



wwPDB X-ray Structure Validation Summary Report ⓘ

May 18, 2020 – 09:58 am BST

PDB ID : 2O5J
Title : Crystal structure of the T. thermophilus RNAP polymerase elongation complex with the NTP substrate analog
Authors : Vassilyev, D.G.; Vassilyeva, M.N.
Deposited on : 2006-12-06
Resolution : 3.00 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
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) : 2.11

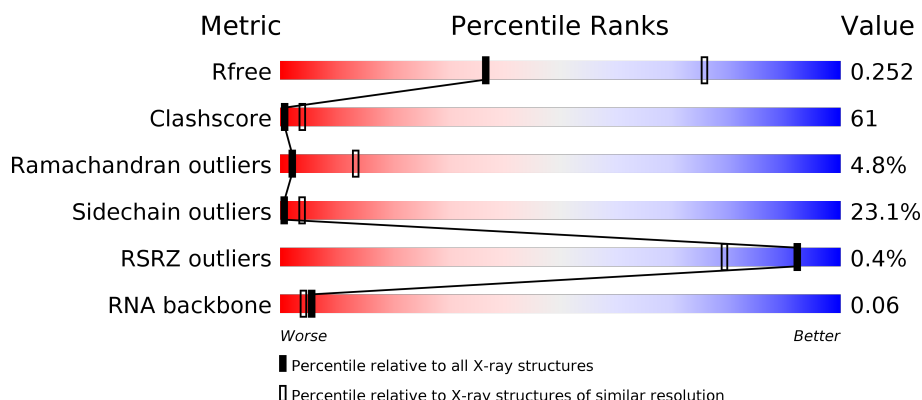
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)
RNA backbone	3102	1173 (3.30-2.70)

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 respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	G	23	<div> <div>17%</div> <div>57%</div> <div>22%</div> <div>.</div> </div>
1	X	23	<div> <div>17%</div> <div>61%</div> <div>13%</div> <div>9%</div> </div>
2	H	16	<div> <div>19%</div> <div>81%</div> </div>
2	Y	16	<div> <div>31%</div> <div>69%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	I	14	<div><div></div><div>21%64%7%7%</div></div>
3	Z	14	<div><div></div><div>29%64%7%</div></div>
4	A	315	<div><div></div><div>21%42%10%27%</div></div>
4	B	315	<div><div></div><div>%22%40%11%27%</div></div>
4	K	315	<div><div></div><div>23%42%8%27%</div></div>
4	L	315	<div><div></div><div>18%42%13%27%</div></div>
5	C	1119	<div><div></div><div>23%57%19%.</div></div>
5	M	1119	<div><div></div><div>%23%58%18%. .</div></div>
6	D	1524	<div><div></div><div>21%47%14%.17%</div></div>
6	N	1524	<div><div></div><div>22%47%14%.17%</div></div>
7	E	99	<div><div></div><div>22%49%23%. .</div></div>
7	O	99	<div><div></div><div>20%59%16%. .</div></div>

2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 51213 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 DNA chain called 5'-D(P*CP*CP*CP*TP*GP*TP*CP*TP*GP*GP*CP*GP*TP*TP*CP*GP*CP*GP*CP*GP*CP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	G	23	Total	C	N	O	P	0	0	0
			467	220	80	144	23			
1	X	23	Total	C	N	O	P	0	0	0
			467	220	80	144	23			

- Molecule 2 is a RNA chain called 5'-R(P*GP*AP*GP*UP*CP*UP*GP*CP*GP*GP*CP*GP*CP*GP*CP*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	16	Total	C	N	O	P	0	0	0
			347	153	64	114	16			
2	Y	16	Total	C	N	O	P	0	0	0
			347	153	64	114	16			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	I	13	Total	C	N	O	P	0	0	0
			270	126	57	74	13			
3	Z	13	Total	C	N	O	P	0	0	0
			270	126	57	74	13			

- Molecule 4 is a protein called DNA-directed RNA polymerase alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	229	Total	C	N	O	S	0	0	0
			1806	1153	313	337	3			
4	B	229	Total	C	N	O	S	0	0	0
			1806	1153	313	337	3			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	K	229	Total	C	N	O	S	0	0	0
			1806	1153	313	337	3			
4	L	229	Total	C	N	O	S	0	0	0
			1806	1153	313	337	3			

- Molecule 5 is a protein called DNA-directed RNA polymerase beta chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	C	1119	Total	C	N	O	S	0	0	0
			8829	5581	1577	1647	24			
5	M	1119	Total	C	N	O	S	0	0	0
			8829	5581	1577	1647	24			

- Molecule 6 is a protein called DNA-directed RNA polymerase beta' chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	D	1264	Total	C	N	O	S	0	0	0
			9960	6302	1773	1852	33			
6	N	1264	Total	C	N	O	S	0	0	0
			9960	6302	1773	1852	33			

- Molecule 7 is a protein called DNA-directed RNA polymerase omega chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	E	95	Total	C	N	O	S	0	0	0
			770	491	133	142	4			
7	O	95	Total	C	N	O	S	0	0	0
			770	491	133	142	4			

- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	D	2	Total	Zn	0	0
			2	2		
8	N	2	Total	Zn	0	0
			2	2		

- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

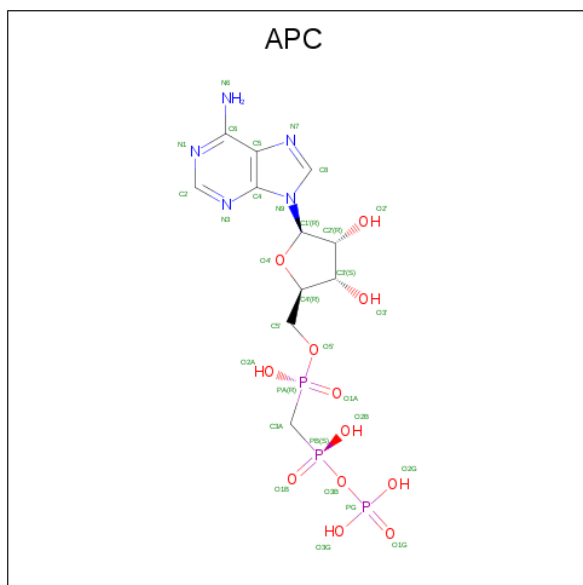
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	2	Total	Mg	0	0
			2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	N	2	Total	Mg	0	0
			2	2		

- Molecule 10 is DIPHOSPHOMETHYLPHOSPHONIC ACID ADENOSYL ESTER (three-letter code: APC) (formula: $C_{11}H_{18}N_5O_{12}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
10	D	1	Total	C	N	O	P	0	0
			31	11	5	12	3		
10	N	1	Total	C	N	O	P	0	0
			31	11	5	12	3		

- Molecule 11 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	G	32	Total	O	0	0
			32	32		
11	H	37	Total	O	0	0
			37	37		
11	I	22	Total	O	0	0
			22	22		
11	X	43	Total	O	0	0
			43	43		
11	Y	30	Total	O	0	0
			30	30		
11	Z	30	Total	O	0	0
			30	30		

Continued on next page...

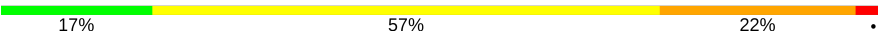
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	A	106	Total 106	O 106	0	0
11	B	82	Total 82	O 82	0	0
11	C	482	Total 482	O 482	0	0
11	D	506	Total 506	O 506	0	0
11	E	60	Total 60	O 60	0	0
11	K	86	Total 86	O 86	0	0
11	L	104	Total 104	O 104	0	0
11	M	483	Total 483	O 483	0	0
11	N	491	Total 491	O 491	0	0
11	O	39	Total 39	O 39	0	0

3 Residue-property plots [i](#)

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

- Molecule 1: 5'-D(P*CP*CP*CP*TP*GP*TP*CP*TP*GP*GP*CP*GP*TP*TP*CP*GP*CP*GP*CP*GP*CP*CP*G)-3'

Chain G: 



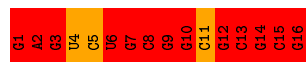
- Molecule 1: 5'-D(P*CP*CP*CP*TP*GP*TP*CP*TP*GP*GP*CP*GP*TP*TP*CP*GP*CP*GP*CP*GP*CP*CP*G)-3'

Chain X: 



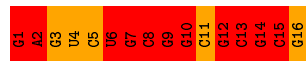
- Molecule 2: 5'-R(P*GP*AP*GP*UP*CP*UP*GP*CP*GP*GP*CP*GP*CP*GP*CP*G)-3',

Chain H: 

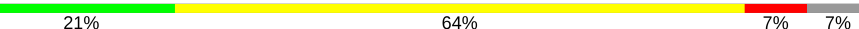


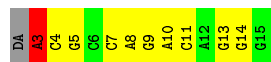
- Molecule 2: 5'-R(P*GP*AP*GP*UP*CP*UP*GP*CP*GP*GP*CP*GP*CP*GP*CP*G)-3',

Chain Y: 



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

Chain I: 

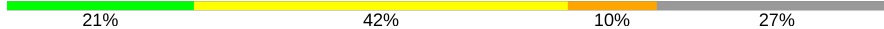


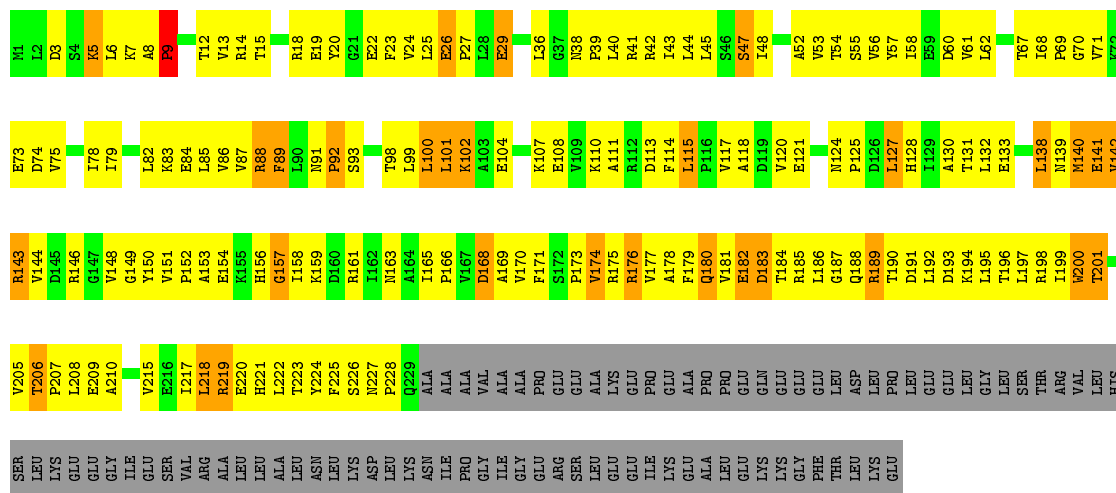
- Molecule 3: 5'-D(*AP*AP*CP*GP*CP*CP*AP*GP*AP*CP*AP*GP*GP*G)-3'

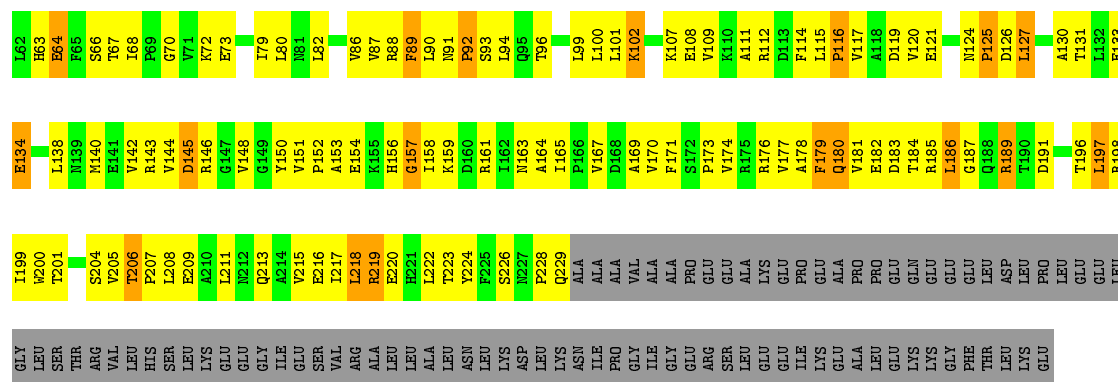
Chain Z: 



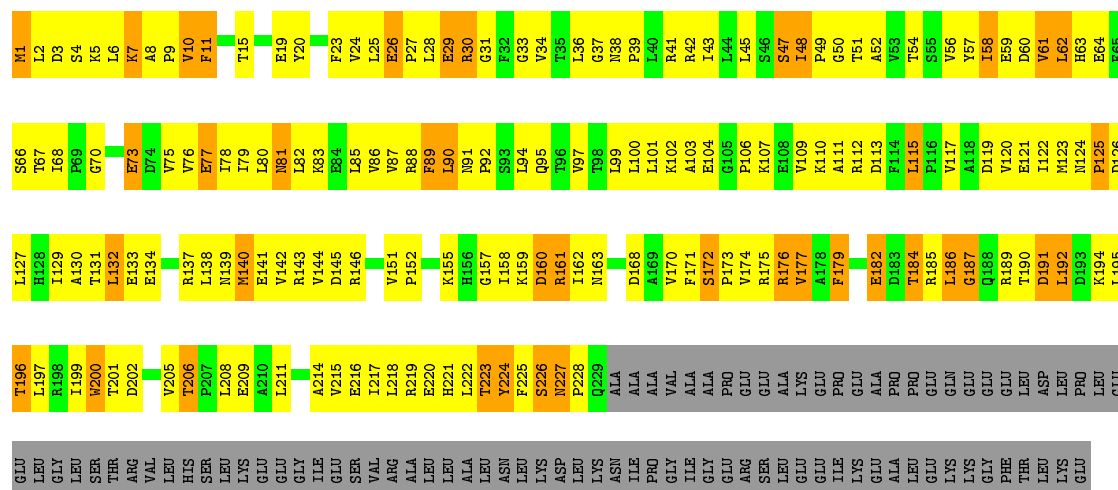
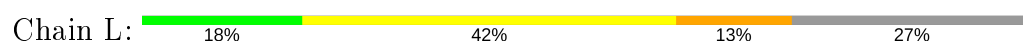
• Molecule 4: DNA-directed RNA polymerase alpha chain

Chain A: 

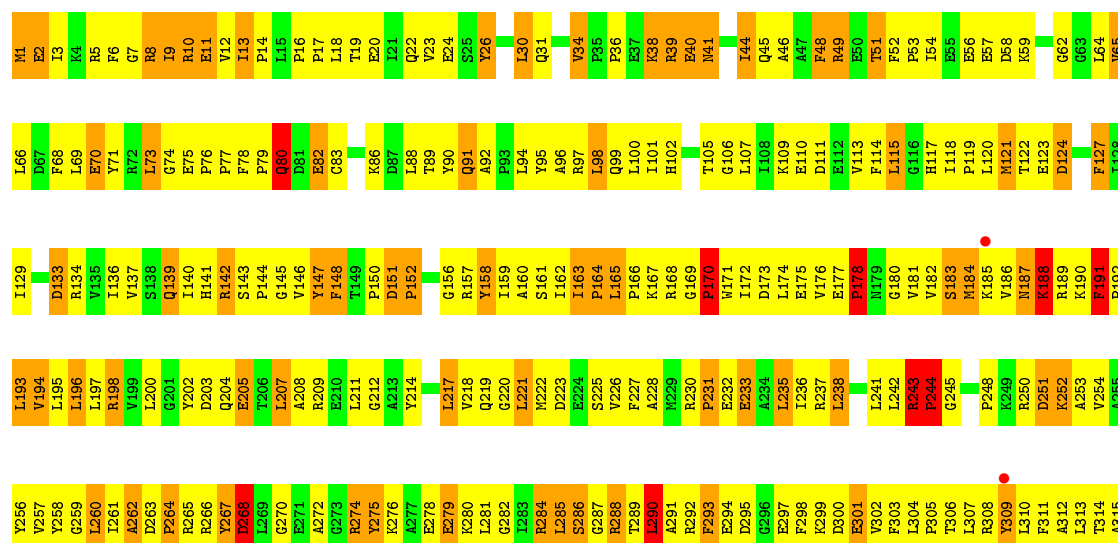


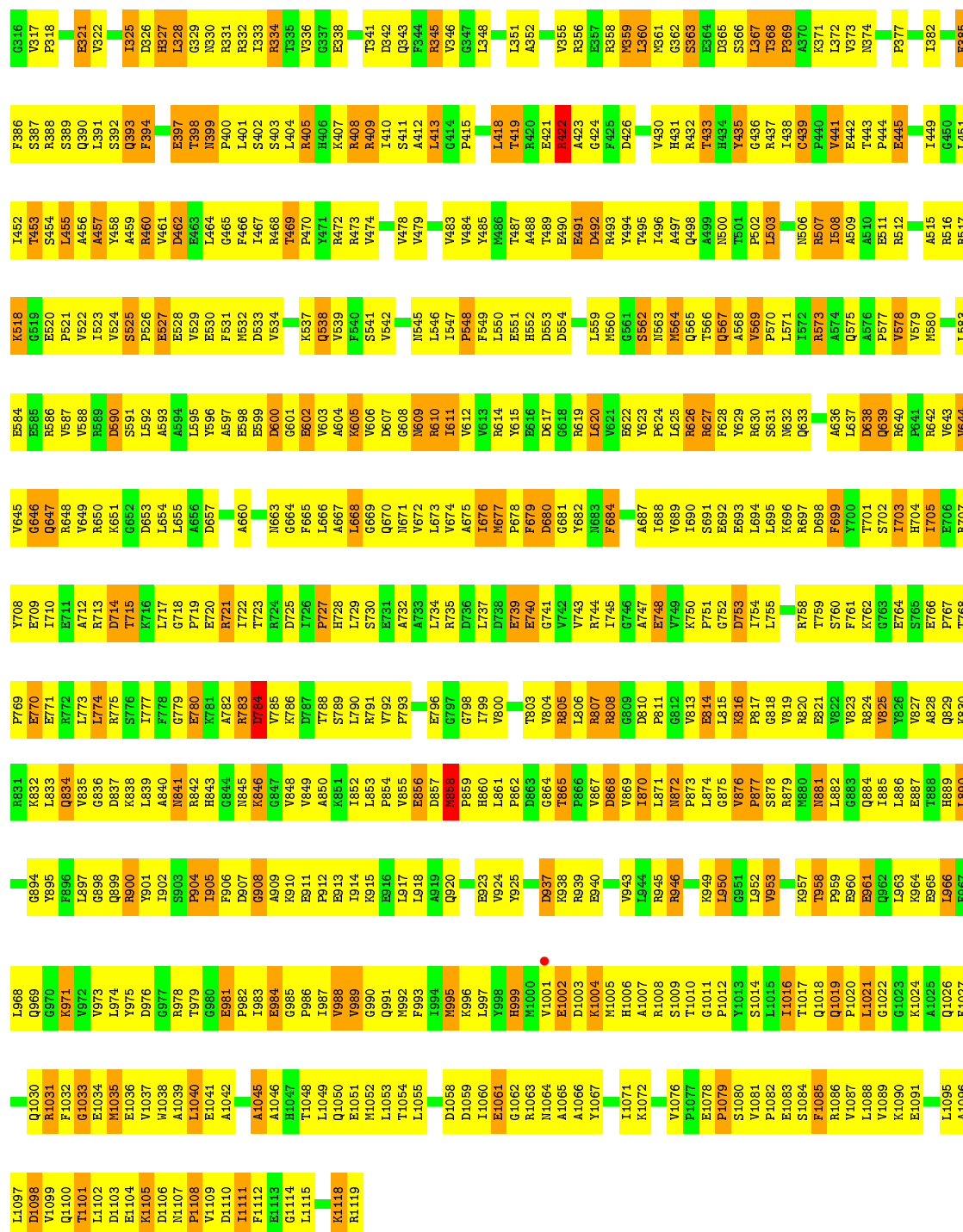


• Molecule 4: DNA-directed RNA polymerase alpha chain

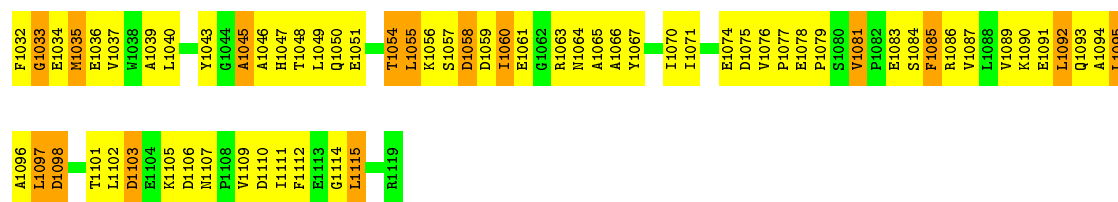


• Molecule 5: DNA-directed RNA polymerase beta chain



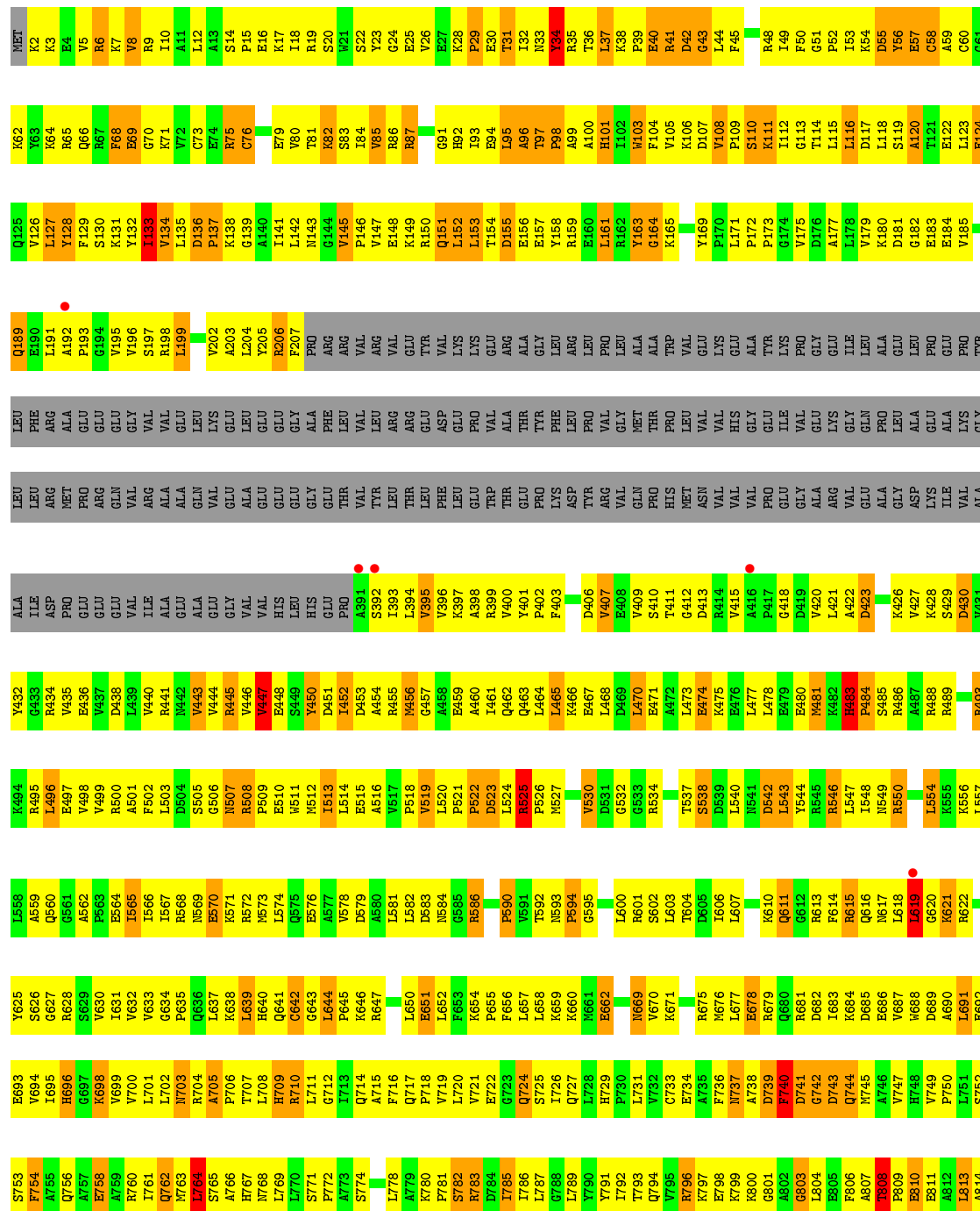






• Molecule 6: DNA-directed RNA polymerase beta' chain

Chain D: 21% 47% 14% 17%





L127	K62	MET
L128	V63	K2
F129	K64	K3
S130	R65	E4
K131	O66	V5
Y132	K67	R6
L133	F68	K7
V134	E69	R9
L135		R9
D136	G73	I10
P137	E74	A11
K138	R75	L12
G139	C76	A13
A140	G77	S14
I141	V78	P15
L142	E79	E16
M143	R80	K17
G144	T81	I18
V145		R19
P146	T84	S20
V147	V85	W21
E148	R86	S22
K149	R87	Y23
R150	T88	G24
K151	R89	E25
L152	N90	E27
L153	G91	K28
T154	H92	P29
D155	I93	E30
E156	E94	T31
E157	L95	A32
V158	A96	N33
R159	T97	V34
E160	P98	R35
L161		R36
L162	H101	L37
V163	I102	K38
G164	W103	P39
K165	F104	E40
O166	V105	R41
E167	K106	D42
T168	D107	G43
V169	V108	L44
P170	P109	P45
L171	S110	
P172	K111	D46
	I112	E47
		R48
V175		L49
D176	L115	F50
L177	L116	G51
L178	D117	P52
V179	L118	I53
	S119	K54
K180	V119	D55
D181	A120	E56
G182	T121	Y57
E183	I122	E58
V184	L123	
E185	E124	C58
	Q125	
L186	V126	G51

Y1145	G1081	L1020	D952	I1885	R818	A755	E693	S629	I566	L503	V443	GLU	PRO	GLU	GLU
G1146	A1082	Y1021	D953	V886	G819	Q756	V694	V630	I567	G506	V444	GLU	ARG	GLU	GLU
R1147	D1083	Y1022	A954	A987	V821	A757	I695	I631	R568	R507	R445	GLU	GLN	GLU	GLU
V1148	T1084	M1023	V955	E888	V821	E758	H696	V632	I569	I507	V446	VAL	VAL	GLY	GLY
L1149	A1085	A1025	I956	A889	L423	A759	A899	V633	R508	R508	V447	VAL	ARG	VAL	VAL
A1150	L1086	Q1024	P957	V890	L423	R760	A822	G634	K571	P509	E448	ALA	ALA	VAL	R198
R1151	R1087	S1026	R958	E933	R824	I761	V700	P835	R572	E510	S449	GLU	ALA	GLU	L199
E1152	T1088	G1027	E959	E893	A825	Q762	L701	Q636	H573	N511	R450	LEU	GLN	LEU	
V1153	A1089	A1028	R960	R994	P826	Q763	L702	L637	H574	N512	R451	LYS	VAL	LYS	V202
E1154	D1090	R1029	K961	V895	I827	L764	N703	V638	Q575	I513	R452	GLY	GLU	GLU	A203
V1155	S1091	G1030	Q962	A896	K828	S765	R704	L639	E576	L514	R453	LEU	ALA	LEU	L204
L1156	G1092	M1031	V963	H997	V829		A705	H840	A577	E515	R454	VAL	GLU	GLU	R205
G1157	Y1083	P1032	L964	E993	A830	N768	P706	G641	H578	A516	R455	HIS	GLU	GLU	R206
V1158	L1094	Q1033	E965	L899	G831	L769	T707	G642	H579	N517	R456	LEU	GLU	GLY	F207
R1159	T1095	Q1034	R966	I900	R832	L770	L708	R832	A580	P518	G457	HIS	GLY	ALA	PRO
L1160	R1096	I1035	A967	Q901	E833	S771	H709	L644	I581	V519	A458	GLU	GLU	PHE	ARG
E1161	K1097	T1036	D968	L902	T834	P772	R710	P645	L582	L520	E459	THR	THR	LEU	ARG
R1162	L1098	Q1037	R969	D903	S835	A773	L711	K646	D583	P521	A460	VAL	VAL	VAL	VAL
G1163	V1099	L1038	K970	V904	V836	S774	G712	R647	H584	P522	I461	LEU	TYR	ARG	ARG
R1164	D1100	C1039	L971	P905	G837	G776	I713	V648	G585	D523	Q482	LEU	TYR	ARG	ARG
V1165	V1101	Q1040	L972	Q906	R838	E776	Q714	V648	R586	I524	Q483	LEU	THR	ARG	GLU
L1166	T1102	R1041	Q973	E907	L839	P777	A715	L652	R587	P526	L464	LEU	THR	ARG	GLU
S1167		R1042	I974	K908	K840	L778	F716	L652	R588	P526	L465	ASP	PHE	GLY	TYR
M1168		G1043	E975		V841	A779	Q717	V653	A589	N527	R466	LEU	GLY	LYS	LYS
D1169	V1106	L1044	Q976	I911		K780	P718	K654	P590	V528	R467	PRO	GLY	LYS	LYS
V1170	V1107	M1045	A977	K912	A844	P781	V719	P655		N529	L488	VAL	TRP	VAL	GLU
V1171	I1108	Q1046		D913	R845	S782	L720	P656	P594	V530	D469	THR	THR	ALA	ARG
	E1109	K1047	P982	L914	P846	R783	V721	L657	G595	D531	L470	GLU	GLU	ALA	ALA
L1174	A1110	P1048	L983	V915	D847	D784	E722	K658	S596	G532	R472	PRO	PRO	THR	THR
	D1111	S1049	T984	V916	E848	T785	G723	R659	D597	G533	A471	LYS	LYS	PHE	PHE
A1177	G1112	G1050	D895	Q917	E848	I786	Q724	K660	R598	B534	L473	LEU	ASP	LEU	ARG
A1178	L1113	E1051	R896	A918	L851	L787	S725	V661	P999		E474	PRO	TYR	LEU	LEU
E1179	T1114	I1052	E987	F919	A852	G788	I726	L600		S538	R476	VAL	ARG	PRO	PRO
	M1116	F1053	R988	L920	V853	L789	Q727	R601	S602	D539	L477	GLY	VAL	LEU	LEU
Q1184	N1117	E1054	R990	R921		Y790	L728	A667	L603	L540	L478	MET	GLN	GLY	ALA
E1185	V1117	V1055	G991	G923	I857	V791	H729	P668	T604	N541	E479	THR	PRO	THR	ALA
V1186	S1118	P1056	I901	G923	V858	I792	P730	N669	R605	D542	E480	MET	HIS	PRO	TRP
P1187	S1119	V1057	I992	N924	D859	T793	L731	R670	L606	L543	R481	VAL	ASN	LEU	VAL
V1188	P1120	R1058	L993	E925	L860	Q794	V732	H671	L606	R544	R482	VAL	VAL	VAL	GLU
R1189	L1122	S1059	Q994	K926	Q861	R796	C733	A672		R545	R483	VAL	VAL	HIS	GLY
F1123	F1061	F1060	L995		D862	K797	E735	A673	K610	R546	P484	VAL	VAL	GLY	ALA
Q1124	R1062	E1063	V996	L930	V864	E798	F736	R675	Q611	L547	S485	PRO	PRO	GLY	TYR
P1125	E1063		E998			K799	N737	R676	R612	I548	R486	GLY	GLY	ILE	LYS
	T1128	G1064	T999	L934	V868	K900	A738	L677	R613	N549	A487	VAL	GLY	VAL	PRO
Q1195	V1129	L1065	E1000	K935			D739	B678	R614	R550	R488	ALA	ALA	GLY	GLY
T1196	R1130	T1066	T1001	Y936	K871	G803	F740	B679	R615	N551	R489	ARG	ARG	LYS	LYS
R1197		V1067	K1002	Y937	R872	L804	D741	Q680	Q616	N552	A490	VAL	VAL	GLY	ILE
Y1198	L1134	L1068	V1003	G938	L873	E805	G742	R681	N617	B553	K491	GLY	GLY	GLN	LEU
G1199	R1135	E1069		F939	G874	F806	D743	D882	L618	L564	R492	ALA	ALA	PRO	ALA
V1200	K1136	Y1070	V1007	T943	T875	A807	Q744	L683	L619		R493	GLY	GLY	LEU	GLU
C1201	R1137	F1071	F1008	T944	S876	T808	H745	R684	G620	I557	R494	ASP	ASP	ALA	LEU
Q1202	A1138	S1073	K1009	T944	P877	P909	A746	D685	R621	L558	R495	LYS	LYS	GLU	PRO
K1203	L1139	S1074	N1010	S945	G878	E810	V747	E586	R622	A559	R496	ILE	ILE	ALA	GLU
C1204	I1140	H1075	F1011	G946	R879	R679	H748	V687	V623	O560	E497	VAL	VAL	LYS	PRO
Y1205	E1141		Y1015	I947	T880	L813	V749	R688	D624	G561	V498	ALA	ALA	GLY	TYR
G1206	A1142	R1078	P1016	T948	L881	A814	P750	P689	V625	R562	V499	ILE	ALA	LEU	LEU
V1207	G1143	K1079	F1017	I949	P882	A815	L751	A690	S626	P563	R500	ARG	ASP	PHE	ARG
D1208	L1144	G1080		G950	R884	H616	F754	L691	R627	E564	R441	PRO	PRO	ALA	ALA



4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α , β , γ	152.34Å 152.34Å 524.57Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 3.00 38.29 – 3.00	Depositor EDS
% Data completeness (in resolution range)	89.3 (40.00-3.00) 83.5 (38.29-3.00)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.51 (at 3.01Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.225 , 0.257 0.227 , 0.252	Depositor DCC
R_{free} test set	11219 reflections (5.70%)	wwPDB-VP
Wilson B-factor (Å ²)	64.2	Xtriage
Anisotropy	0.049	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 136.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.40$, $\langle L^2 \rangle = 0.23$	Xtriage
Estimated twinning fraction	0.146 for h,-k,-l	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	51213	wwPDB-VP
Average B, all atoms (Å ²)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.44% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: APC, ZN, MG

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	G	1.04	1/520 (0.2%)	1.12	1/798 (0.1%)
1	X	1.12	2/520 (0.4%)	1.14	1/798 (0.1%)
2	H	1.39	3/387 (0.8%)	2.45	39/601 (6.5%)
2	Y	1.36	3/387 (0.8%)	2.44	33/601 (5.5%)
3	I	0.72	0/304	0.92	1/467 (0.2%)
3	Z	0.73	0/304	0.91	0/467
4	A	0.69	0/1838	0.76	0/2498
4	B	0.76	0/1838	0.76	2/2498 (0.1%)
4	K	0.73	0/1838	0.82	3/2498 (0.1%)
4	L	0.73	0/1838	0.78	4/2498 (0.2%)
5	C	0.79	1/8997 (0.0%)	0.89	17/12164 (0.1%)
5	M	0.78	1/8997 (0.0%)	0.90	17/12164 (0.1%)
6	D	0.79	1/10128 (0.0%)	0.91	18/13681 (0.1%)
6	N	0.79	2/10128 (0.0%)	0.89	22/13681 (0.2%)
7	E	0.83	1/784 (0.1%)	1.07	3/1057 (0.3%)
7	O	0.78	1/784 (0.1%)	1.07	3/1057 (0.3%)
All	All	0.80	16/49592 (0.0%)	0.95	164/67528 (0.2%)

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	G	0	6
1	X	0	5
2	H	0	2
2	Y	0	1
3	I	0	1
All	All	0	15

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	1	G	OP3-P	10.56	1.73	1.61
1	G	1	DC	OP3-P	-7.74	1.51	1.61
1	X	1	DC	OP3-P	-7.13	1.52	1.61
7	E	94	PRO	N-CA	6.34	1.58	1.47
5	C	439	CYS	CB-SG	-6.08	1.72	1.82

The worst 5 of 164 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	Y	2	A	N9-C1'-C2'	-18.09	90.48	114.00
5	M	409	ARG	NE-CZ-NH1	15.09	127.85	120.30
2	Y	7	G	N9-C1'-C2'	-12.29	98.02	114.00
7	E	94	PRO	CA-N-CD	-11.28	95.71	111.50
2	H	1	G	N9-C1'-C2'	11.20	128.56	114.00

There are no chirality outliers.

5 of 15 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	G	13	DT	Sidechain
1	G	15	DC	Sidechain
1	G	16	DG	Sidechain
1	G	17	DC	Sidechain
1	G	18	DG	Sidechain

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	467	0	259	43	0
1	X	467	0	259	43	0
2	H	347	0	174	58	0
2	Y	347	0	174	81	0
3	I	270	0	144	14	0
3	Z	270	0	144	12	0
4	A	1806	0	1861	169	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	1806	0	1861	174	0
4	K	1806	0	1861	182	0
4	L	1806	0	1861	199	0
5	C	8829	0	8933	1208	0
5	M	8829	0	8933	1204	0
6	D	9960	0	10183	1379	0
6	N	9960	0	10183	1351	0
7	E	770	0	784	108	0
7	O	770	0	784	101	0
8	D	2	0	0	0	0
8	N	2	0	0	0	0
9	D	2	0	0	0	0
9	N	2	0	0	0	0
10	D	31	0	14	2	0
10	N	31	0	14	2	0
11	A	106	0	0	16	0
11	B	82	0	0	21	0
11	C	482	0	0	120	0
11	D	506	0	0	138	0
11	E	60	0	0	6	0
11	G	32	0	0	3	0
11	H	37	0	0	3	0
11	I	22	0	0	3	0
11	K	86	0	0	19	0
11	L	104	0	0	23	0
11	M	483	0	0	129	0
11	N	491	0	0	115	0
11	O	39	0	0	6	0
11	X	43	0	0	4	0
11	Y	30	0	0	6	0
11	Z	30	0	0	4	0
All	All	51213	0	48426	5871	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 61.

The worst 5 of 5871 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)
2:Y:7:G:N1	5:M:1014:SER:HA	1.62	1.13
2:Y:16:G:H21	6:N:705:ALA:HB1	1.11	1.12

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:N:18:ILE:HG23	6:N:518:PRO:HG3	1.33	1.10
5:C:409:ARG:HA	5:C:454:SER:HA	1.27	1.10
6:N:1189:ARG:HB3	6:N:1204:CYS:HA	1.34	1.09

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

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

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

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	A	227/315 (72%)	206 (91%)	14 (6%)	7 (3%)	4	23
4	B	227/315 (72%)	206 (91%)	15 (7%)	6 (3%)	5	27
4	K	227/315 (72%)	208 (92%)	13 (6%)	6 (3%)	5	27
4	L	227/315 (72%)	206 (91%)	15 (7%)	6 (3%)	5	27
5	C	1117/1119 (100%)	919 (82%)	136 (12%)	62 (6%)	2	10
5	M	1117/1119 (100%)	923 (83%)	133 (12%)	61 (6%)	2	10
6	D	1258/1524 (82%)	1051 (84%)	149 (12%)	58 (5%)	2	14
6	N	1258/1524 (82%)	1058 (84%)	140 (11%)	60 (5%)	2	13
7	E	93/99 (94%)	76 (82%)	11 (12%)	6 (6%)	1	7
7	O	93/99 (94%)	74 (80%)	12 (13%)	7 (8%)	1	5
All	All	5844/6744 (87%)	4927 (84%)	638 (11%)	279 (5%)	2	13

5 of 279 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	29	GLU
4	A	187	GLY
4	B	29	GLU
4	B	187	GLY

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
5	C	178	PRO

5.3.2 Protein sidechains ⓘ

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

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

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	A	202/273 (74%)	162 (80%)	40 (20%)	1	7
4	B	202/273 (74%)	159 (79%)	43 (21%)	1	5
4	K	202/273 (74%)	162 (80%)	40 (20%)	1	7
4	L	202/273 (74%)	150 (74%)	52 (26%)	0	3
5	C	941/941 (100%)	704 (75%)	237 (25%)	0	3
5	M	941/941 (100%)	713 (76%)	228 (24%)	0	3
6	D	1063/1279 (83%)	825 (78%)	238 (22%)	1	4
6	N	1063/1279 (83%)	833 (78%)	230 (22%)	1	5
7	E	84/88 (96%)	59 (70%)	25 (30%)	0	1
7	O	84/88 (96%)	68 (81%)	16 (19%)	1	8
All	All	4984/5708 (87%)	3835 (77%)	1149 (23%)	1	4

5 of 1149 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	D	1256	LEU
4	L	67	THR
6	N	1042	ARG
6	D	1365	ASP
4	K	1	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 154 such sidechains are listed below:

Mol	Chain	Res	Type
6	D	1103	HIS
4	K	156	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
6	N	906	GLN
6	D	1227	GLN
7	E	29	GLN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	H	16/16 (100%)	10 (62%)	8 (50%)
2	Y	16/16 (100%)	10 (62%)	8 (50%)
All	All	32/32 (100%)	20 (62%)	16 (50%)

5 of 20 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	H	2	A
2	H	3	G
2	H	6	U
2	H	7	G
2	H	8	C

5 of 16 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	H	15	C
2	Y	1	G
2	Y	9	G
2	H	13	C
2	Y	12	G

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 carbohydrates in this entry.

5.6 Ligand geometry

Of 10 ligands modelled in this entry, 8 are monoatomic - leaving 2 for Mogul analysis.

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

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
10	APC	D	3999	9	27,33,33	1.39	3 (11%)	31,52,52	1.97	8 (25%)
10	APC	N	4999	9	27,33,33	1.25	4 (14%)	31,52,52	1.93	7 (22%)

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
10	APC	D	3999	9	-	8/15/38/38	0/3/3/3
10	APC	N	4999	9	-	8/15/38/38	0/3/3/3

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
10	D	3999	APC	PB-O2B	-3.54	1.48	1.56
10	D	3999	APC	PB-O3B	3.53	1.62	1.58
10	D	3999	APC	PA-O2A	-3.13	1.49	1.56
10	N	4999	APC	PB-O3B	3.13	1.61	1.58
10	N	4999	APC	PB-O2B	-3.02	1.49	1.56

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
10	N	4999	APC	C1'-N9-C4	-6.01	116.08	126.64
10	D	3999	APC	C1'-N9-C4	-5.65	116.72	126.64
10	D	3999	APC	C5-C6-N6	4.52	127.21	120.35
10	N	4999	APC	C5-C6-N6	4.29	126.87	120.35
10	N	4999	APC	PG-O3B-PB	-3.73	119.47	132.62

There are no chirality outliers.

5 of 16 torsion outliers are listed below:

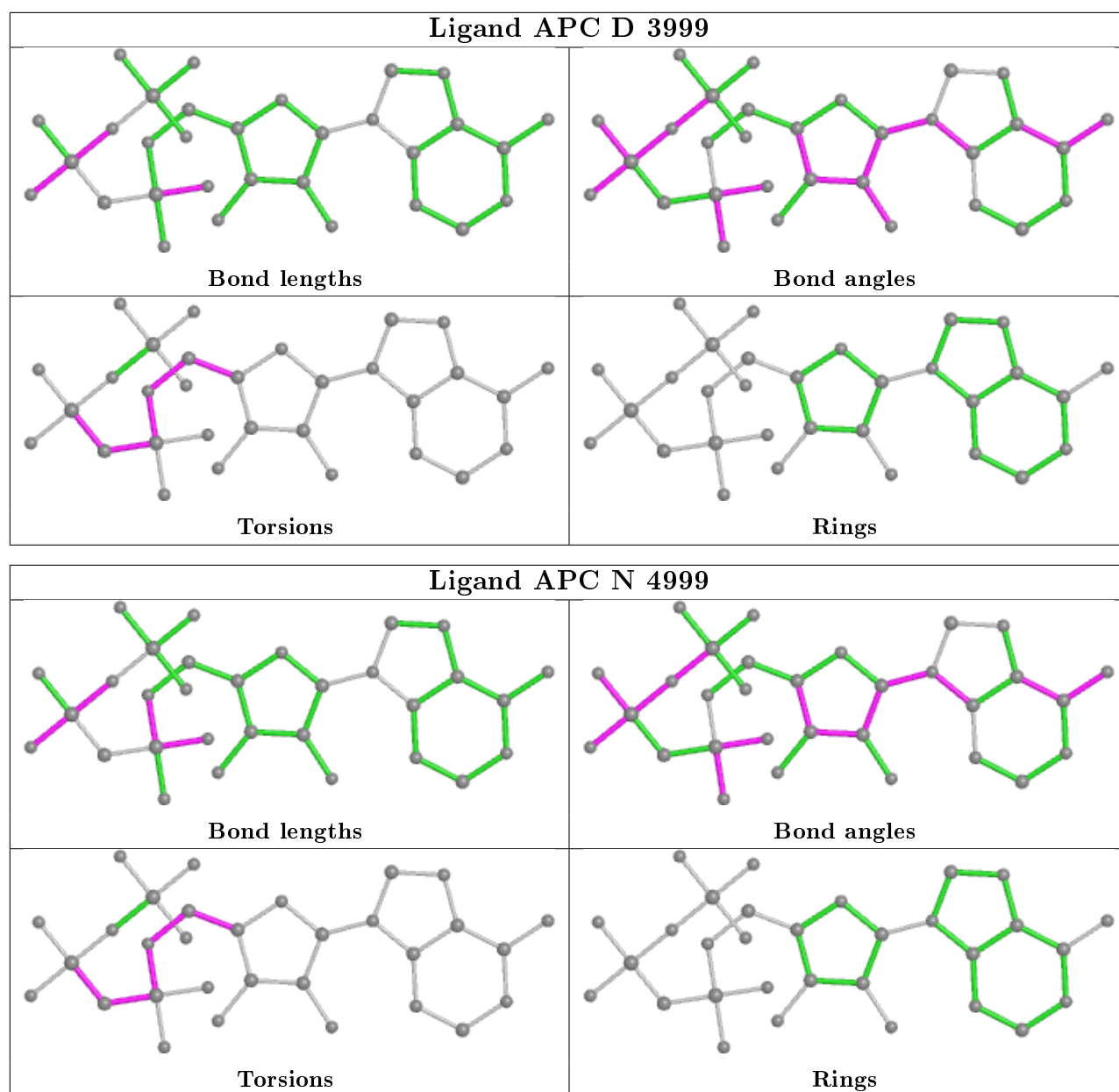
Mol	Chain	Res	Type	Atoms
10	D	3999	APC	PA-C3A-PB-O1B
10	D	3999	APC	PA-C3A-PB-O3B
10	D	3999	APC	PB-C3A-PA-O1A
10	D	3999	APC	PB-C3A-PA-O5'
10	D	3999	APC	O4'-C4'-C5'-O5'

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	D	3999	APC	2	0
10	N	4999	APC	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

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

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

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	G	23/23 (100%)	-0.73	0 100 100	22, 46, 76, 79	0
1	X	23/23 (100%)	-0.67	0 100 100	21, 38, 71, 90	0
2	H	16/16 (100%)	-0.58	0 100 100	36, 56, 100, 102	0
2	Y	16/16 (100%)	-0.62	0 100 100	21, 73, 101, 103	0
3	I	13/14 (92%)	-0.89	0 100 100	49, 65, 74, 85	0
3	Z	13/14 (92%)	-0.86	0 100 100	54, 67, 80, 81	0
4	A	229/315 (72%)	-0.62	0 100 100	41, 66, 86, 94	0
4	B	229/315 (72%)	-0.55	2 (0%) 84 63	43, 70, 85, 101	0
4	K	229/315 (72%)	-0.61	1 (0%) 92 79	41, 65, 81, 93	0
4	L	229/315 (72%)	-0.60	0 100 100	38, 68, 80, 90	0
5	C	1119/1119 (100%)	-0.60	3 (0%) 94 84	26, 64, 86, 101	0
5	M	1119/1119 (100%)	-0.58	6 (0%) 91 75	25, 65, 89, 109	0
6	D	1264/1524 (82%)	-0.65	5 (0%) 92 79	23, 61, 83, 100	0
6	N	1264/1524 (82%)	-0.66	5 (0%) 92 79	26, 60, 84, 100	0
7	E	95/99 (95%)	-0.66	0 100 100	35, 59, 74, 82	0
7	O	95/99 (95%)	-0.67	0 100 100	45, 66, 80, 84	0
All	All	5976/6850 (87%)	-0.62	22 (0%) 92 79	21, 64, 86, 109	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
6	D	416	ALA	4.9
4	B	147	GLY	4.3
6	D	391	ALA	3.5
6	N	427	VAL	3.2
6	D	192	ALA	3.1

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

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

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

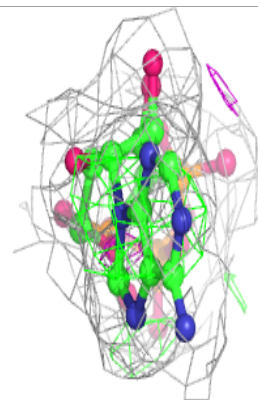
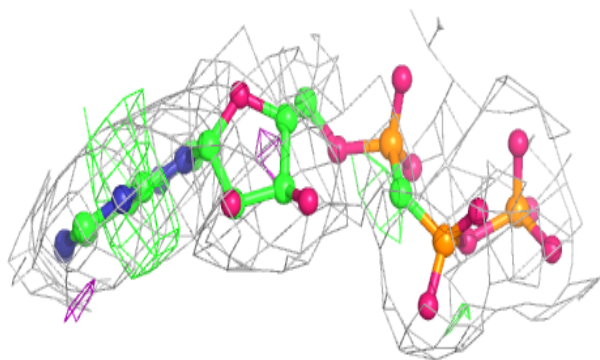
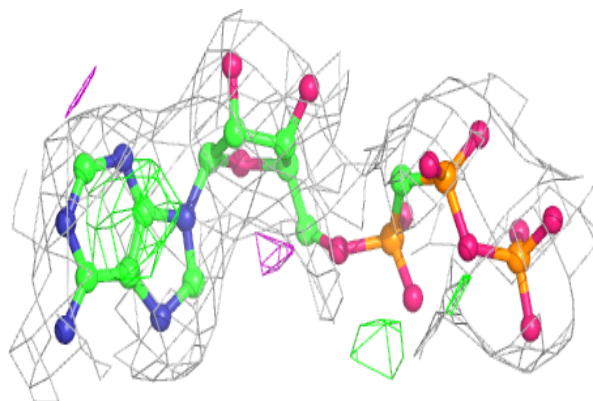
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
8	ZN	N	5058	1/1	0.97	0.09	66,66,66,66	0
10	APC	D	3999	31/31	0.97	0.16	41,49,52,54	0
9	MG	N	9002	1/1	0.98	0.09	23,23,23,23	0
10	APC	N	4999	31/31	0.98	0.16	45,49,51,54	0
8	ZN	D	4058	1/1	0.99	0.09	66,66,66,66	0
8	ZN	D	6112	1/1	0.99	0.14	59,59,59,59	0
9	MG	N	9001	1/1	0.99	0.11	21,21,21,21	0
9	MG	D	8002	1/1	0.99	0.09	25,25,25,25	0
9	MG	D	8001	1/1	0.99	0.10	23,23,23,23	0
8	ZN	N	7112	1/1	1.00	0.13	65,65,65,65	0

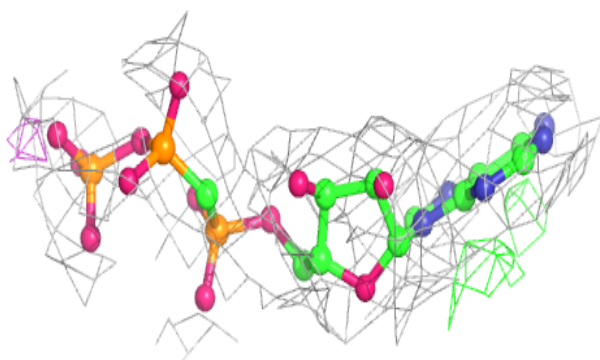
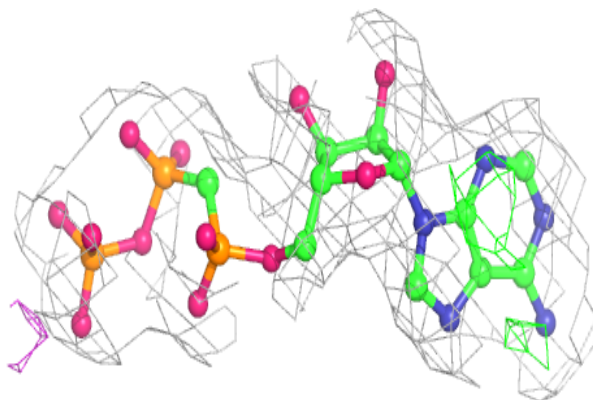
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around APC D 3999:

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

**Electron density around APC N 4999:**

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



6.5 Other polymers [i](#)

There are no such residues in this entry.