wwPDB X-ray Structure Validation Summary Report (i)

Jun 25, 2024 – 04:26 AM EDT

PDB ID	:	5TMC
Title	:	Re-refinement of Thermus thermopiles DNA-directed RNA polymerase structure
Authors	:	Wang, J.
Deposited on	:	2016-10-12
Resolution	:	2.71 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:

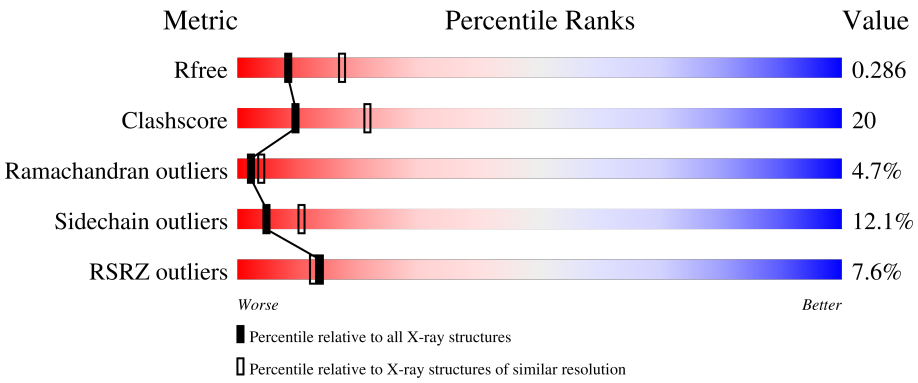
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.71 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	3359 (2.74-2.70)
Clashscore	141614	3686 (2.74-2.70)
Ramachandran outliers	138981	3622 (2.74-2.70)
Sidechain outliers	138945	3623 (2.74-2.70)
RSRZ outliers	127900	3276 (2.74-2.70)

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 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	315	3%; 34%, 31%, 8%, 27%
1	B	315	10%; 46%, 27%, ·, 24%
2	C	1119	6%; 52%, 41%, 7%
3	D	1524	8%; 52%, 40%, 7%, ·
4	E	99	7%; 69%, 23%, ·, ·, ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	423	10% 56% 23% • 17%
6	Z	48	96% •

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 28419 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 subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	231	1816	1159	315	339	3	0	0	0
1	B	238	1863	1188	322	350	3	0	0	0

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

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

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	1504	11864	7518	2091	2219	36	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	95	769	488	133	144	4	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	61	GLU	VAL	conflict	UNP Q72ID6
E	92	ILE	LEU	conflict	UNP Q72ID6
E	95	GLY	VAL	conflict	UNP Q72ID6

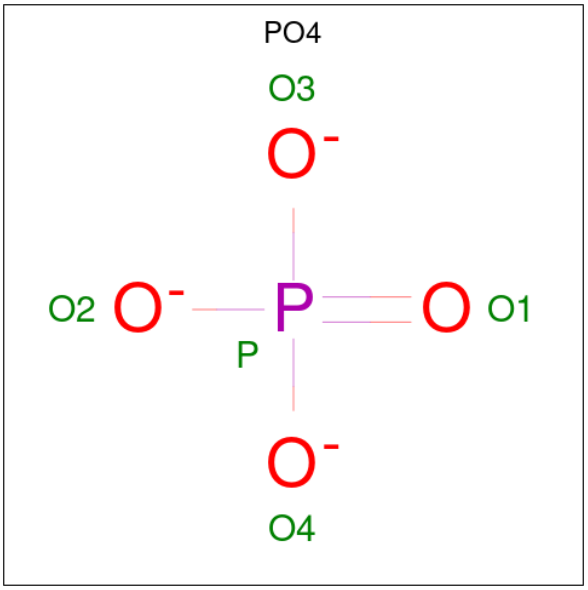
- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	351	2844	1794	515	531	4	0	0	0

- Molecule 6 is a protein called unknown protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
6	Z	48	240	144	48	48	0	0	0

- Molecule 7 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	P		
7	A	1	5	4	1	0	0

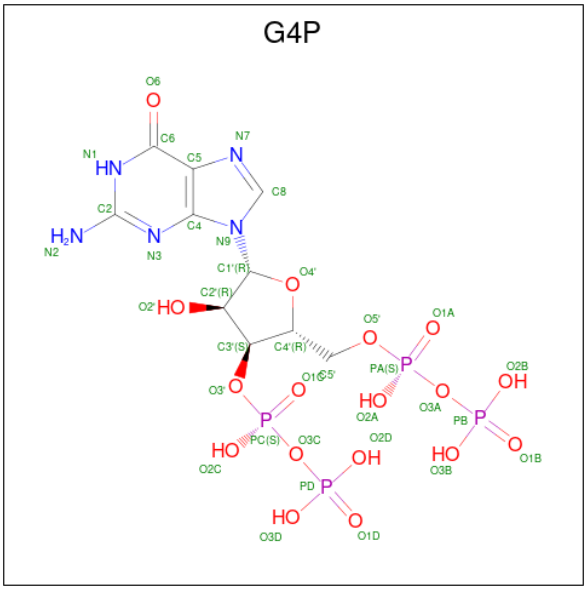
- Molecule 8 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
8	C	1	1	1	0	0
8	D	3	3	3	0	0

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 10 is GUANOSINE-5',3'-TETRAPHOSPHATE (three-letter code: G4P) (formula: $C_{10}H_{17}N_5O_{17}P_4$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
10	D	1	36	10	5	17	4	0	0

- Molecule 11 is water.

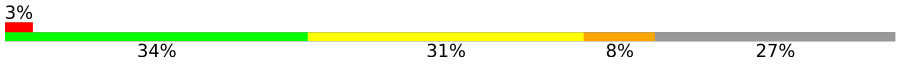
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
11	A	15	15	15	0	0
11	B	2	2	2	0	0
11	C	54	54	54	0	0
11	D	62	62	62	0	0
11	E	9	9	9	0	0
11	F	5	5	5	0	0

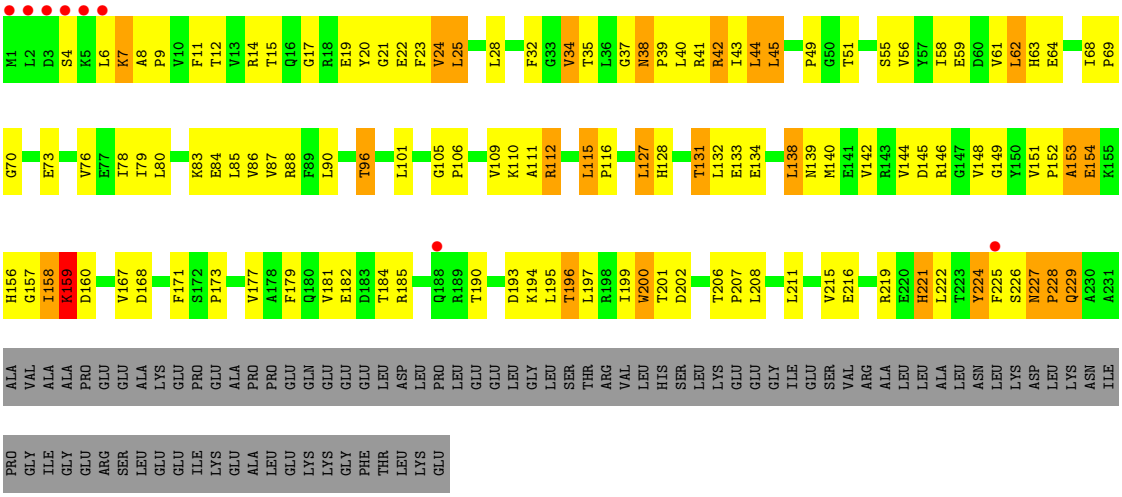
3 Residue-property plots (i)

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 subunit alpha

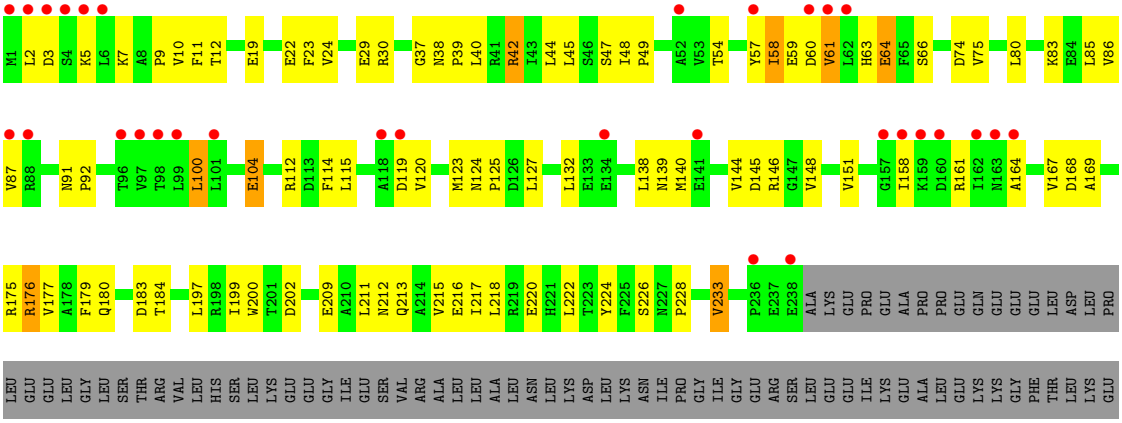
Chain A:





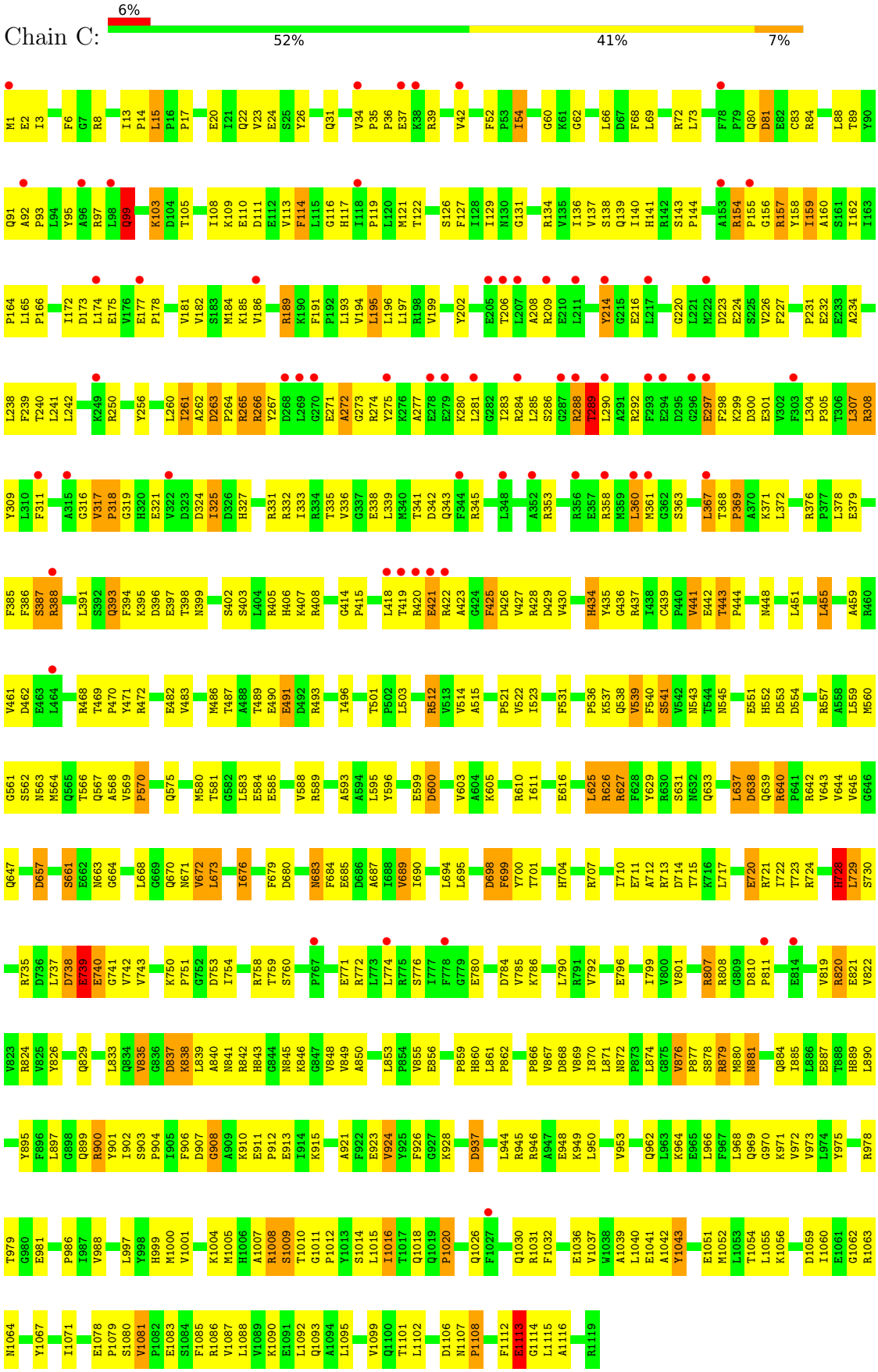
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain B:



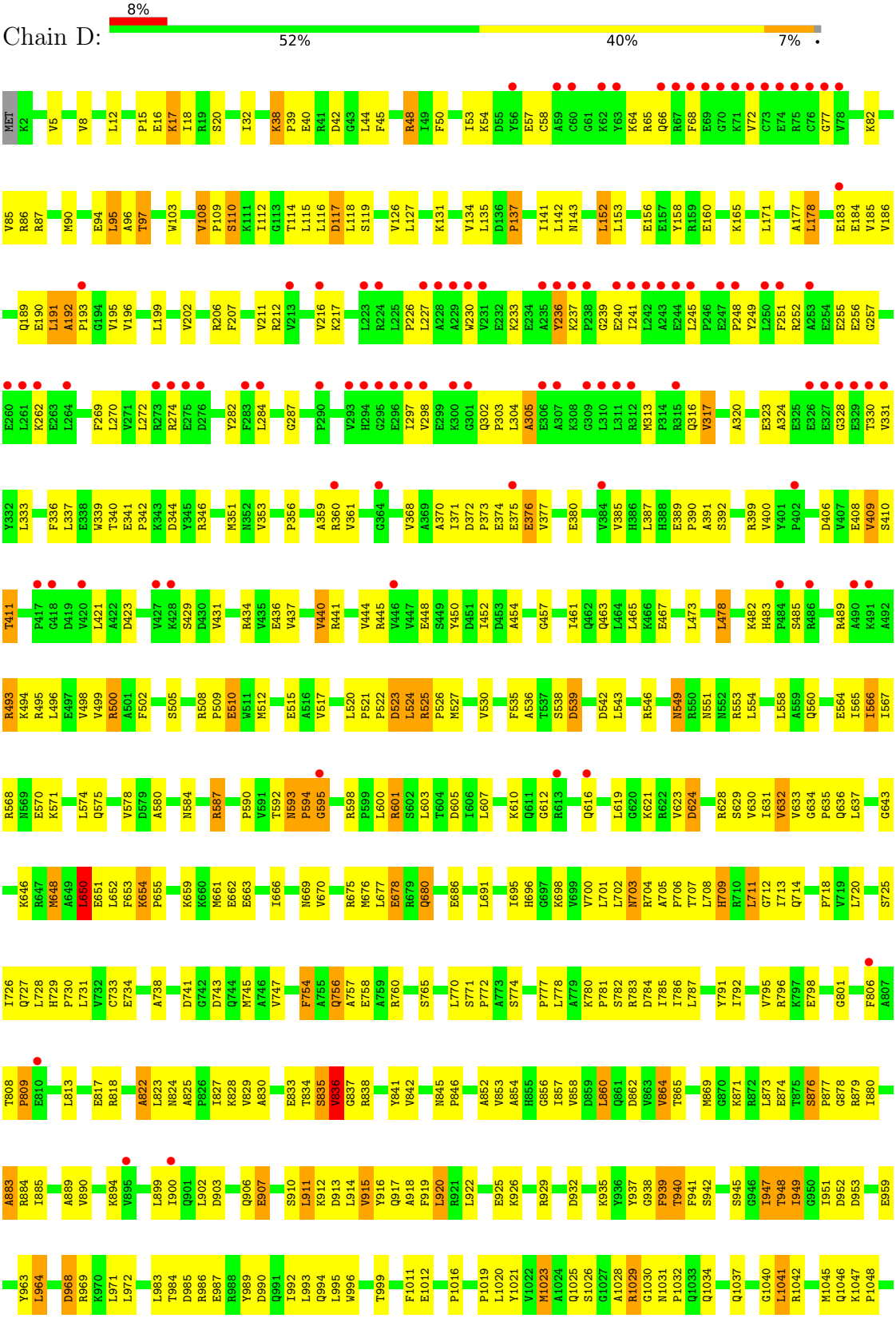
• Molecule 2: DNA-directed RNA polymerase subunit beta

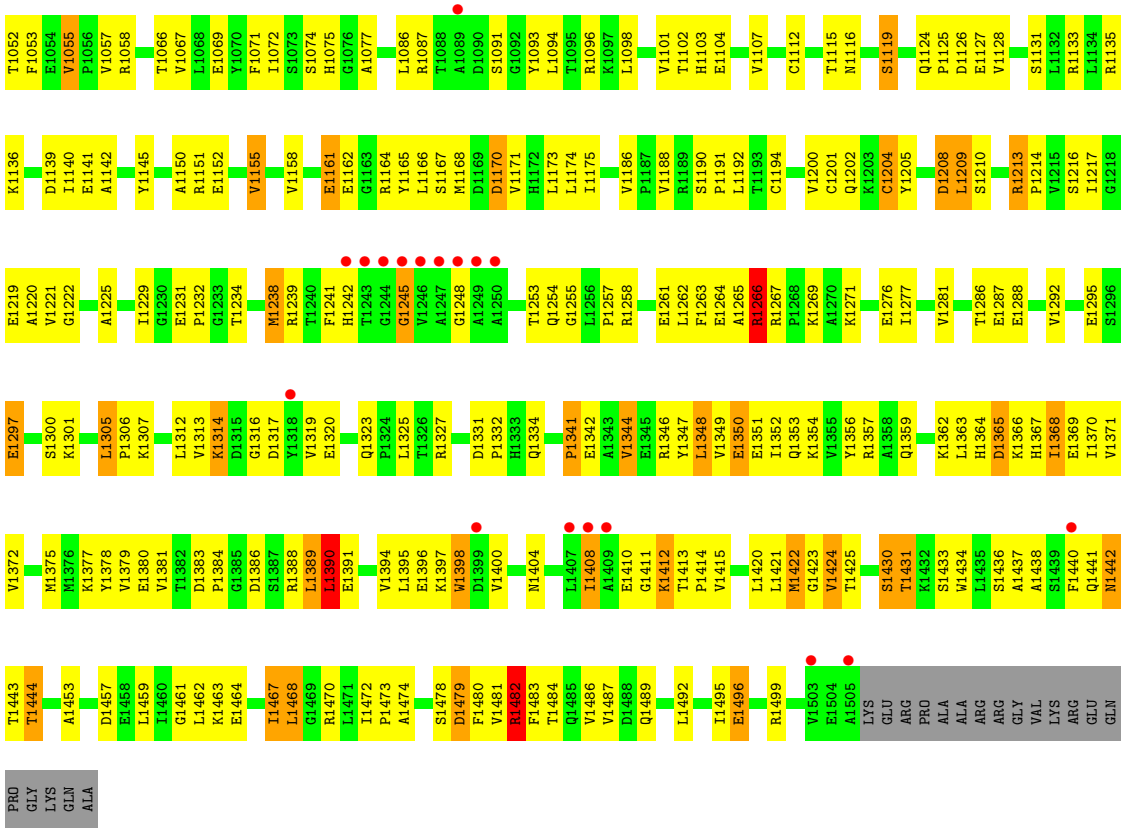
Chain C:



• Molecule 3: DNA-directed RNA polymerase subunit beta'

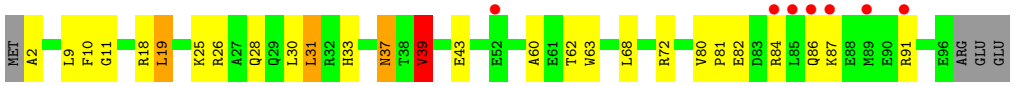
Chain D:





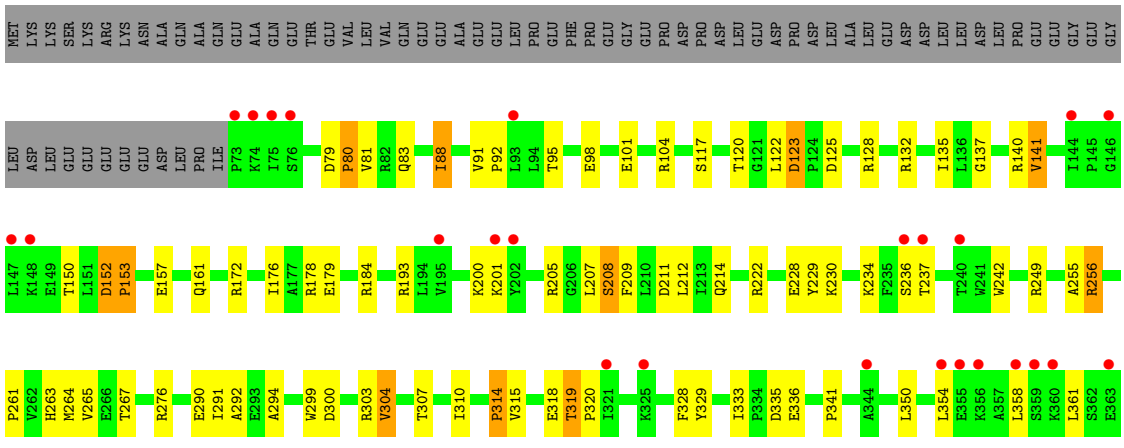
- Molecule 4: DNA-directed RNA polymerase subunit omega

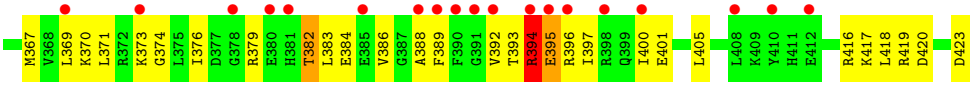
Chain E:



- Molecule 5: RNA polymerase sigma factor SigA

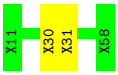
Chain F:





- Molecule 6: unknown protein

Chain Z: 96%



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α, β, γ	236.35Å 236.35Å 249.04Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.51 – 2.71 48.50 – 2.71	Depositor EDS
% Data completeness (in resolution range)	96.9 (48.51-2.71) 97.1 (48.50-2.71)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.06 (at 2.73Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.265 , 0.286 0.265 , 0.286	Depositor DCC
R_{free} test set	7447 reflections (3.61%)	wwPDB-VP
Wilson B-factor (Å^2)	51.2	Xtriage
Anisotropy	0.049	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , 89.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.38$, $< L^2 > = 0.20$	Xtriage
Estimated twinning fraction	0.086 for h,-h-k,-l	Xtriage
Reported twinning fraction	0.875 for H, K, L 0.125 for -H-K, K, -L	Depositor
Outliers	1 of 206105 reflections (0.000%)	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	28419	wwPDB-VP
Average B, all atoms (Å^2)	121.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.68% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: MG, PO4, G4P, ZN

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

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\|>5$	RMSZ	$\#\|Z\|>5$
1	A	0.47	0/1848	0.64	0/2512
1	B	0.39	0/1896	0.57	0/2579
2	C	0.46	0/8997	0.65	0/12164
3	D	0.45	0/12073	0.65	2/16324 (0.0%)
4	E	0.44	0/783	0.63	0/1054
5	F	0.35	0/2890	0.55	0/3888
All	All	0.44	0/28487	0.63	2/38521 (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
3	D	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	95	LEU	CA-CB-CG	5.47	127.89	115.30
3	D	650	LEU	CA-CB-CG	5.43	127.79	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	1126	ASP	Peptide

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	1816	0	1871	90	0
1	B	1863	0	1914	59	0
2	C	8829	0	8933	408	0
3	D	11864	0	12094	544	0
4	E	769	0	775	20	0
5	F	2844	0	2926	68	0
6	Z	240	0	50	1	0
7	A	5	0	0	0	0
8	C	1	0	0	0	0
8	D	3	0	0	0	0
9	D	2	0	0	0	0
10	D	36	0	11	1	0
11	A	15	0	0	0	0
11	B	2	0	0	2	0
11	C	54	0	0	4	0
11	D	62	0	0	7	0
11	E	9	0	0	1	0
11	F	5	0	0	0	0
All	All	28419	0	28574	1111	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 20.

The worst 5 of 1111 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)
3:D:1265:ALA:HA	3:D:1266:ARG:CB	1.60	1.29
3:D:1265:ALA:HA	3:D:1266:ARG:HB2	1.23	1.13
3:D:96:ALA:HB3	3:D:554:LEU:HD23	1.32	1.07
2:C:93:PRO:HA	2:C:117:HIS:HB2	1.36	1.06
2:C:567:GLN:HB2	2:C:997:LEU:HD22	1.41	1.01

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	229/315 (73%)	175 (76%)	40 (18%)	14 (6%)	1 2
1	B	236/315 (75%)	186 (79%)	45 (19%)	5 (2%)	7 16
2	C	1117/1119 (100%)	877 (78%)	187 (17%)	53 (5%)	2 4
3	D	1502/1524 (99%)	1160 (77%)	260 (17%)	82 (6%)	2 3
4	E	93/99 (94%)	76 (82%)	14 (15%)	3 (3%)	4 8
5	F	349/423 (82%)	286 (82%)	54 (16%)	9 (3%)	5 12
All	All	3526/3795 (93%)	2760 (78%)	600 (17%)	166 (5%)	2 4

5 of 166 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	229	GLN
2	C	42	VAL
2	C	263	ASP
2	C	272	ALA
2	C	738	ASP

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	202/273 (74%)	171 (85%)	31 (15%)	2 6
1	B	206/273 (76%)	183 (89%)	23 (11%)	6 13
2	C	941/941 (100%)	816 (87%)	125 (13%)	4 8

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
3	D	1264/1279 (99%)	1111 (88%)	153 (12%)	5 10
4	E	83/87 (95%)	75 (90%)	8 (10%)	8 19
5	F	306/370 (83%)	283 (92%)	23 (8%)	13 30
All	All	3002/3223 (93%)	2639 (88%)	363 (12%)	5 10

5 of 363 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	648	MET
3	D	1202	GLN
3	D	708	LEU
3	D	922	LEU
3	D	1287	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 49 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	463	GLN
3	D	729	HIS
3	D	507	ASN
3	D	703	ASN
3	D	906	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 6 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	$\#\|Z\| > 2$
7	PO4	A	401	-	4,4,4	1.00	0	6,6,6	0.40	0
10	G4P	D	1605	8	30,38,38	1.46	1 (3%)	42,61,61	1.33	5 (11%)

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	G4P	D	1605	8	-	6/23/43/43	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
10	D	1605	G4P	O6-C6	6.65	1.36	1.23

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
10	D	1605	G4P	PA-O3A-PB	-3.63	120.36	132.83
10	D	1605	G4P	PC-O3C-PD	-2.92	122.80	132.83
10	D	1605	G4P	C5-C6-N1	2.55	118.45	113.95
10	D	1605	G4P	C2-N1-C6	-2.45	120.58	125.10
10	D	1605	G4P	C8-N7-C5	2.45	107.66	102.99

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
10	D	1605	G4P	C5'-O5'-PA-O3A

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
10	D	1605	G4P	C5'-O5'-PA-O1A
10	D	1605	G4P	C5'-O5'-PA-O2A
10	D	1605	G4P	O4'-C4'-C5'-O5'
10	D	1605	G4P	C4'-C3'-O3'-PC

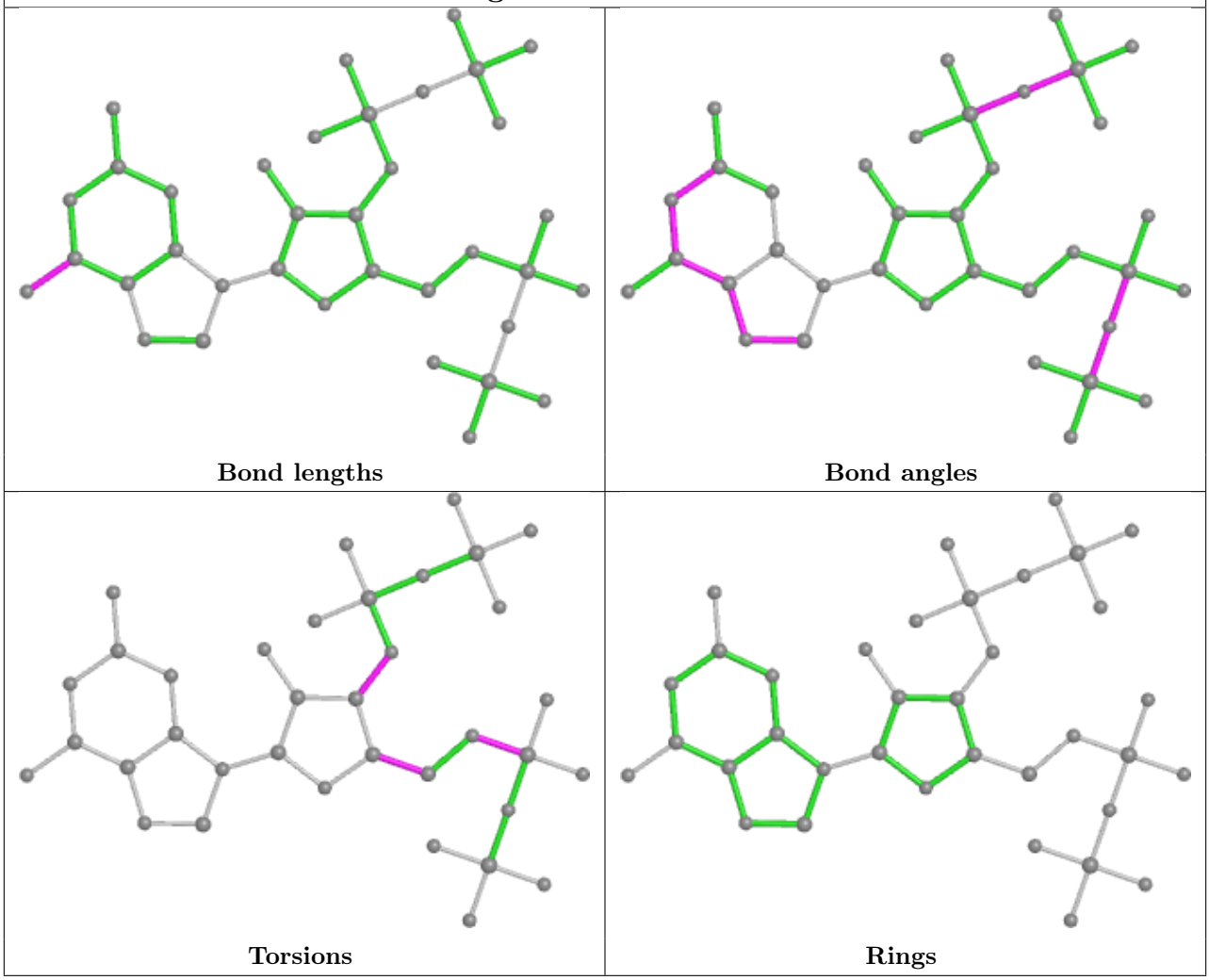
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	D	1605	G4P	1	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 then 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.

Ligand G4P D 1605



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	A	231/315 (73%)	-0.02	8 (3%) 44 44	51, 78, 162, 212	1 (0%)
1	B	238/315 (75%)	0.53	31 (13%) 3 3	70, 149, 236, 260	0
2	C	1119/1119 (100%)	0.07	64 (5%) 23 23	53, 97, 222, 263	2 (0%)
3	D	1504/1524 (98%)	0.22	116 (7%) 13 12	52, 95, 245, 296	4 (0%)
4	E	95/99 (95%)	0.06	7 (7%) 14 13	58, 104, 192, 203	0
5	F	351/423 (82%)	0.60	44 (12%) 3 3	68, 124, 245, 277	2 (0%)
6	Z	0/48	-	-	-	-
All	All	3538/3843 (92%)	0.21	270 (7%) 13 12	51, 102, 232, 296	9 (0%)

The worst 5 of 270 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	391	GLY	25.7
3	D	1247	ALA	17.6
1	A	1	MET	14.6
5	F	390	PHE	13.7
1	A	3	ASP	13.7

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

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

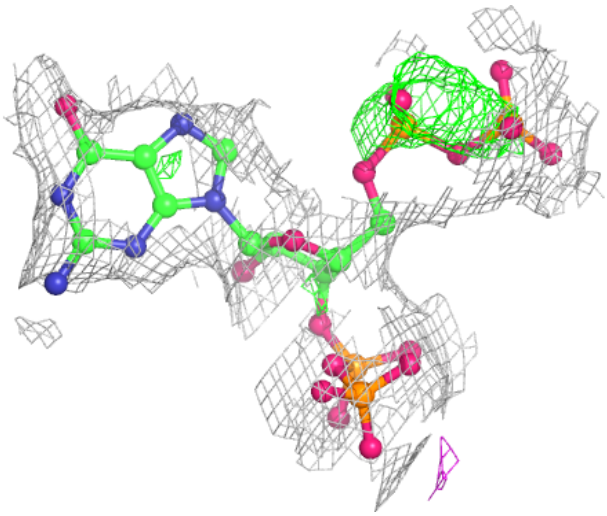
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

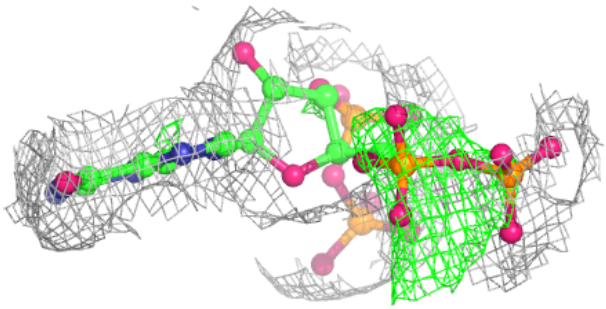
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
9	ZN	D	1602	1/1	0.60	0.08	225,225,225,225	0
8	MG	C	1201	1/1	0.90	0.22	131,131,131,131	0
10	G4P	D	1605	36/36	0.91	0.21	183,197,222,224	0
8	MG	D	1604	1/1	0.94	0.24	86,86,86,86	0
7	PO4	A	401	5/5	0.95	0.31	143,144,148,149	0
9	ZN	D	1603	1/1	0.96	0.11	168,168,168,168	0
8	MG	D	1606	1/1	0.97	0.17	155,155,155,155	0
8	MG	D	1601	1/1	0.98	0.35	133,133,133,133	0

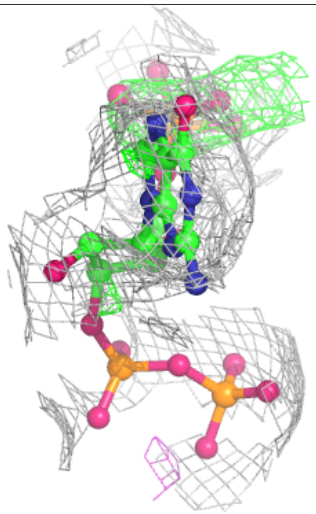
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around G4P D 1605:

$2mF_o$-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.