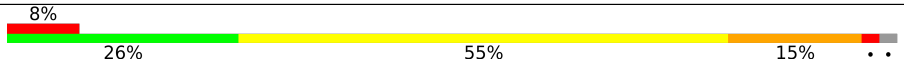
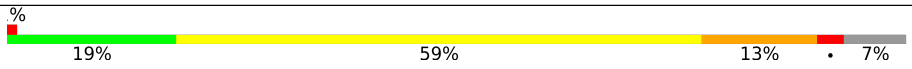
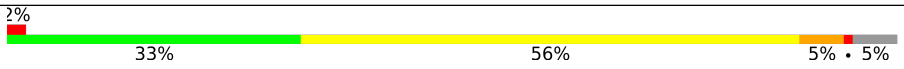
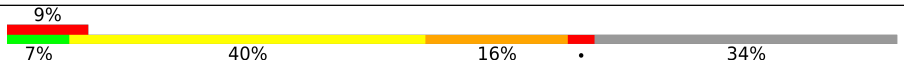
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)
RNA backbone	3102	1006 (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 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	A	1733	2% 28% 43% 10% • 18%
2	B	1224	3% 27% 51% 11% • 10%
3	C	347	% 22% 43% 12% 23%
4	D	221	29% 42% 10% 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	E	215	
6	F	155	
7	G	171	
8	H	146	
9	I	122	
10	J	70	
11	K	120	
12	L	70	
13	T	41	
14	N	41	
15	P	20	

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
13	BRU	T	20	-	-	X	-

2 Entry composition (i)

There are 17 unique types of molecules in this entry. The entry contains 31803 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA-directed RNA polymerase II subunit RPB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1419	Total	C	N	O	S	0	0	0
			11166	7036	1953	2115	62			

- Molecule 2 is a protein called DNA-directed RNA polymerase II subunit RPB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1105	Total	C	N	O	S	0	0	0
			8786	5564	1541	1627	54			

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

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

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-28	MET	-	expression tag	UNP P16370
C	-27	GLY	-	expression tag	UNP P16370
C	-26	SER	-	expression tag	UNP P16370
C	-25	HIS	-	expression tag	UNP P16370
C	-24	HIS	-	expression tag	UNP P16370
C	-23	HIS	-	expression tag	UNP P16370
C	-22	HIS	-	expression tag	UNP P16370
C	-21	HIS	-	expression tag	UNP P16370
C	-20	HIS	-	expression tag	UNP P16370
C	-19	SER	-	expression tag	UNP P16370
C	-18	ASN	-	expression tag	UNP P16370
C	-17	SER	-	expression tag	UNP P16370
C	-16	GLY	-	expression tag	UNP P16370
C	-15	LEU	-	expression tag	UNP P16370

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-14	ASN	-	expression tag	UNP P16370
C	-13	ASP	-	expression tag	UNP P16370
C	-12	ILE	-	expression tag	UNP P16370
C	-11	PHE	-	expression tag	UNP P16370
C	-10	GLU	-	expression tag	UNP P16370
C	-9	ALA	-	expression tag	UNP P16370
C	-8	GLN	-	expression tag	UNP P16370
C	-7	LYS	-	expression tag	UNP P16370
C	-6	ILE	-	expression tag	UNP P16370
C	-5	GLU	-	expression tag	UNP P16370
C	-4	TRP	-	expression tag	UNP P16370
C	-3	HIS	-	expression tag	UNP P16370
C	-2	GLU	-	expression tag	UNP P16370
C	-1	ASP	-	expression tag	UNP P16370
C	0	THR	-	expression tag	UNP P16370
C	1	GLY	-	expression tag	UNP P16370

- Molecule 4 is a protein called DNA-directed RNA polymerase II subunit RPB4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	178	Total	C	N	O	S	0	0	0
			1365	845	242	276	2			

- Molecule 5 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	213	Total	C	N	O	S	0	0	0
			1744	1107	308	318	11			

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	87	Total	C	N	O	S	0	0	0
			705	451	119	132	3			

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB7.

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

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

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

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB9.

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

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

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

- Molecule 11 is a protein called DNA-directed RNA polymerase II subunit RPB11.

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

- Molecule 12 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

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

- Molecule 13 is a DNA chain called 5'-D(*CP*CP*AP*AP*GP*CP*TP*CP*AP*AP*G*TP *AP*CP*TP*TP*AP*CP*GP*CP*CP*(BRU)P*GP*GP*TP*CP*AP*TP*TP*AP*CP*T P*AP*GP*TP*AP*CP*TP*GP*CP*C)-3'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
13	T	19	Total	Br	C	N	O	P	8	0	0
			382	1	184	64	115	18			

- Molecule 14 is a DNA chain called 5'-D(*CP*CP*GP*GP*CP*AP*GP*TP*AP*CP*TP*A P*GP*TP*AP*AP*AP*CP*TP*AP*GP*TP*AP*TP*T*GP*AP*AP*AP*GP*TP*AP*C P*TP*TP*GP*AP*GP*CP*TP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
14	N	7	Total	C	N	O	P	11	0	0
			145	70	32	37	6			

- Molecule 15 is a RNA chain called 5'-R(*UP*AP*UP*AP*UP*GP*CP*A*UP*AP*AP*AP *GP*AP*CP*CP*AP*GP*GP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
15	P	10	Total	C	N	O	P	0	0	0
			213	97	43	64	9			

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

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

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

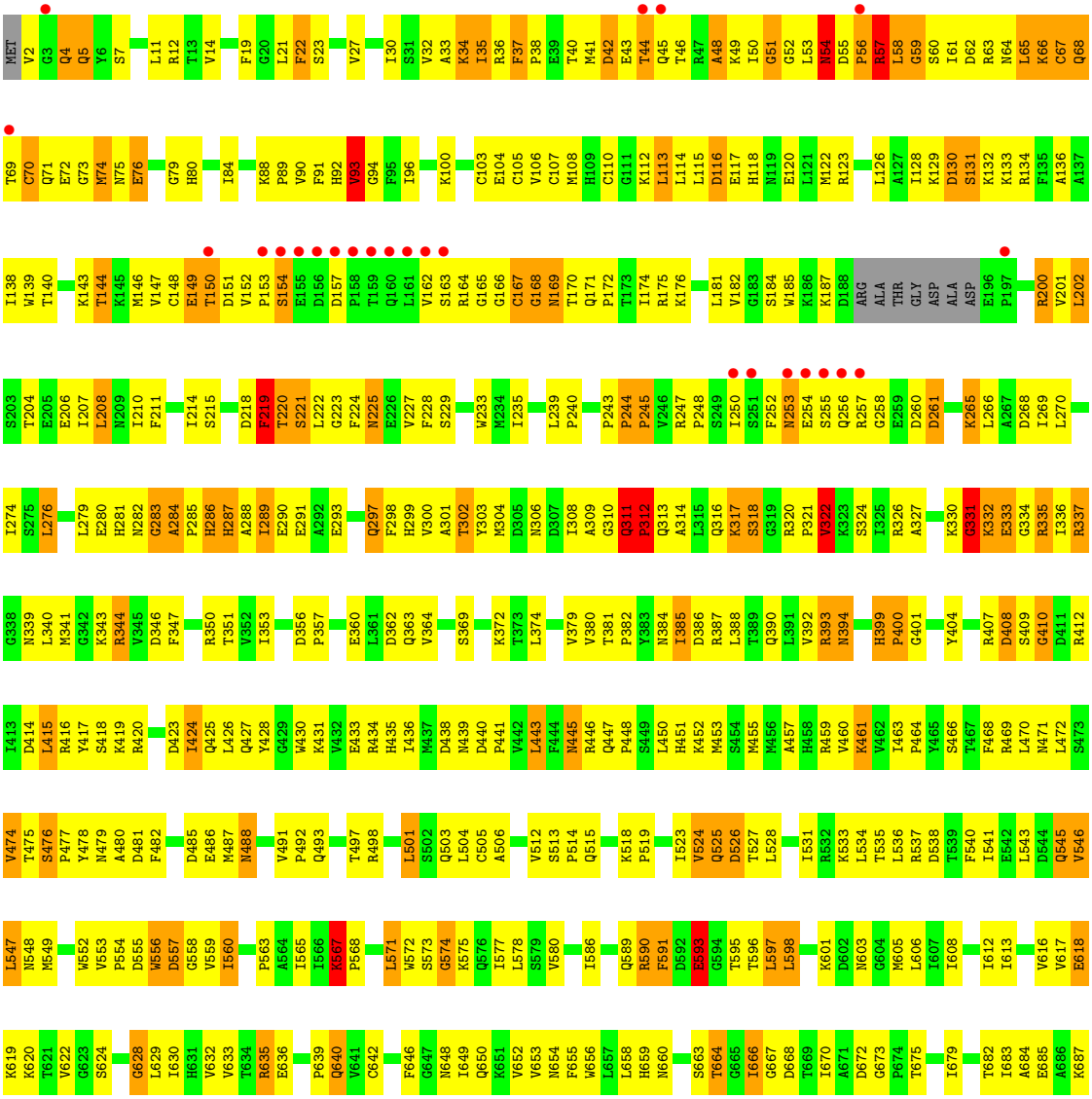
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
17	A	1	Total	Mg	0	0
			1	1		

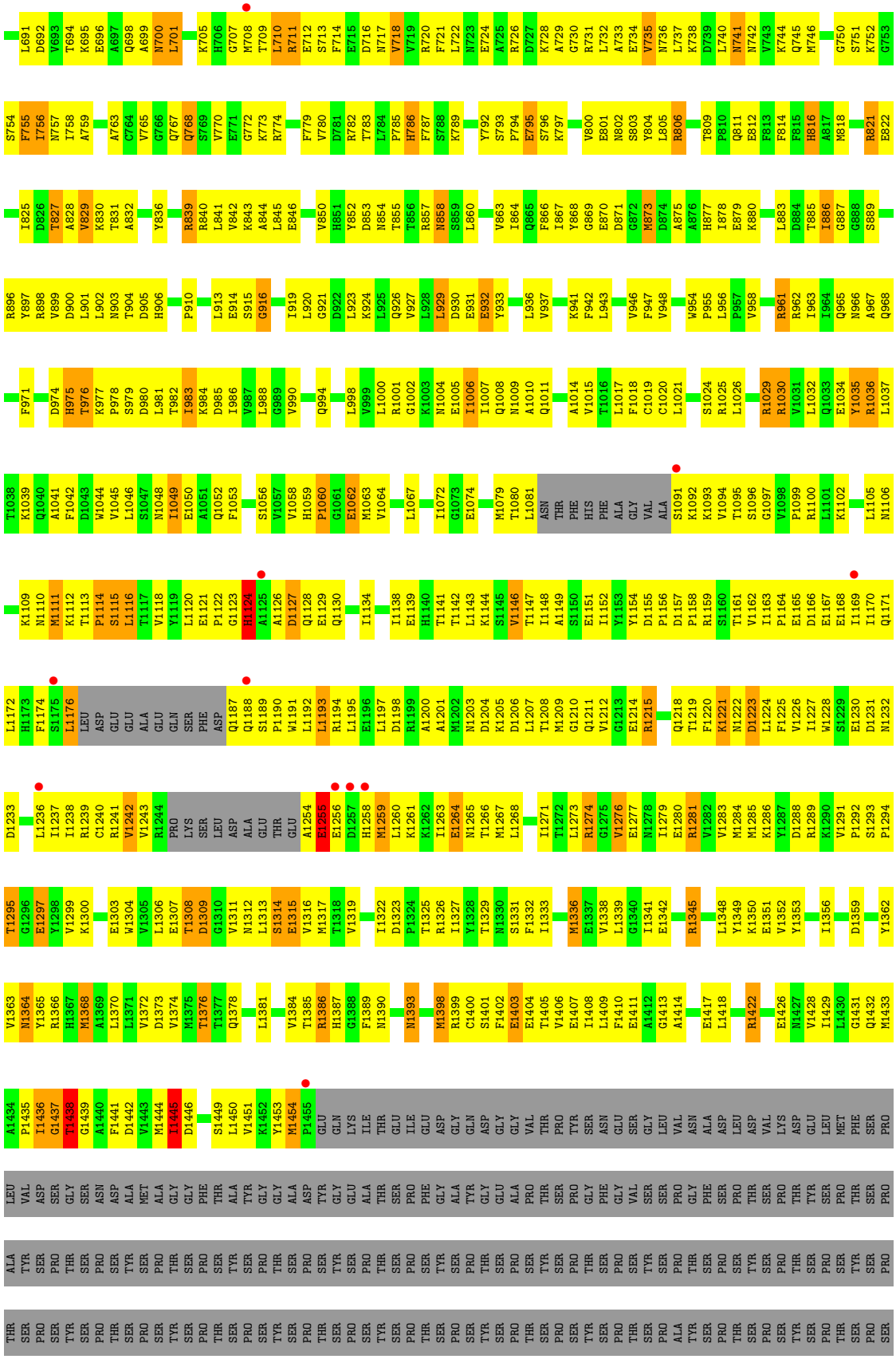
3 Residue-property plots (i)

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: DNA-directed RNA polymerase II subunit RPB1

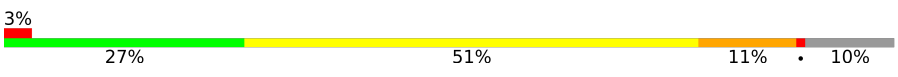
Chain A:

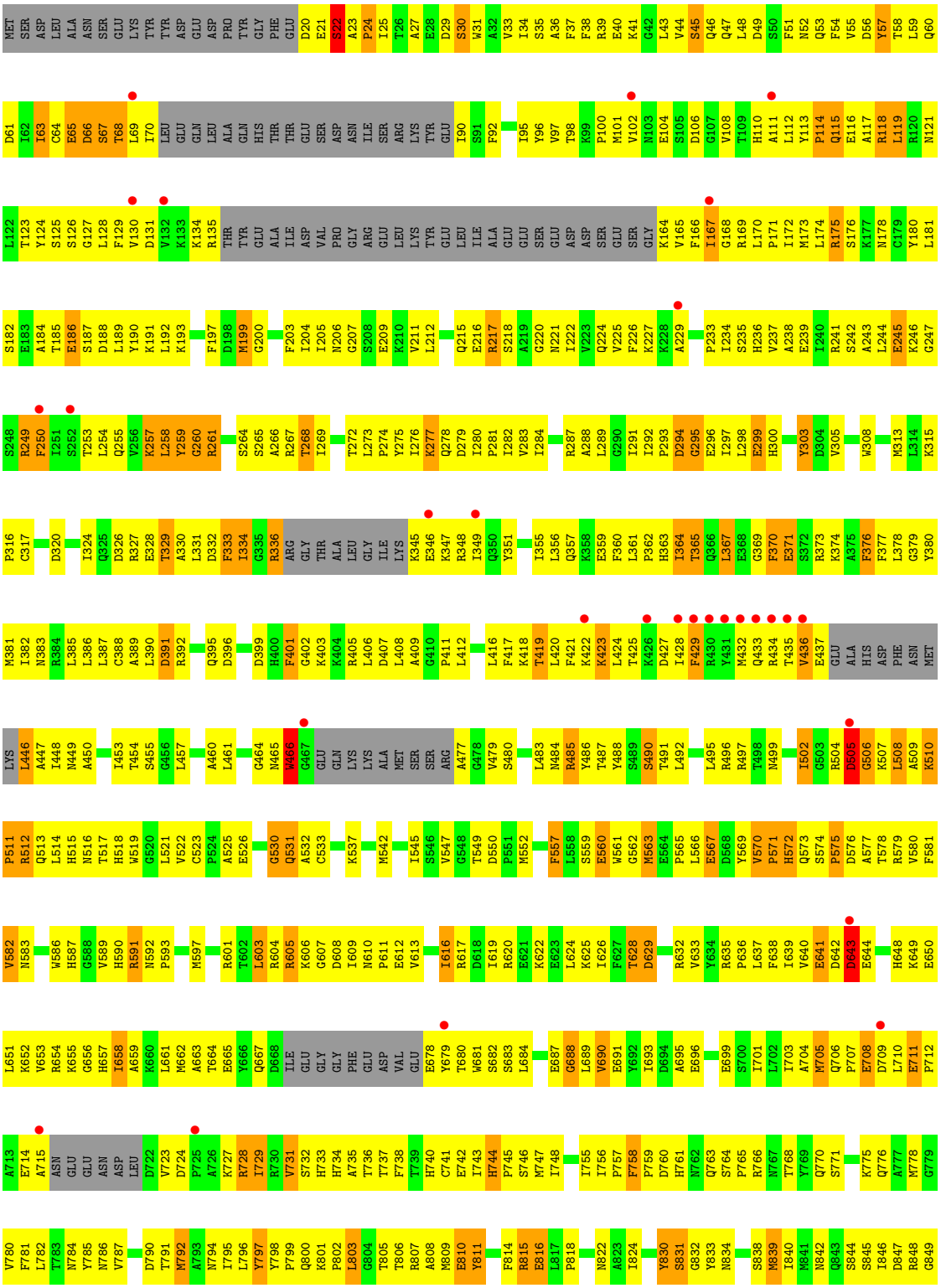


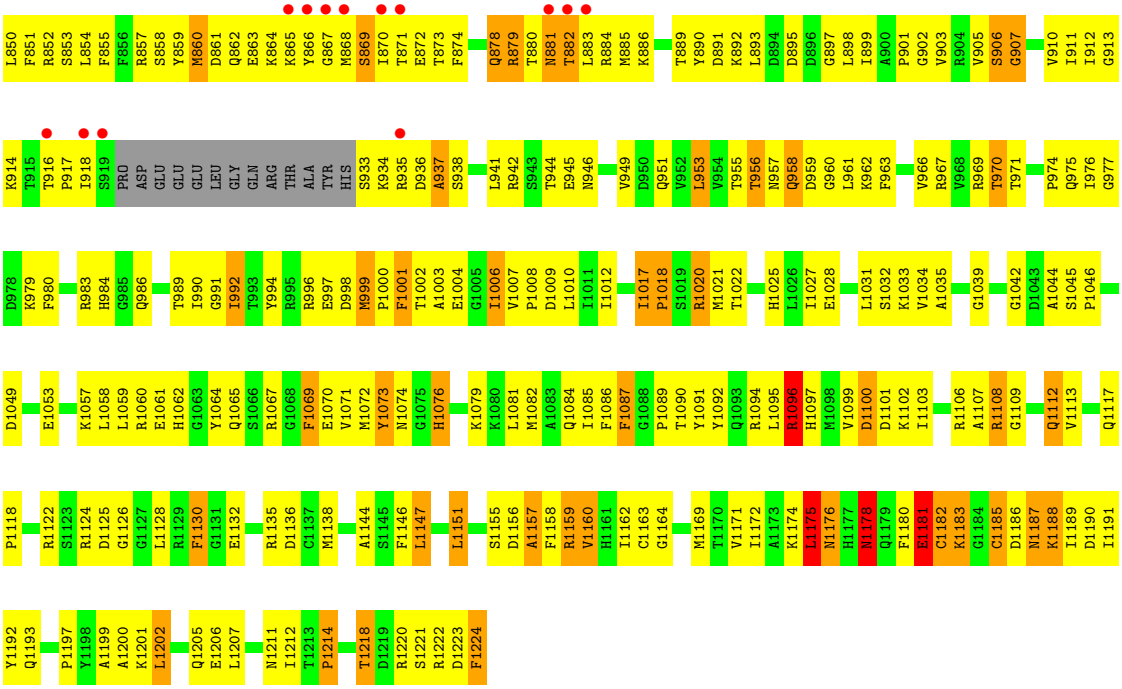

L691 D692 V693 T694 K695 E696 A697 Q698 A699 N700 L701 K705 H706 G707 M708 T709 L710 R711 E712 S713 F714 E715 D716 N717 V718 V719 R720 F721 L722 N723 E724 A725 R726 D727 K728 A729 G730 R731 L732 A733 E734 V735 N736 L737 K738 D739 L740 N741 N742 V743 K744 Q745 M746 G750 S751 K752 G753
S754 F755 I756 N757 I758 A759 A763 C764 V765 G766 Q767 Q768 S769 V770 E771 G772 K773 R774 F779 V780 D781 R782 T783 L784 P785 H786 F787 S788 K789 Y792 S793 P794 E795 S796 K797 V800 E801 N802 S803 Y804 L805 R806 T809 P810 Q811 E812 F813 F814 F815 H816 A817 M818 R821 E822
I825 D826 T827 A828 V829 K830 T831 A832 Y836 R839 R840 L841 V842 K843 A844 L845 E846 V850 H851 Y852 D853 N854 T855 T856 R857 N858 S859 L860 V863 I864 Q865 F866 I867 Y868 G869 E870 D871 G872 M873 D874 A875 A876 H877 I878 E879 K880 L883 D884 T885 I886 G887 G888 S889
R896 Y897 R898 V899 D900 L901 L902 N903 T904 D905 H906 P910 L913 E914 S915 G916 I919 L920 G921 D922 L923 K924 L925 Q926 V927 L928 L929 D930 E931 E932 Y933 L936 V937 K941 F942 L943 V946 F947 V948 W954 P955 L956 P957 V958 R961 R962 I963 I964 Q965 N966 A967 Q968
F971 D974 H975 T976 K977 P978 S979 D980 L981 T982 I983 K984 D985 I986 V987 L988 G989 V990 Q994 L998 V999 L1000 R1001 G1002 K1003 N1004 E1005 I1006 I1007 Q1008 N1009 A1010 Q1011 A1014 V1015 T1016 L1017 F1018 C1019 C1020 L1021 S1024 R1025 L1026 R1029 R1030 V1031 L1032 Q1033 E1034 Y1035 R1036 L1037
T1038 K1039 Q1040 A1041 F1042 D1043 W1044 V1045 L1046 S1047 N1048 I1049 E1050 A1051 Q1052 F1053 S1056 V1057 V1058 H1059 P1060 G1061 E1062 M1063 V1064 L1067 I1072 G1073 E1074 M1079 T1080 L1081 ASN THR PHE HIS PHE ALA GLY VAL ALA S1091 K1092 K1093 V1094 T1095 S1096 G1097 V1098 P1099 R1100 L1101 K1102 L1105 N1106
K1109 N1110 M1111 K1112 T1113 P1114 S1115 L1116 T1117 V1118 Y1119 L1120 E1121 P1122 G1123 H1124 A1125 A1126 D1127 Q1128 E1129 Q1130 I1134 I1138 E1139 H1140 T1141 T1142 L1143 K1144 S1145 V1146 T1147 I1148 A1149 S1150 E1151 I1152 Y1153 Y1154 D1155 P1156 D1157 P1158 R1159 S1160 T1161 V1162 I1163 P1164 E1165 D1166 E1167 E1168 I1169 Q1171
L1172 H1173 F1174 S1175 L1176 LEU ASP GLU GLU ALA GLU GLN SER PHE ASP Q1187 Q1188 S1189 P1190 W1191 L1192 L1193 R1194 L1195 E1196 L1197 D1198 R1199 A1200 A1201 M1202 N1203 D1204 K1205 D1206 L1207 T1208 M1209 G1210 Q1211 V1212 G1213 E1214 R1215 Q1218 T1219 F1220 K1221 N1222 D1223 L1224 F1225 V1226 I1227 W1228 S1229 E1230 D1231 N1232
D1233 L1236 I1237 I1238 R1239 C1240 R1241 V1242 V1243 R1244 PRO LYS SER LEU ASP ALA GLU THR GLU A1254 E1255 E1256 D1257 H1258 M1259 L1260 K1261 K1262 I1263 E1264 N1265 T1266 M1267 L1268 I1271 T1272 L1273 R1274 G1275 V1276 E1277 N1278 I1279 E1280 R1281 V1282 V1283 M1284 M1285 K1286 Y1287 D1288 R1289 K1290 V1291 P1292 S1293 P1294
T1295 G1296 E1297 Y1298 V1299 K1300 E1303 W1304 V1305 L1306 E1307 T1308 D1309 G1310 V1311 N1312 L1313 S1314 E1315 V1316 M1317 T1318 V1319 I1322 D1323 P1324 T1325 R1326 I1327 Y1328 T1329 N1330 S1331 F1332 I1333 M1336 E1337 V1338 L1339 G1340 I1341 E1342 R1345 L1348 Y1349 K1350 E1351 V1352 Y1353 I1356 D1359 Y1362
V1363 N1364 Y1365 R1366 H1367 M1368 A1369 L1370 L1371 V1372 D1373 V1374 M1375 T1376 T1377 Q1378 L1381 V1384 T1385 R1386 H1387 C1388 F1389 N1390 N1393 M1398 R1399 C1400 S1401 F1402 E1403 E1404 T1405 V1406 E1407 I1408 L1409 F1410 E1411 A1412 G1413 A1414 E1417 L1418 R1422 E1426 N1427 V1428 I1429 L1430 G1431 Q1432 M1433
A1434 P1435 I1436 G1437 T1438 G1439 A1440 F1441 D1442 V1443 M1444 I1445 D1446 S1449 L1450 V1451 K1452 Y1453 M1454 P1455 GLU GLN LYS ILE THR GLU ILE GLU ASP GLY GLN ASP GLY GLU VAL THR PRO TYR SER ASN GLU GLY LEU VAL ASN ALA ASP PHE GLY GLU SER LEU ASP LYS VAL ASP SER PRO
LEU VAL ASP SER GLY SER ASN ASP ALA MET ALA GLY GLY PHE THR ALA TYR GLY GLY ALA ASP TYR GLY GLU ALA THR SER PRO PHE GLY ALA TYR SER PRO GLY GLN ASP TYR GLY GLU ALA PRO THR SER PRO GLY GLY VAL PHE GLY VAL SER SER PRO GLY PHE SER PRO GLY ASP LEU ASP VAL LYS ASP THR TYR GLU LEU PHE SER PRO
ALA TYR SER PRO THR SER PRO SER TYR SER PRO THR SER PRO SER TYR SER PRO THR SER PRO SER

• Molecule 2: DNA-directed RNA polymerase II subunit RPB2

Chain B:

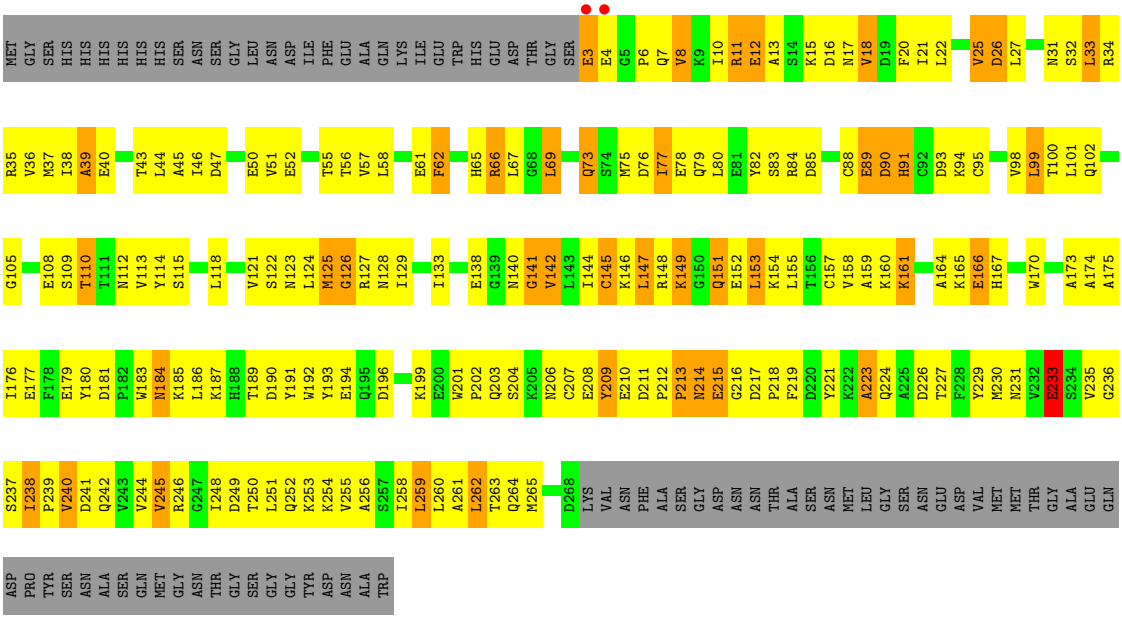






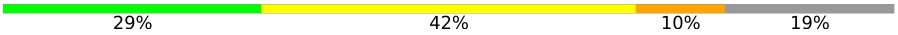
• Molecule 3: DNA-directed RNA polymerase II subunit RPB3

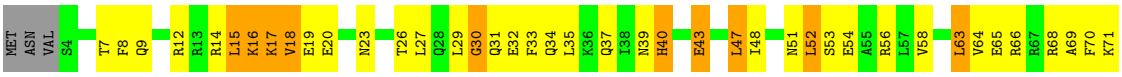
Chain C:

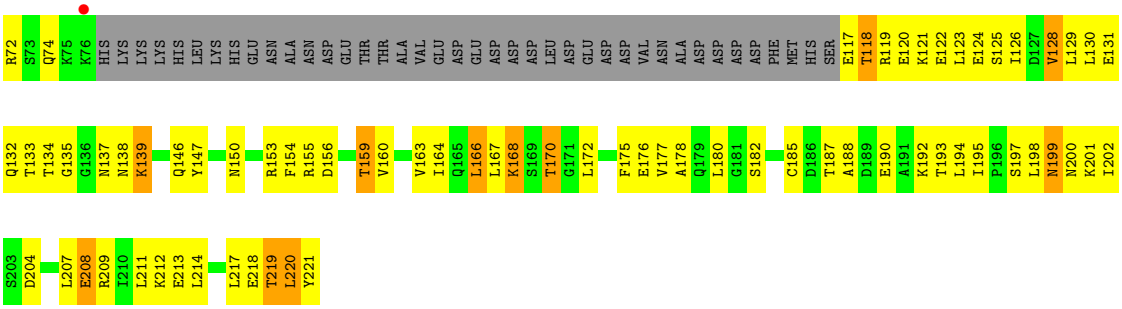


• Molecule 4: DNA-directed RNA polymerase II subunit RPB4

Chain D:

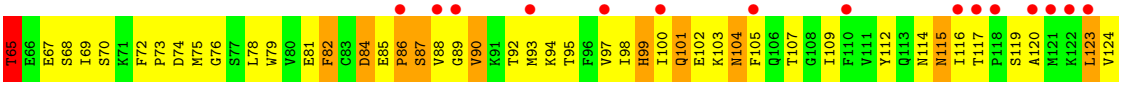


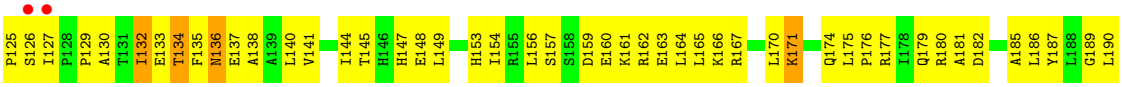




• Molecule 5: DNA-directed RNA polymerases I, II, and III subunit RPABC1

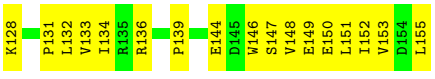
Chain E:





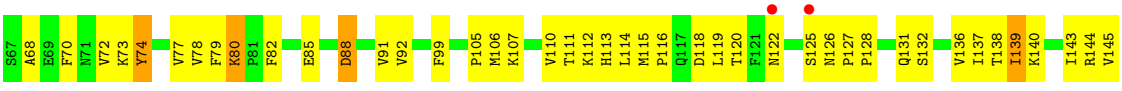
• Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC2

Chain F:

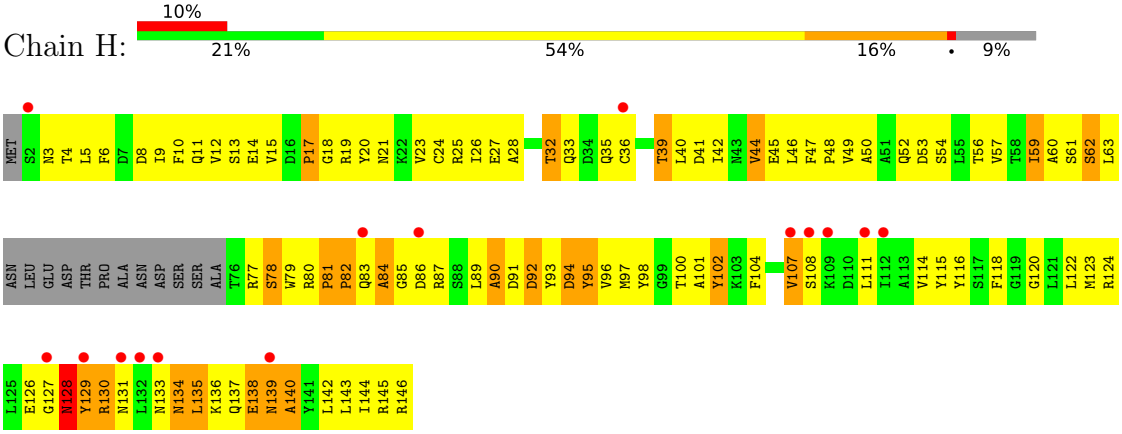


• Molecule 7: DNA-directed RNA polymerase II subunit RPB7

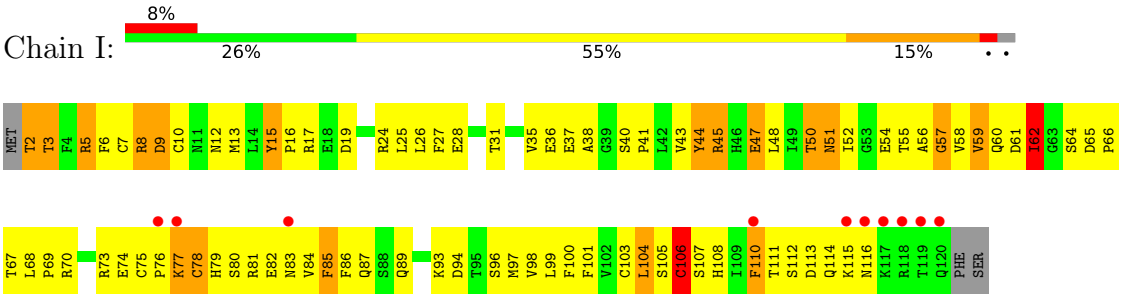
Chain G:



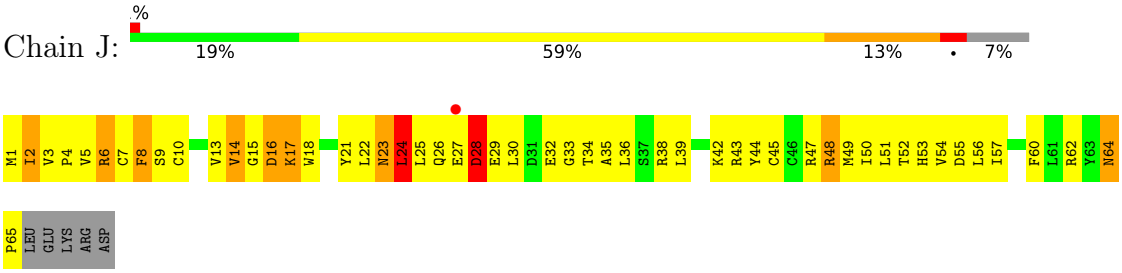
• Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC3



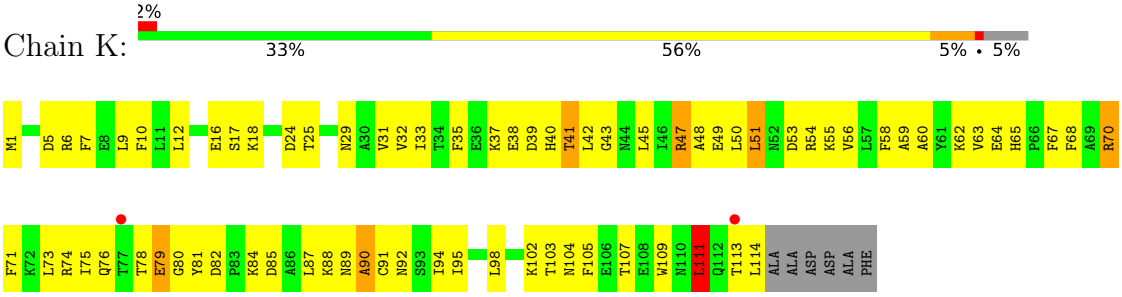
• Molecule 9: DNA-directed RNA polymerase II subunit RPB9



• Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC5

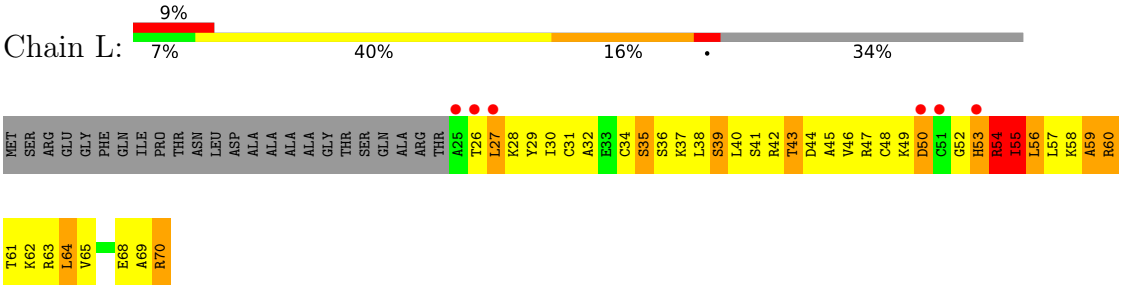


• Molecule 11: DNA-directed RNA polymerase II subunit RPB11



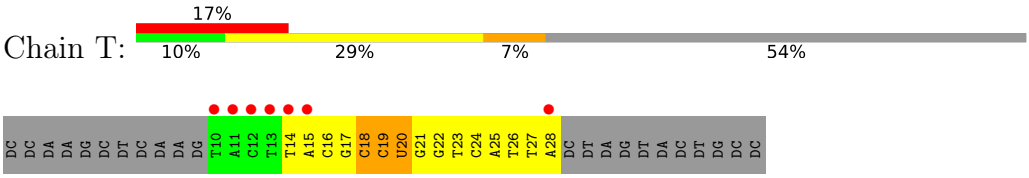
• Molecule 12: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L:



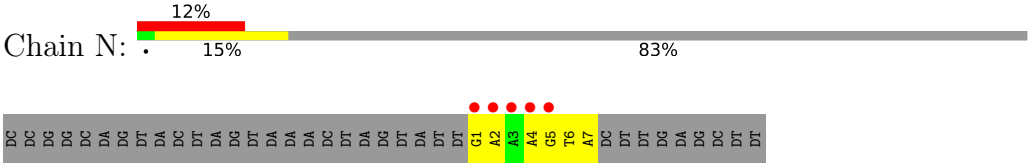
• Molecule 13: 5'-D(*CP*CP*AP*AP*GP*CP*TP*CP*AP*AP*G*TP*AP*CP*TP*TP*AP*CP*GP*CP*CP*(BRU)P*GP*GP*TP*CP*AP*TP*TP*AP*CP*TP*AP*GP*TP*AP*CP*TP*GP*CP*C)-3'

Chain T:



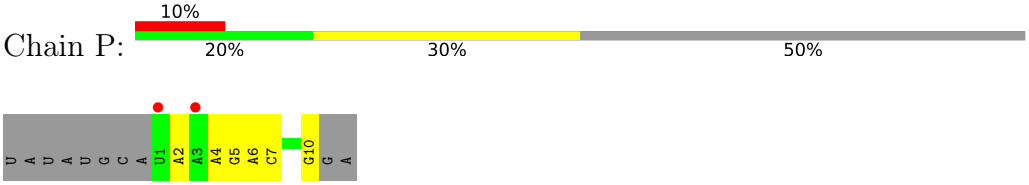
• Molecule 14: 5'-D(*CP*CP*GP*GP*CP*AP*GP*TP*AP*CP*TP*AP*GP*TP*AP*AP*AP*CP*TP*AP*GP*TP*AP*TP*T*GP*AP*AP*AP*GP*TP*AP*CP*TP*TP*GP*AP*GP*CP*TP*T)-3'

Chain N:



• Molecule 15: 5'-R(*UP*AP*UP*AP*UP*GP*CP*A*UP*AP*AP*AP*GP*AP*CP*CP*AP*GP*GP*A)-3'

Chain P:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	222.14Å 392.69Å 282.38Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.40 49.09 – 3.40	Depositor EDS
% Data completeness (in resolution range)	99.9 (50.00-3.40) 99.9 (49.09-3.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.88 (at 3.40Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.216 , 0.254 0.216 , 0.256	Depositor DCC
R_{free} test set	3325 reflections (1.98%)	wwPDB-VP
Wilson B-factor (Å^2)	89.1	Xtriage
Anisotropy	0.021	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 85.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.013 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.018 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	31803	wwPDB-VP
Average B, all atoms (Å^2)	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.04% 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: MG, ZN, BRU

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

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.50	0/11365	0.78	7/15367 (0.0%)
2	B	0.49	0/8957	0.74	1/12078 (0.0%)
3	C	0.51	0/2133	0.75	1/2891 (0.0%)
4	D	0.46	0/1374	0.74	1/1849 (0.1%)
5	E	0.46	0/1780	0.68	1/2395 (0.0%)
6	F	0.57	0/717	0.86	1/967 (0.1%)
7	G	0.52	0/1368	0.80	0/1844
8	H	0.47	0/1086	0.76	0/1470
9	I	0.45	0/989	0.68	0/1331
10	J	0.50	0/541	0.88	0/727
11	K	0.49	0/937	0.72	0/1265
12	L	0.60	0/365	0.84	0/485
13	T	0.56	0/403	0.82	0/617
14	N	0.79	0/164	0.77	0/252
15	P	0.58	0/239	0.80	0/371
All	All	0.50	0/32418	0.76	12/43909 (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	T	0	3

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	26	THR	N-CA-C	-8.03	89.31	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	39	ALA	N-CA-C	6.23	127.81	111.00
5	E	171	LYS	N-CA-C	-6.18	94.32	111.00
1	A	56	PRO	N-CA-C	-6.06	96.34	112.10
1	A	1445	ILE	CB-CA-C	-5.83	99.94	111.60

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
13	T	18	DC	Sidechain
13	T	19	DC	Sidechain
13	T	21	DG	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	A	11166	0	11248	1160	0
2	B	8786	0	8819	1073	0
3	C	2095	0	2051	275	0
4	D	1365	0	1325	116	0
5	E	1744	0	1772	190	0
6	F	705	0	731	72	0
7	G	1340	0	1357	149	0
8	H	1068	0	1040	174	0
9	I	971	0	929	128	0
10	J	532	0	542	84	0
11	K	919	0	929	97	0
12	L	363	0	388	74	0
13	T	382	0	215	32	0
14	N	145	0	80	9	0
15	P	213	0	111	4	0
16	A	2	0	0	0	0
16	B	1	0	0	0	0
16	C	1	0	0	0	0
16	I	2	0	0	0	0
16	J	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
16	L	1	0	0	0	0
17	A	1	0	0	0	0
All	All	31803	0	31537	3317	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 52.

The worst 5 of 3317 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)
14:N:5:DG:H2"	14:N:6:DT:H71	1.17	1.15
3:C:57:VAL:HG11	10:J:60:PHE:HB3	1.22	1.14
2:B:345:LYS:HE2	2:B:349:ILE:HD11	1.30	1.14
1:A:868:TYR:CE1	1:A:1064:VAL:HG11	1.83	1.14
3:C:189:THR:HG22	3:C:190:ASP:H	1.11	1.14

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	1409/1733 (81%)	1031 (73%)	264 (19%)	114 (8%)	1 5
2	B	1087/1224 (89%)	798 (73%)	191 (18%)	98 (9%)	1 4
3	C	264/347 (76%)	192 (73%)	49 (19%)	23 (9%)	1 5
4	D	174/221 (79%)	121 (70%)	35 (20%)	18 (10%)	0 3
5	E	211/215 (98%)	155 (74%)	41 (19%)	15 (7%)	1 7
6	F	85/155 (55%)	72 (85%)	10 (12%)	3 (4%)	3 21
7	G	169/171 (99%)	143 (85%)	20 (12%)	6 (4%)	3 21
8	H	129/146 (88%)	83 (64%)	26 (20%)	20 (16%)	0 0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
9	I	117/122 (96%)	78 (67%)	27 (23%)	12 (10%)	0 3
10	J	63/70 (90%)	39 (62%)	15 (24%)	9 (14%)	0 1
11	K	112/120 (93%)	91 (81%)	15 (13%)	6 (5%)	2 13
12	L	44/70 (63%)	18 (41%)	14 (32%)	12 (27%)	0 0
All	All	3864/4594 (84%)	2821 (73%)	707 (18%)	336 (9%)	1 5

5 of 336 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	57	ARG
1	A	59	GLY
1	A	67	CYS
1	A	70	CYS
1	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
1	A	1242/1520 (82%)	1106 (89%)	136 (11%)	6 23
2	B	958/1061 (90%)	859 (90%)	99 (10%)	7 26
3	C	234/299 (78%)	209 (89%)	25 (11%)	6 24
4	D	141/200 (70%)	124 (88%)	17 (12%)	5 18
5	E	195/197 (99%)	175 (90%)	20 (10%)	7 26
6	F	77/137 (56%)	70 (91%)	7 (9%)	9 32
7	G	152/152 (100%)	140 (92%)	12 (8%)	12 39
8	H	117/128 (91%)	108 (92%)	9 (8%)	13 40
9	I	113/116 (97%)	95 (84%)	18 (16%)	2 10
10	J	60/65 (92%)	52 (87%)	8 (13%)	4 15
11	K	99/102 (97%)	94 (95%)	5 (5%)	24 54
12	L	40/57 (70%)	34 (85%)	6 (15%)	3 12

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	3428/4034 (85%)	3066 (89%)	362 (11%)	6 24

5 of 362 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	1187	ASN
5	E	101	GLN
3	C	33	LEU
4	D	35	LEU
6	F	111	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 100 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	986	GLN
3	C	135	GLN
11	K	104	ASN
2	B	1065	GLN
3	C	31	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
15	P	9/20 (45%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

5.4 Non-standard residues in protein, DNA, RNA chains (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	$\#\|Z\|>2$
13	BRU	T	20	13,15	18,21,22	3.90	1 (5%)	26,30,33	1.16	2 (7%)

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
13	BRU	T	20	13,15	-	0/7/21/22	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
13	T	20	BRU	BR-C5	-16.43	1.50	1.88

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
13	T	20	BRU	C2'-C1'-N1	-2.83	107.25	113.77
13	T	20	BRU	C6-C5-C4	-2.80	117.83	120.67

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
13	T	20	BRU	8	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 9 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	A	1419/1733 (81%)	0.13	36 (2%) 57 55	21, 68, 106, 122	0
2	B	1105/1224 (90%)	0.20	42 (3%) 40 39	28, 78, 111, 122	0
3	C	266/347 (76%)	0.09	2 (0%) 86 85	38, 66, 91, 115	0
4	D	178/221 (80%)	0.10	1 (0%) 89 89	46, 77, 106, 115	0
5	E	213/215 (99%)	0.50	20 (9%) 8 10	47, 91, 115, 121	0
6	F	87/155 (56%)	-0.20	0 100 100	29, 49, 75, 86	0
7	G	171/171 (100%)	0.12	2 (1%) 79 77	45, 67, 96, 108	0
8	H	133/146 (91%)	0.71	15 (11%) 5 6	81, 98, 112, 118	0
9	I	119/122 (97%)	0.52	10 (8%) 11 13	64, 97, 116, 124	0
10	J	65/70 (92%)	-0.03	1 (1%) 73 72	42, 63, 88, 99	0
11	K	114/120 (95%)	0.18	2 (1%) 68 67	33, 70, 88, 101	0
12	L	46/70 (65%)	0.81	6 (13%) 3 4	53, 105, 121, 123	0
13	T	18/41 (43%)	1.66	7 (38%) 0 0	79, 107, 129, 134	1 (5%)
14	N	7/41 (17%)	2.38	5 (71%) 0 0	113, 116, 124, 126	1 (14%)
15	P	10/20 (50%)	1.01	2 (20%) 1 1	80, 99, 124, 127	0
All	All	3951/4696 (84%)	0.21	151 (3%) 40 39	21, 74, 111, 134	2 (0%)

The worst 5 of 151 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
13	T	28	DA	5.7
2	B	883	LEU	5.2
1	A	161	LEU	4.9
12	L	26	THR	4.8
1	A	1455	PRO	4.6

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{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
13	BRU	T	20	20/21	0.80	0.26	75,80,83,84	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
17	MG	A	2458	1/1	0.62	0.18	100,100,100,100	0
16	ZN	L	1071	1/1	0.95	0.06	119,119,119,119	0
16	ZN	I	1122	1/1	0.95	0.09	134,134,134,134	0
16	ZN	A	2456	1/1	0.97	0.09	85,85,85,85	0
16	ZN	I	1121	1/1	0.98	0.10	75,75,75,75	0
16	ZN	A	2457	1/1	0.99	0.15	53,53,53,53	0
16	ZN	B	2225	1/1	0.99	0.23	73,73,73,73	0
16	ZN	C	1269	1/1	1.00	0.11	65,65,65,65	0
16	ZN	J	1066	1/1	1.00	0.22	64,64,64,64	0

6.5 Other polymers (i)

There are no such residues in this entry.