



wwPDB X-ray Structure Validation Summary Report i

Feb 26, 2014 – 03:29 PM GMT

PDB ID : 3HOZ
Title : Complete RNA polymerase II elongation complex IV with a T-U mismatch and a frayed RNA 3'-guanine
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 : 3.65 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

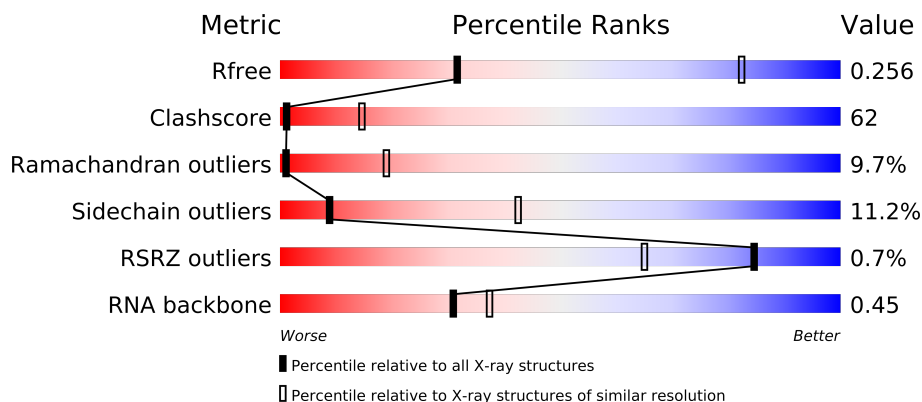
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 3.65 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.



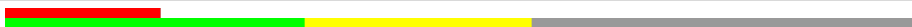


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	1021 (3.94-3.38)
Clashscore	79885	1025 (3.90-3.42)
Ramachandran outliers	78287	1195 (3.92-3.40)
Sidechain outliers	78261	1193 (3.92-3.40)
RSRZ outliers	66119	1021 (3.94-3.38)
RNA backbone	1838	1008 (4.52-2.80)

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 ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	1733	
2	B	1224	
3	C	347	
4	D	221	
5	E	215	
6	F	155	
7	G	171	
8	H	146	
9	I	122	
10	J	70	
11	K	120	
12	L	70	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
13	N	12	
14	T	26	
15	P	18	

2 Entry composition

There are 17 unique types of molecules in this entry. The entry contains 31961 atoms, of which 0 are hydrogen and 0 are deuterium.

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	1418	Total	C	N	O	S	0	0	0
			11158	7030	1951	2115	62			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1109	Total	C	N	O	S	0	0	0
			8821	5584	1546	1636	55			

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

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

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	179	Total	C	N	O	S	0	0	0
			1443	892	258	291	2			

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	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	136	Total	C	N	O	S	0	0	0
			1092	688	184	215	5			

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

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

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	115	Total	C	N	O	S	0	0	0
			924	593	157	172	2			

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

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	N	7	Total	C	N	O	P	0	0	0
			137	68	22	41	6			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
14	T	19	Total	Br	C	N	O	P	0	0	0
			387	1	185	69	114	18			

- Molecule 15 is a RNA chain called 5'-R(*UP*GP*CP*AP*UP*UP*U*CP*AP*AP*CP*CP

*AP*GP*GP*CP*UP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
15	P	11	Total	C	N	O	P	0	0	0
			232	105	44	73	10			

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

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

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

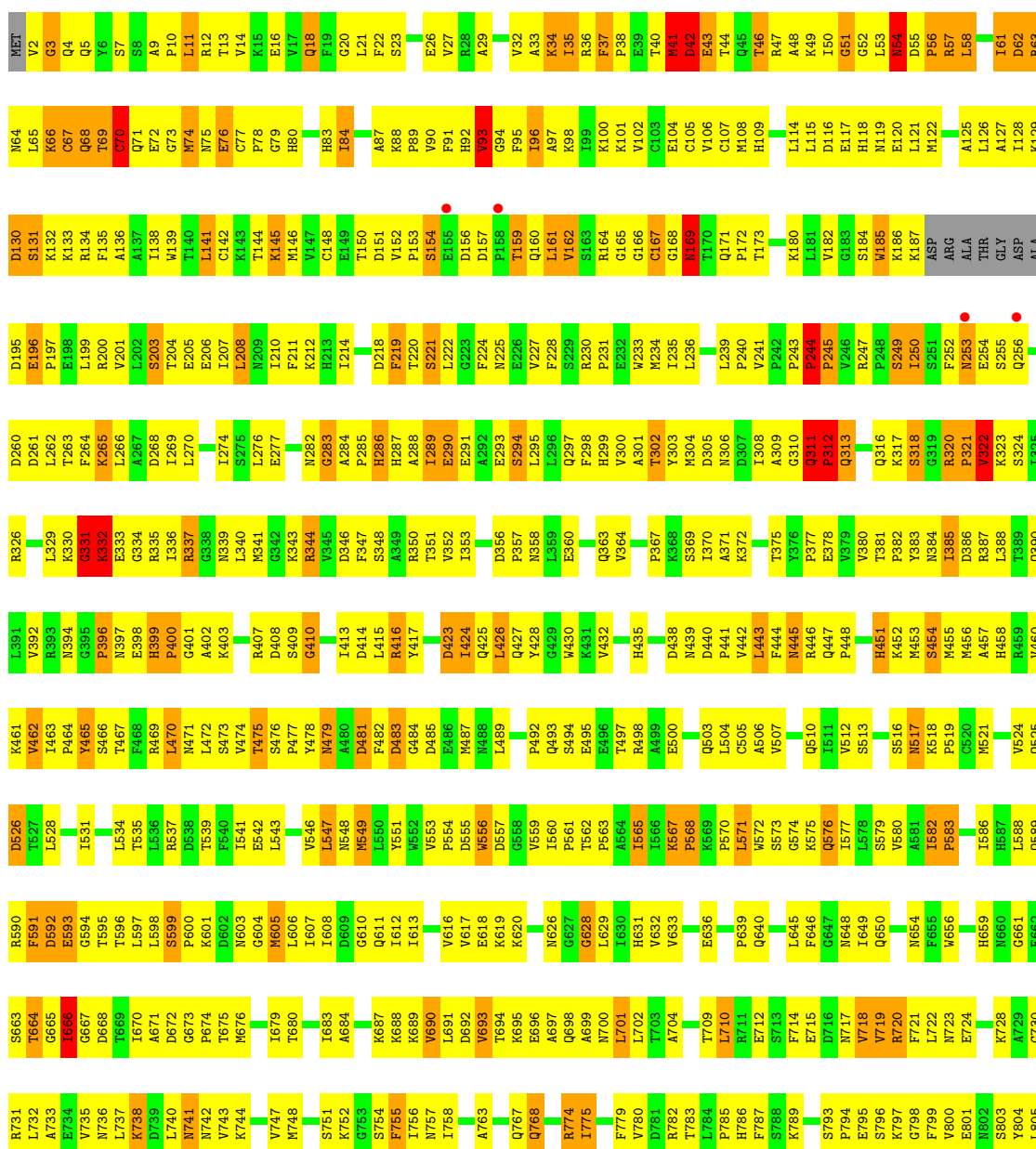
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
17	P	1	Total	Mg	0	0
			1	1		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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:



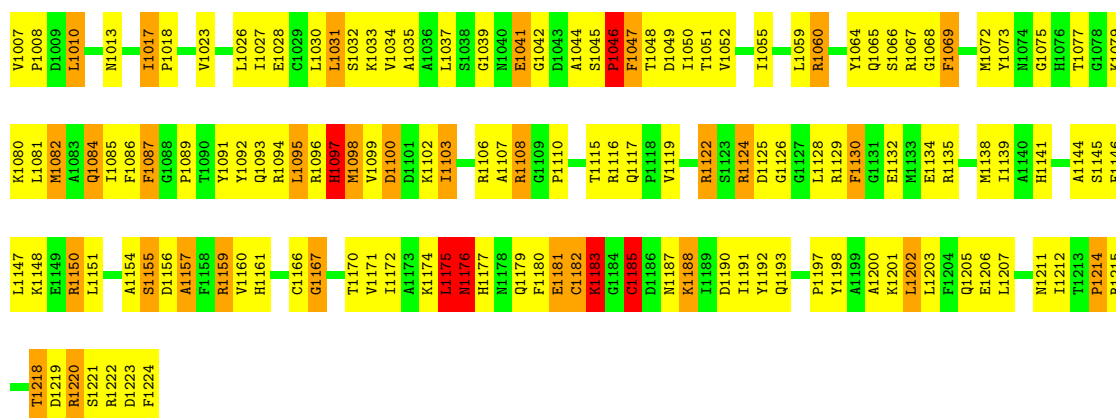
THR	SER	THR	SER	THR	GLY	GLN	GLU	F1389	T1325	E1263	M1202	E1139	T1077	Q1011	R940	Q872	R806
PRO	PRO	GLY	GLN	M1390	R1326	E1264	M1203	H1140	R1326	E1265	M1203	H1140	Q1078	R1012	K941	M873	T809
THR	THR	ALA	ILE	R1391	T1329	T1266	D1204	T1141	T1329	T1266	D1204	T1141	M1079	V1015	F942	A875	P810
THR	THR	THR	THR	M1393	S1331	M1267	D1205	K1144	M1393	M1267	D1205	K1144	T1080	L1081	R944	A876	Q811
PRO	PRO	SER	GLU	T1394	S1331	L1268	D1206	S1145	T1394	L1268	D1206	S1145	L1016	L1017	R945	R877	E812
PRO	PRO	PRO	ILE	G1395	F1332	E1269	T1208	V1146	G1395	E1269	T1208	V1146	F1018	F1019	V946	I878	F813
THR	THR	PHE	GLU	A1396	S1333	M1270	M1209	T1147	A1396	M1270	M1209	T1147	C1020	C1021	F947	S882	F814
THR	THR	GLY	ASP	L1397	D1334	I1271	G1210	I1148	L1397	I1271	G1210	I1148	L1021	L1022	A952	D884	H816
PRO	PRO	ALA	GLN	M1398	M1336	T1272	V1212	S1150	M1398	T1272	V1212	S1150	PHE	N953	A953	N953	H816
THR	THR	THR	GLN	M1399	M1336	T1273	V1212	S1150	M1399	T1273	V1212	S1150	ALA	L1022	N953	N953	A817
THR	THR	GLY	ASP	C1400	E1337	R1274	G1213	E1151	C1400	E1337	R1274	G1213	GLY	R1023	W954	T885	M818
SER	SER	GLY	GLY	S1401	V1338	G1275	E1214	I1152	S1401	V1338	G1275	E1214	VAL	S1024	G819	G819	G819
PRO	PRO	GLY	GLY	F1402	L1339	V1276	R1215	V1153	F1402	L1339	V1276	R1215	ALA	R1025	V958	G857	G820
THR	THR	VAL	VAL	E1403	G1340	E1277	I1216	V1154	E1403	G1340	E1277	I1216	SER	L1026	N959	G888	R821
THR	THR	THR	THR	E1404	L1341	M1278	I1217	D1155	E1404	L1341	M1278	I1217	LYS	L1027	I960	S889	E822
PRO	PRO	PRO	PRO	T1405	E1342	I1279	Q1218	D1155	T1405	E1342	I1279	Q1218	SER	T1028	R961	D890	G823
PRO	PRO	SER	PRO	V1406	A1343	R1280	T1219	P1158	V1406	A1343	R1280	T1219	VAL	R1029	R962	L824	G823
THR	THR	GLY	SER	F1407	G1344	R1281	F1220	P1159	F1407	G1344	R1281	F1220	GLY	R1030	I963	L824	L825
SER	SER	PHE	ASN	I1408	R1345	V1282	K1221	S1160	I1408	R1345	V1282	K1221	ASN	R1031	I964	R826	R826
PRO	PRO	GLY	GLY	L1409	R1345	V1283	K1222	S1160	L1409	R1345	V1283	K1222	GLY	R1032	I965	R896	T827
THR	THR	VAL	SER	F1410	A1347	M1284	D1223	T1161	F1410	A1347	M1284	D1223	VAL	Q1033	N966	R898	A828
SER	SER	SER	GLY	E1411	L1348	M1285	L1224	I1163	E1411	L1348	M1285	L1224	THR	E1034	N967	V829	V829
PRO	PRO	SER	LEU	A1412	Y1349	K1286	F1225	I1163	A1412	Y1349	K1286	F1225	PRO	P1099	A967	V899	V899
ALA	PRO	PRO	VAL	G1413	K1350	V1287	V1226	H1172	G1413	K1350	V1287	V1226	LYS	R1100	Q968	D900	K830
THR	THR	ASN	ASN	A1414	E1351	D1288	D1233	H1173	A1414	E1351	D1288	D1233	GLY	R1036	L901	T831	T831
SER	SER	PHE	ALA	E1417	V1352	D1289	D1233	F1174	E1417	V1352	D1289	D1233	ASN	L1037	L902	A832	A832
PRO	PRO	GLY	ASP	L1418	R1353	K1290	S1229	S1175	L1418	R1353	K1290	S1229	GLY	T1038	N903	T903	T903
THR	THR	PRO	LEU	F1419	M1354	V1291	D1230	LEU	F1419	M1354	V1291	D1230	THR	T1039	N903	T903	T903
SER	SER	THR	THR	D1419	V1355	P1292	D1231	ASP	D1419	V1355	P1292	D1231	THR	Q1040	N905	H906	H906
PRO	PRO	SER	VAL	A1420	I1356	S1293	N1232	GLU	A1420	I1356	S1293	N1232	PRO	A1041	K977	T907	T907
SER	SER	PRO	LYS	C1421	I1356	S1294	N1232	H1172	C1421	I1356	S1294	N1232	THR	F1042	P978	T907	T907
THR	THR	THR	ASP	D1422	D1359	T1295	D1233	H1173	D1422	D1359	T1295	D1233	THR	T1045	S979	L908	R839
PRO	PRO	THR	GLY	G1423	G1296	G1296	L1236	F1174	G1423	G1296	G1296	L1236	PRO	V1046	D980	D909	R840
PRO	PRO	SER	ASP	V1424	V1362	E1297	I1237	LEU	V1424	V1362	E1297	I1237	GLY	S1047	L981	P910	P910
THR	THR	PRO	LEU	S1425	V1363	I1238	I1238	LEU	S1425	V1363	I1238	I1238	THR	K1112	T982	S911	K843
SER	SER	THR	PHE	E1426	N1364	V1299	R1239	ASP	E1426	N1364	V1299	R1239	THR	N1048	I983	L912	A844
PRO	PRO	SER	SER	E1429	V1365	K1300	C1240	GLU	E1429	V1365	K1300	C1240	THR	P1114	K984	L913	L845
SER	SER	PRO	PRO	I1432	R1366	R1241	R1241	GLU	I1432	R1366	R1241	R1241	PRO	S1115	D985	E914	E846
THR	THR	VAL	VAL	Q1432	H1367	E1242	V1242	ALA	Q1432	H1367	E1242	V1242	THR	F1053	I986	S915	S915
PRO	PRO	SER	ASP	L1436	M1368	L1306	R1243	GLN	L1436	M1368	L1306	R1243	PRO	L1054	L988	G916	G916
THR	THR	PRO	ASP	L1437	L1370	E1307	P1245	SER	L1437	L1370	E1307	P1245	THR	R1065	T982	I919	T919
PRO	PRO	GLY	GLY	V1372	L1371	T1308	LYS	PHE	V1372	L1371	T1308	LYS	THR	V1057	L993	G921	G921
THR	THR	SER	SER	D1437	V1372	D1309	SER	D1186	D1437	V1372	D1309	SER	THR	G1061	Q994	D922	D922
PRO	PRO	ASN	ASN	T1438	D1373	V1310	LEU	Q1187	T1438	D1373	V1310	LEU	PRO	E1062	E995	L923	T855
THR	THR	SER	ASP	G1439	V1374	V1311	ASP	Q1188	G1439	V1374	V1311	ASP	THR	K924	N996	K924	T856
SER	SER	ALA	ALA	A1440	M1375	N1312	ALA	S1189	A1440	M1375	N1312	ALA	THR	M1063	L997	L925	R857
PRO	PRO	MET	MET	F1441	T1376	L1313	GLU	P1190	F1441	T1376	L1313	GLU	PRO	V1064	L998	Q926	N858
THR	THR	ALA	ALA	D1442	T1377	S1314	THR	V1191	D1442	T1377	S1314	THR	THR	G1065	V999	V927	S859
SER	SER	GLY	GLY	V1443	Q1378	E1315	GLU	L1192	V1443	Q1378	E1315	GLU	THR	V1066	L1000	L928	L860
PRO	PRO	PHE	PHE	I1445	G1379	V1316	ALA	L1193	I1445	G1379	V1316	ALA	PRO	L1067	R1001	L929	L860
THR	THR	SER	SER	L1381	M1380	M1317	E1254	L1193	L1381	M1380	M1317	E1254	THR	G1002	T982	L929	L860
SER	SER	THR	THR	V1384	V1319	T1318	E1255	L1194	V1384	V1319	T1318	E1255	THR	K1003	G933	Y933	Q865
PRO	PRO	ALA	ALA	V1384	V1319	T1318	E1255	L1194	V1384	V1319	T1318	E1255	PRO	S1071	K934	Y933	Q865
THR	THR	SER	SER	V1384	V1319	T1318	E1255	L1194	V1384	V1319	T1318	E1255	THR	E1005	K935	Q935	T867
GLY	GLY	GLY	GLY	T1385	G1321	G1321	M1259	L1197	T1385	G1321	G1321	M1259	GLY	G1073	I1006	L936	T868
THR	THR	THR	THR	T1386	G1322	I1322	K1260	D1198	T1386	G1322	I1322	K1260	THR	E1074	N1009	V937	G869
ALA	ALA	ALA	ALA	H1387	D1323	K1261	A1200	A1200	H1387	D1323	K1261	A1200	ALA	P1075	K938	E870	E870
PRO	PRO	ASP	ASP	G1388	P1324	K1262	K1262	A1201	G1388	P1324	K1262	K1262	PRO	A1076	A1010	D939	D871

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

Chain B:

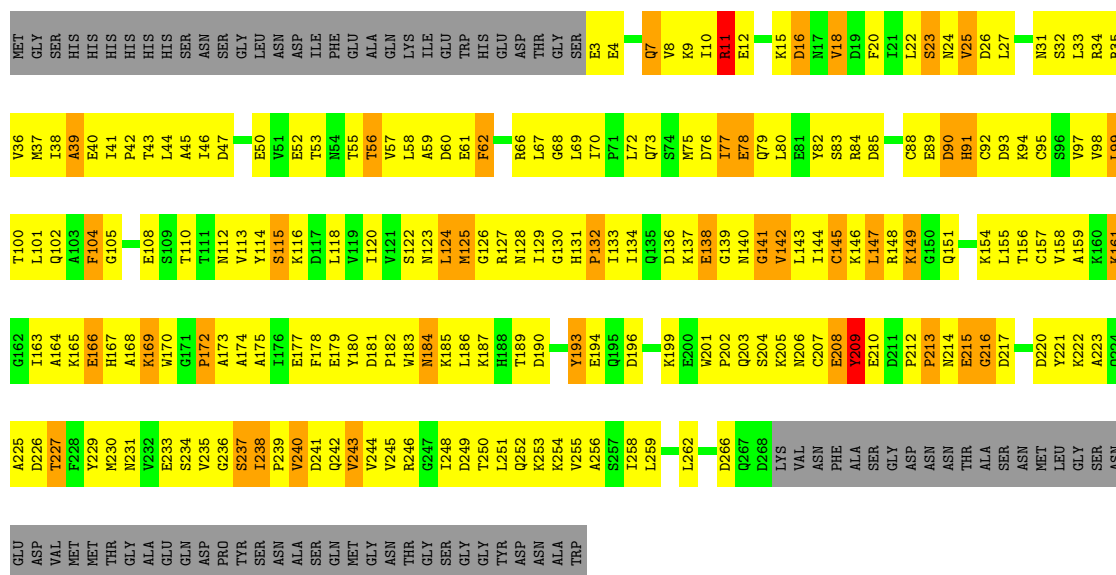






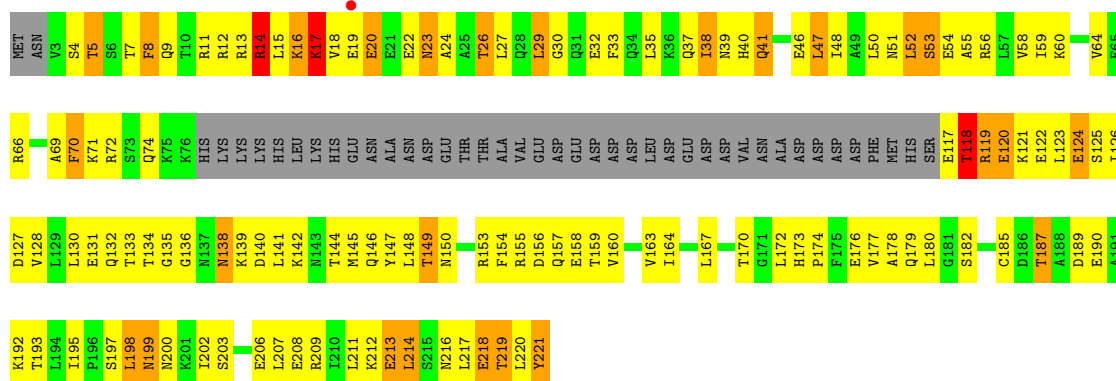
• Molecule 3: DNA-directed RNA polymerase II subunit RPB3

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

Frequency	Percentage
Daily	35%
Weekly	45%
Monthly	20%



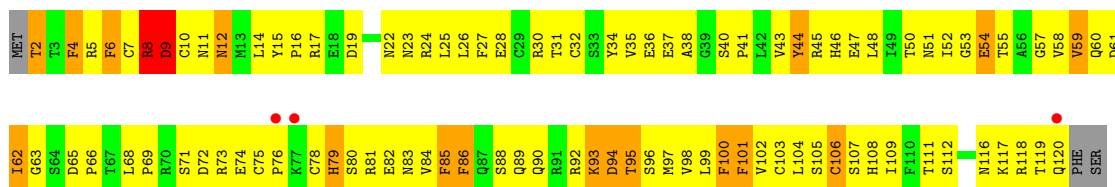
Frequency	Percentage
Daily	25%
Weekly	45%
Monthly	10%
Other	20%



Frequency	Percentage
Daily	~2%
Weekly	~18%
Monthly	~78%
Quarterly	~2%
Other	~0%

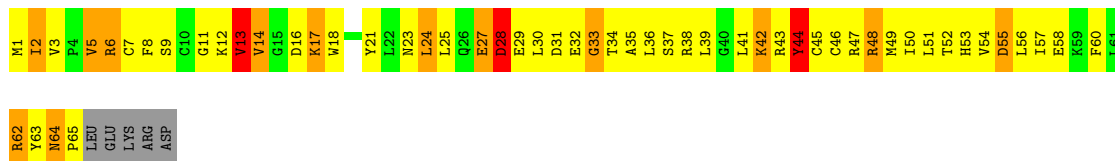


Chain I:



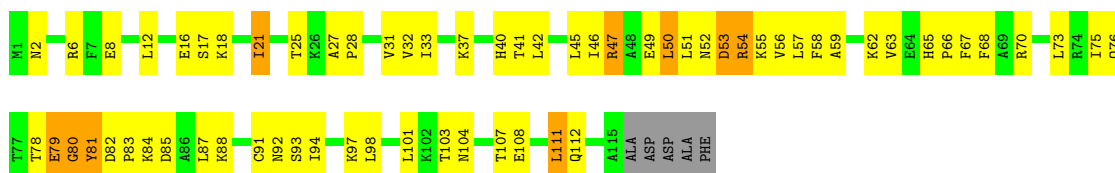
- Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J:



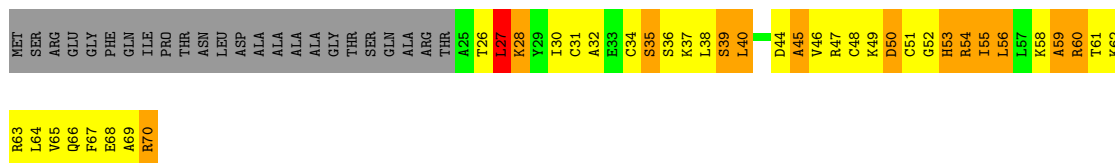
- Molecule 11: DNA-directed RNA polymerase II subunit RPB11

Chain K:



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

Chain L:



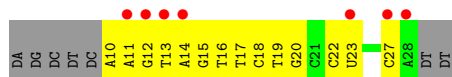
- Molecule 13: 5'-D(*AP*CP*TP*AP*CP*TP*TP*GP*AP*GP*CP*T)-3'

Chain N:



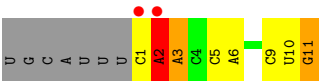
- Molecule 14: 5'-D(*AP*GP*CP*TP*C*AP*AP*GP*TP*AP*GP*TP*TP*CP*TP*GP*CP*C P*(BRU)P*GP*GP*TP*CP*AP*TP*T)-3'

Chain T:



- Molecule 15: 5'-R(*UP*GP*CP*AP*UP*UP*U*CP*AP*AP*CP*CP*AP*GP*GP*CP*UP*G)-3'

Chain P: 



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	221.43Å 393.75Å 281.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.65 49.84 – 3.65	Depositor EDS
% Data completeness (in resolution range)	99.9 (50.00-3.65) 100.0 (49.84-3.65)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.21 (at 3.67Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.210 , 0.253 0.212 , 0.256	Depositor DCC
R_{free} test set	2674 reflections (2.01%)	DCC
Wilson B-factor (Å ²)	88.8	Xtriage
Anisotropy	0.057	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 75.8	EDS
Estimated twinning fraction	0.017 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.025 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
L-test for twinning	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 135971 reflections	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	31961	wwPDB-VP
Average B, all atoms (Å ²)	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.01% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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	0.51	0/11358	0.79	4/15360 (0.0%)
2	B	0.49	0/8991	0.74	4/12121 (0.0%)
3	C	0.50	0/2133	0.74	1/2891 (0.0%)
4	D	0.48	0/1453	0.77	1/1947 (0.1%)
5	E	0.48	0/1788	0.71	2/2406 (0.1%)
6	F	0.57	0/717	0.83	1/967 (0.1%)
7	G	0.54	0/1368	0.81	1/1844 (0.1%)
8	H	0.45	0/1110	0.74	0/1502
9	I	0.44	0/989	0.72	0/1331
10	J	0.51	0/541	0.85	1/727 (0.1%)
11	K	0.49	0/942	0.68	0/1272
12	L	0.56	0/365	0.82	0/485
13	N	0.60	0/152	0.90	0/232
14	T	0.58	0/410	0.82	0/629
15	P	0.57	0/259	0.82	1/402 (0.2%)
All	All	0.50	0/32576	0.76	16/44116 (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
10	J	0	1

There are no bond length outliers.

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	26	THR	N-CA-C	-6.48	93.51	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	39	ALA	N-CA-C	6.32	128.07	111.00
1	A	331	GLY	N-CA-C	5.96	128.00	113.10
7	G	65	ASP	N-CA-C	-5.92	95.02	111.00
1	A	3	GLY	N-CA-C	-5.78	98.65	113.10

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
10	J	44	TYR	Sidechain

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	11158	0	11228	1381	0
2	B	8821	0	8850	1234	0
3	C	2095	0	2051	306	0
4	D	1443	0	1466	213	0
5	E	1752	0	1776	214	0
6	F	705	0	731	92	0
7	G	1340	0	1357	168	0
8	H	1092	0	1069	179	0
9	I	971	0	929	137	0
10	J	532	0	542	112	0
11	K	924	0	934	105	0
12	L	363	0	388	83	0
13	N	137	0	82	4	0
14	T	387	0	214	25	0
15	P	232	0	122	14	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	P	1	0	0	0	0
All	All	31961	0	31739	3920	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 62.

The worst 5 of 3920 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:B:508:LEU:HD13	2:B:510:LYS:HE2	1.26	1.16
1:A:53:LEU:HD23	1:A:54:ASN:N	1.61	1.16
2:B:744:HIS:HD2	2:B:745:PRO:HD2	1.07	1.14
1:A:1208:THR:HB	1:A:1211:GLN:HG3	1.24	1.13
2:B:559:SER:HA	2:B:563:MET:HB3	1.15	1.13

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

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	1408/1733 (81%)	1012 (72%)	262 (19%)	134 (10%)	1	20
2	B	1089/1224 (89%)	779 (72%)	201 (18%)	109 (10%)	1	17
3	C	264/347 (76%)	186 (70%)	51 (19%)	27 (10%)	1	17
4	D	175/221 (79%)	121 (69%)	39 (22%)	15 (9%)	1	23
5	E	212/215 (99%)	154 (73%)	42 (20%)	16 (8%)	2	28
6	F	85/155 (55%)	69 (81%)	14 (16%)	2 (2%)	9	62
7	G	169/171 (99%)	145 (86%)	13 (8%)	11 (6%)	2	33
8	H	132/146 (90%)	85 (64%)	23 (17%)	24 (18%)	0	4
9	I	117/122 (96%)	79 (68%)	29 (25%)	9 (8%)	1	26
10	J	63/70 (90%)	39 (62%)	11 (18%)	13 (21%)	0	3

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
11	K	113/120 (94%)	87 (77%)	22 (20%)	4 (4%)	6	53
12	L	44/70 (63%)	23 (52%)	9 (20%)	12 (27%)	0	1
All	All	3871/4594 (84%)	2779 (72%)	716 (18%)	376 (10%)	1	19

5 of 376 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	4	GLN
1	A	41	MET
1	A	43	GLU
1	A	48	ALA
1	A	54	ASN

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	1241/1520 (82%)	1109 (89%)	132 (11%)	10	46
2	B	963/1061 (91%)	841 (87%)	122 (13%)	6	35
3	C	234/299 (78%)	206 (88%)	28 (12%)	7	38
4	D	161/200 (80%)	139 (86%)	22 (14%)	5	31
5	E	196/197 (100%)	180 (92%)	16 (8%)	17	62
6	F	77/137 (56%)	73 (95%)	4 (5%)	32	79
7	G	152/152 (100%)	137 (90%)	15 (10%)	11	50
8	H	120/128 (94%)	104 (87%)	16 (13%)	6	33
9	I	113/116 (97%)	97 (86%)	16 (14%)	5	30
10	J	60/65 (92%)	55 (92%)	5 (8%)	16	61
11	K	99/102 (97%)	92 (93%)	7 (7%)	21	69
12	L	40/57 (70%)	36 (90%)	4 (10%)	11	50
All	All	3456/4034 (86%)	3069 (89%)	387 (11%)	9	42

5 of 387 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	493	SER
2	B	909	ASP
9	I	6	PHE
2	B	557	PHE
2	B	737	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 106 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1432	GLN
2	B	657	HIS
7	G	131	GLN
2	B	60	GLN
2	B	366	GLN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
15	P	10/18 (55%)	2 (20%)	1 (10%)

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
15	P	3	A
15	P	11	G

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
15	P	2	A

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

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is

the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
14	BRU	T	23	15,14	19,21,22	3.79	3 (15%)	22,30,33	1.78	3 (13%)

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	BRU	T	23	15,14	-	0/5/21/22	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
14	T	23	BRU	BR-C5	-15.72	1.50	1.90
14	T	23	BRU	C4-C5	3.22	1.45	1.39
14	T	23	BRU	P-OP1	2.66	1.49	1.46

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
14	T	23	BRU	C5-C6-N1	6.08	123.67	119.67
14	T	23	BRU	C6-N1-C2	-4.38	121.16	122.41
14	T	23	BRU	BR-C5-C6	2.54	123.02	117.98

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues

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, 95th 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	1418/1733 (81%)	0.02	7 (0%) 88 72	23, 70, 112, 141	0
2	B	1109/1224 (90%)	0.05	4 (0%) 90 77	23, 81, 123, 141	0
3	C	266/347 (76%)	0.01	0 100 100	34, 69, 105, 119	0
4	D	179/221 (80%)	0.12	1 (0%) 86 68	37, 79, 118, 131	0
5	E	214/215 (99%)	0.21	0 100 100	41, 97, 129, 137	0
6	F	87/155 (56%)	-0.20	0 100 100	19, 46, 77, 86	0
7	G	171/171 (100%)	0.06	0 100 100	48, 64, 104, 113	0
8	H	136/146 (93%)	0.48	1 (0%) 84 65	80, 106, 127, 135	0
9	I	119/122 (97%)	0.32	3 (2%) 54 33	65, 100, 125, 143	0
10	J	65/70 (92%)	-0.15	0 100 100	49, 65, 92, 105	0
11	K	115/120 (95%)	0.03	0 100 100	34, 73, 93, 122	0
12	L	46/70 (65%)	0.23	0 100 100	48, 108, 125, 132	0
13	N	7/12 (58%)	1.67	2 (28%) 1 1	135, 140, 151, 157	0
14	T	19/26 (73%)	1.66	7 (36%) 1 1	117, 142, 155, 155	0
15	P	11/18 (61%)	1.36	2 (18%) 2 2	125, 133, 152, 156	0
All	All	3962/4650 (85%)	0.08	27 (0%) 84 65	19, 76, 122, 157	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
14	T	28	DA	4.7
15	P	1	C	4.1
14	T	13	DT	3.4
2	B	868	MET	2.9
1	A	155	GLU	2.8

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	RSR	LLDF	B-factors(Å ²)	Q<0.9
14	BRU	T	23	20/21	0.34	2.74	136,142,145,146	0

6.3 Carbohydrates ⓘ

There are no carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	RSR	LLDF	B-factors(Å ²)	Q<0.9
16	ZN	J	1066	1/1	0.20	0.39	53,53,53,53	0
16	ZN	B	2225	1/1	0.16	-0.73	30,30,30,30	0
17	MG	P	2458	1/1	0.24	-0.73	180,180,180,180	0
16	ZN	C	1269	1/1	0.10	-1.41	40,40,40,40	0
16	ZN	A	2457	1/1	0.12	-1.60	31,31,31,31	0
16	ZN	L	1071	1/1	0.05	-1.82	100,100,100,100	0
16	ZN	I	1121	1/1	0.08	-2.18	76,76,76,76	0
16	ZN	I	1122	1/1	0.12	-2.37	135,135,135,135	0
16	ZN	A	2456	1/1	0.06	-3.91	88,88,88,88	0

6.5 Other polymers ⓘ

There are no such residues in this entry.