wwPDB X-ray Structure Validation Summary Report (i)

Jun 24, 2024 – 06:36 AM EDT

PDB ID	:	6TYF
Title	:	Crystal structure of MTB sigma L transcription initiation complex with 6 nt long RNA primer
Authors	:	Molodtsov, V.; Ebright, R.H.
Deposited on	:	2019-08-08
Resolution	:	3.80 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 (1)) were used in the production of this report:

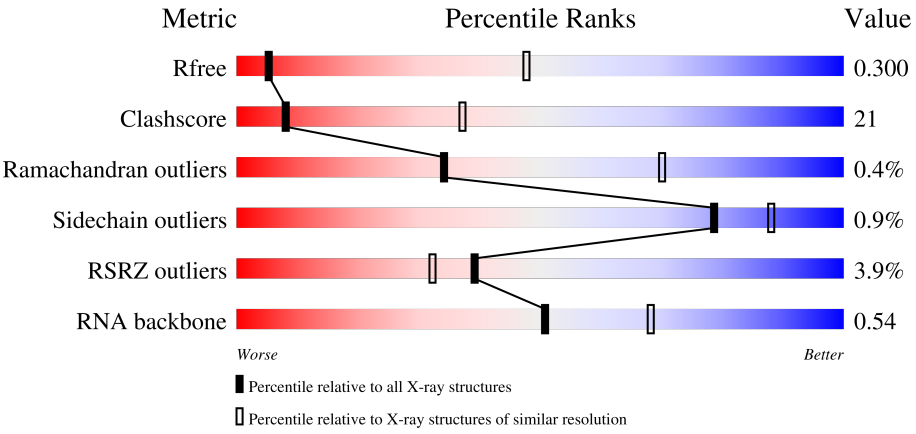
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.80 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.



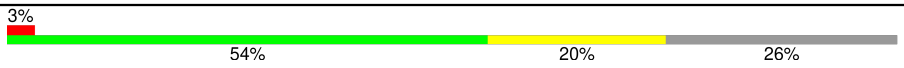
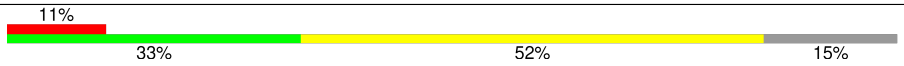
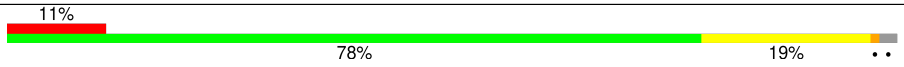
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)
RNA backbone	3102	1036 (4.60-3.00)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=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	347	45% 20% 35%
1	B	347	2% 36% 30% • 33%
2	C	1178	% 60% 34% 5%
3	D	1316	6% 61% 34% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	110	
5	G	19	
6	H	27	
7	I	6	
8	F	177	

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 25308 atoms, of which 372 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
1	A	225	Total	C	N	O	S	0	0	0
			1716	1080	296	338	2			
1	B	232	Total	C	N	O	S	0	0	0
			1732	1093	296	341	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1114	Total	C	N	O	S	0	0	0
			8643	5411	1512	1681	39			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1262	Total	C	N	O	S	0	0	0
			9872	6182	1790	1860	40			

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	E	81	Total	C	N	O	0	0	0
			630	403	106	121			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
5	G	19	Total	C	H	N	O	P	0	0	0
			504	186	113	75	112	18			

- Molecule 6 is a DNA chain called DNA (5'-D(P*GP*TP*GP*TP*CP*AP*GP*TP*AP*GP*CP*TP*GP*TP*CP*AP*CP*GP*GP*AP*TP*GP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
6	H	23	Total	C	H	N	O	P	0	0	0
			734	225	259	87	140	23			

- Molecule 7 is a RNA chain called RNA (5'-R(P*CP*CP*UP*CP*GP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	I	6	Total	C	N	O	P	0	0	0
			125	56	21	42	6			

- Molecule 8 is a protein called RNA polymerase sigma factor.

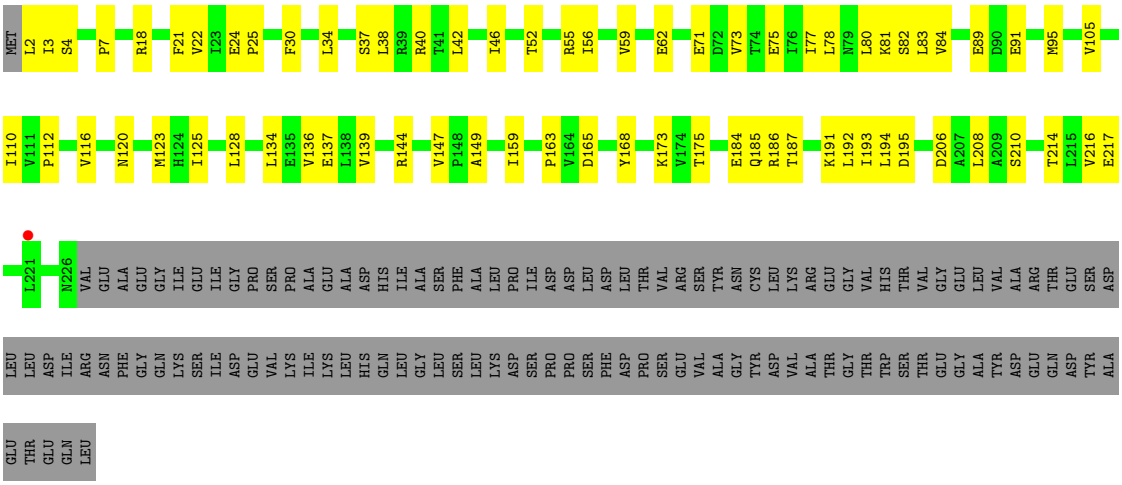
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	F	174	Total	C	N	O	S	0	0	0
			1352	840	256	254	2			

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.

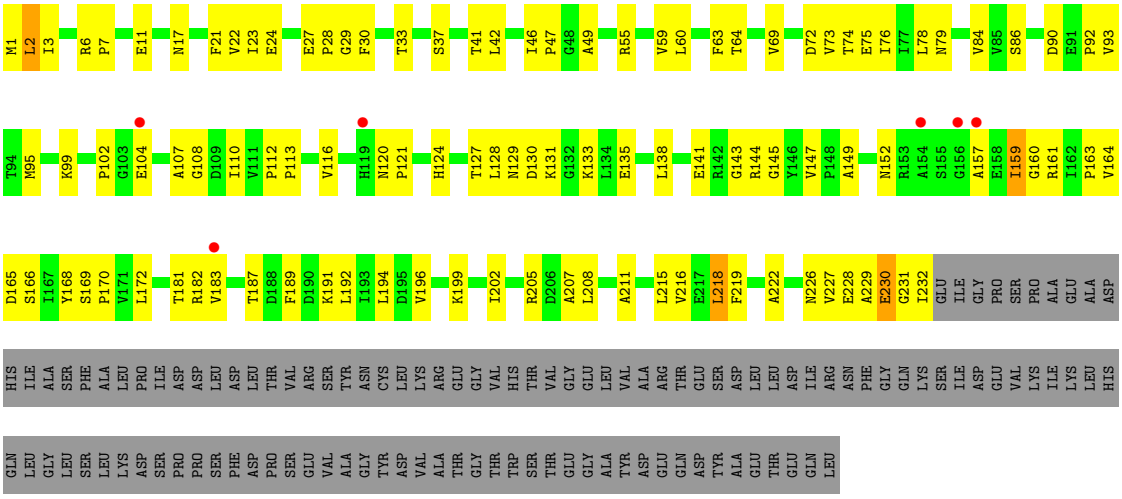
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



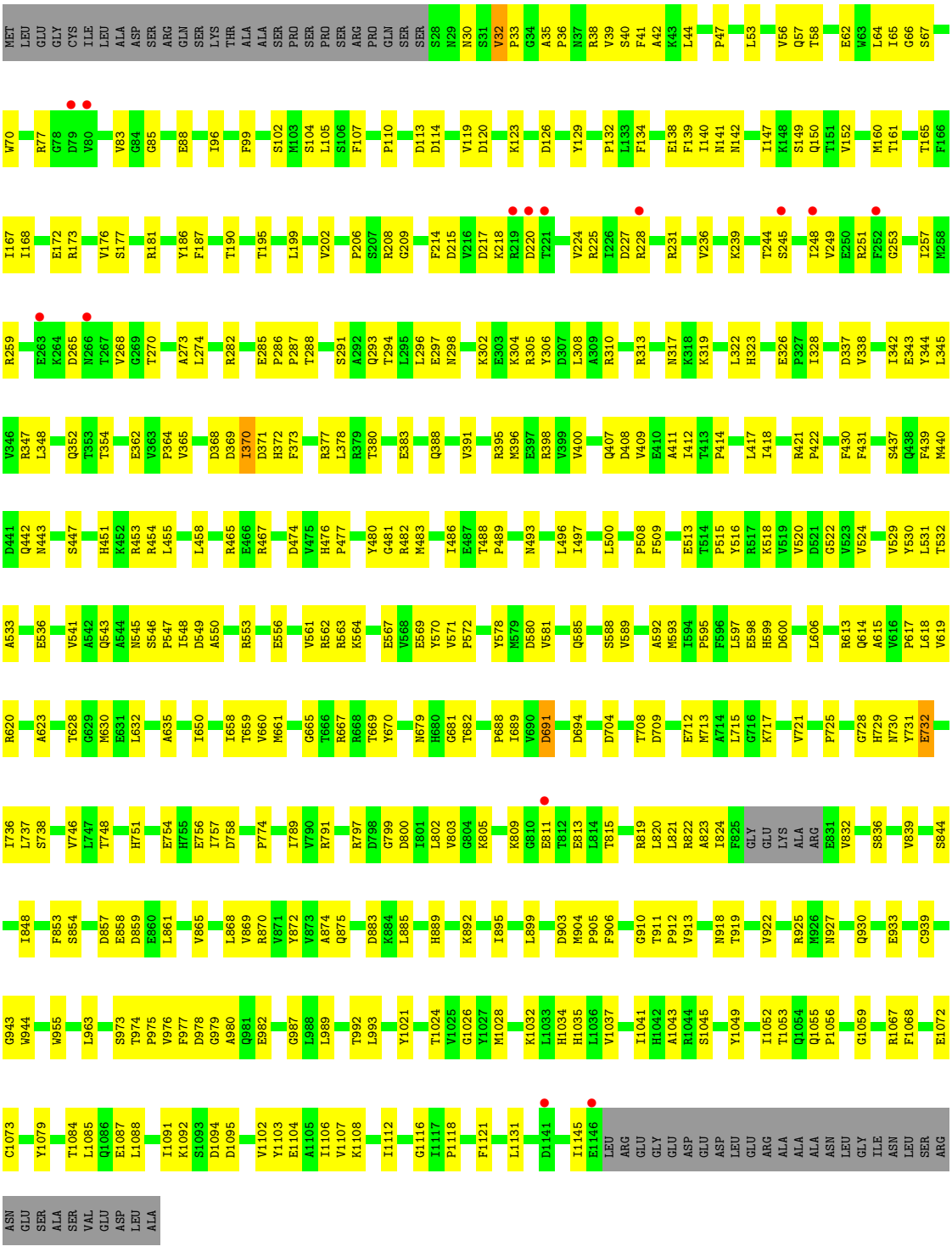
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain B:



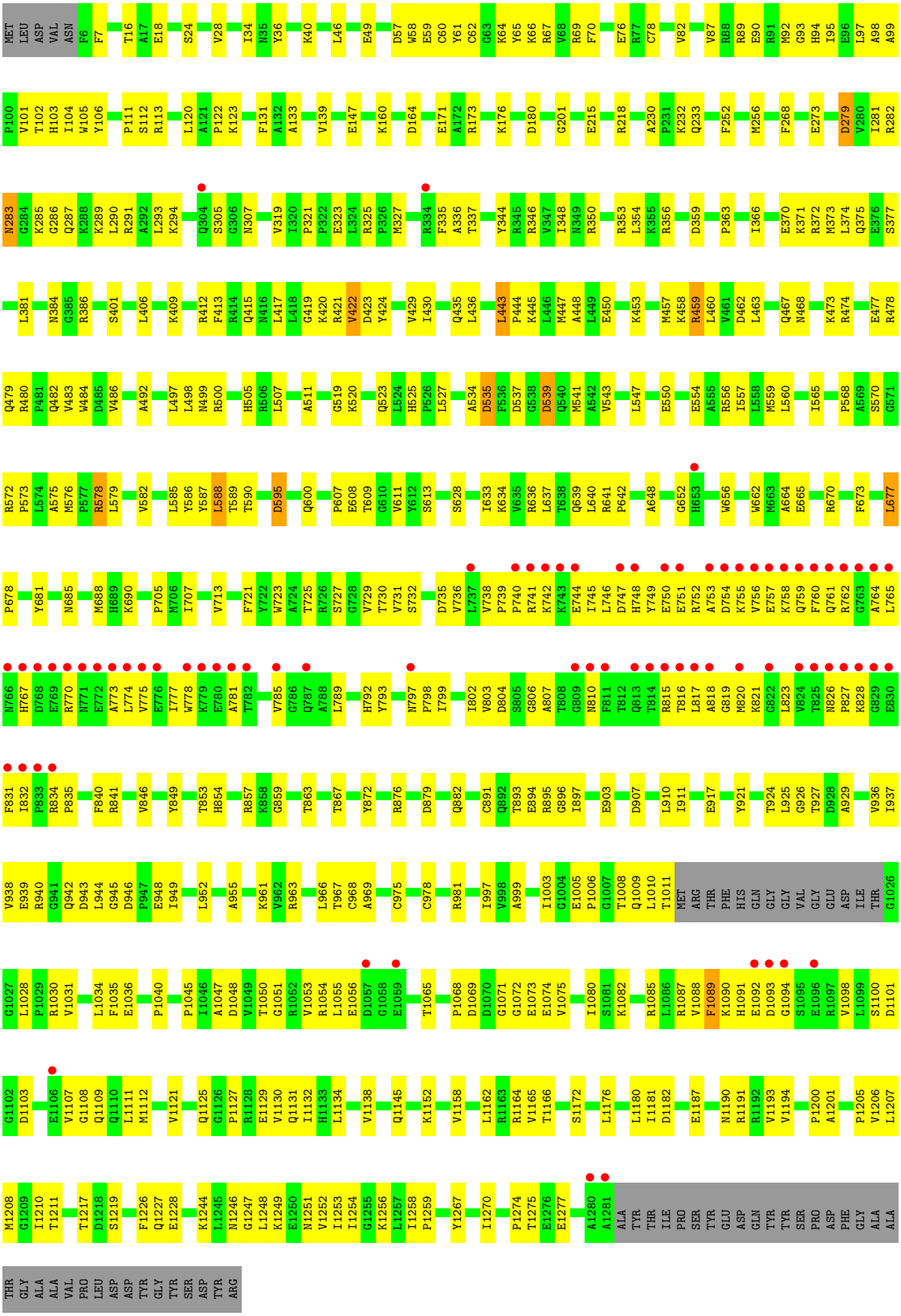
• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain C:



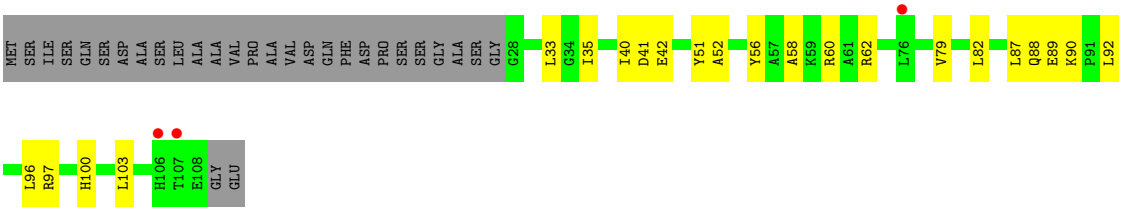
• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D:



• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:

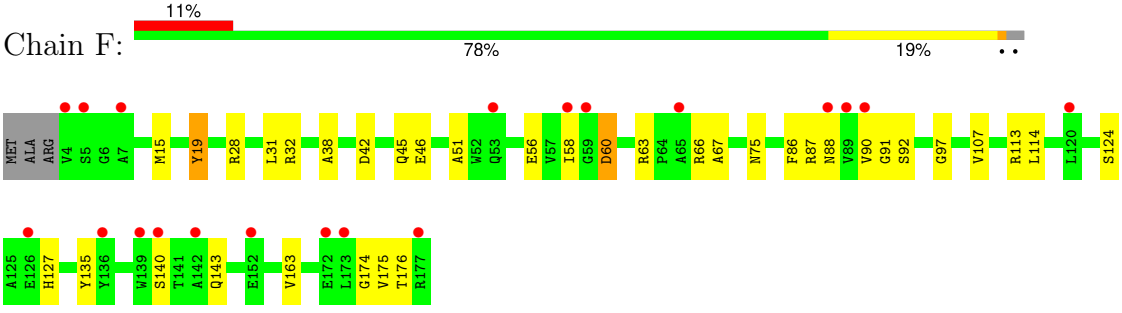


• Molecule 5: DNA (5'-D(*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*AP*TP*CP*GP*AP*GP *GP*GP*T)-3')

• Molecule 6: DNA (5'-D(P*GP*TP*GP*TP*CP*AP*GP*TP*AP*GP*CP*TP*GP*TP*CP*A P*CP*GP*GP*AP*TP*GP*C)-3')

• Molecule 7: RNA (5'-R(P*CP*CP*UP*CP*GP*A)-3')

• Molecule 8: RNA polymerase sigma factor



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	126.43Å 161.78Å 213.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.26 – 3.80 49.26 – 3.80	Depositor EDS
% Data completeness (in resolution range)	96.8 (49.26-3.80) 96.8 (49.26-3.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.87 (at 3.77Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.256 , 0.300 0.257 , 0.300	Depositor DCC
R_{free} test set	1657 reflections (3.90%)	wwPDB-VP
Wilson B-factor (Å^2)	66.1	Xtriage
Anisotropy	0.308	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 32.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	25308	wwPDB-VP
Average B, all atoms (Å^2)	80.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.25% 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

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	0.32	0/1742	0.50	0/2370
1	B	0.32	0/1758	0.52	0/2397
2	C	0.33	0/8801	0.48	0/11933
3	D	0.35	0/10038	0.51	1/13568 (0.0%)
4	E	0.31	0/643	0.46	0/877
5	G	0.89	1/439 (0.2%)	1.12	2/677 (0.3%)
6	H	0.62	0/532	1.03	0/820
7	I	1.45	2/138 (1.4%)	1.19	3/212 (1.4%)
8	F	0.22	0/1374	0.38	0/1869
All	All	0.37	3/25465 (0.0%)	0.54	6/34723 (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

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	I	4	C	O3'-P	-10.26	1.48	1.61
7	I	3	C	O3'-P	-9.50	1.49	1.61
5	G	12	DG	C2'-C1'	6.49	1.58	1.52

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	G	12	DG	C3'-C2'-C1'	-13.45	86.36	102.50
5	G	12	DG	O4'-C4'-C3'	-8.83	100.70	106.00
7	I	4	C	C1'-C2'-O2'	-6.92	89.85	110.60
3	D	422	VAL	CB-CA-C	-5.78	100.42	111.40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	I	3	C	P-O3'-C3'	-5.22	113.43	119.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	419	GLY	Mainchain

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	1716	0	1756	74	0
1	B	1732	0	1754	107	0
2	C	8643	0	8575	400	2
3	D	9872	0	9942	517	4
4	E	630	0	622	28	0
5	G	391	113	215	19	0
6	H	475	259	260	40	0
7	I	125	0	66	13	0
8	F	1352	0	1346	64	0
All	All	24936	372	24536	1052	4

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 21.

The worst 5 of 1052 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:40:ARG:HE	1:B:33:THR:HG22	1.14	1.12
3:D:111:PRO:HB3	6:H:22:DA:H5"	1.14	1.11
3:D:111:PRO:HB3	6:H:22:DA:C5'	1.82	1.09
2:C:1103:TYR:CE2	8:F:107:VAL:HG13	1.89	1.07
3:D:1055:LEU:H	3:D:1101:ASP:HB3	1.17	1.06

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:747:ASP:N	3:D:894:GLU:OE2[4_545]	1.99	0.21
3:D:754:ASP:O	3:D:939:GLU:OE2[4_545]	2.06	0.14
2:C:791:ARG:NE	3:D:171:GLU:OE2[3_554]	2.13	0.07
2:C:667:ARG:NH2	3:D:147:GLU:OE1[3_554]	2.17	0.03

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	223/347 (64%)	215 (96%)	7 (3%)	1 (0%)	34 70
1	B	230/347 (66%)	211 (92%)	15 (6%)	4 (2%)	9 43
2	C	1110/1178 (94%)	1055 (95%)	50 (4%)	5 (0%)	29 66
3	D	1258/1316 (96%)	1201 (96%)	55 (4%)	2 (0%)	47 79
4	E	79/110 (72%)	76 (96%)	3 (4%)	0	100 100
8	F	172/177 (97%)	167 (97%)	5 (3%)	0	100 100
All	All	3072/3475 (88%)	2925 (95%)	135 (4%)	12 (0%)	34 70

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	159	ILE
2	C	370	ILE
1	B	2	LEU
2	C	732	GLU
3	D	1089	PHE

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	194/297 (65%)	194 (100%)	0	100 100
1	B	191/297 (64%)	189 (99%)	2 (1%)	76 86
2	C	945/998 (95%)	941 (100%)	4 (0%)	91 95
3	D	1047/1095 (96%)	1034 (99%)	13 (1%)	71 84
4	E	66/90 (73%)	66 (100%)	0	100 100
8	F	134/136 (98%)	129 (96%)	5 (4%)	34 62
All	All	2577/2913 (88%)	2553 (99%)	24 (1%)	78 88

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	588	LEU
3	D	804	ASP
3	D	677	LEU
3	D	940	ARG
3	D	279	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	748	HIS
3	D	1109	GLN
8	F	45	GLN
2	C	679	ASN
3	D	307	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
7	I	5/6 (83%)	2 (40%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
7	I	4	C
7	I	8	A

There are no RNA pucker outliers to report.

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)

There are no ligands in this entry.

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	225/347 (64%)	-0.17	1 (0%) 92 89	38, 61, 104, 157	0
1	B	232/347 (66%)	0.10	6 (2%) 56 47	48, 74, 109, 168	0
2	C	1114/1178 (94%)	-0.10	14 (1%) 77 70	24, 57, 132, 167	0
3	D	1262/1316 (95%)	0.17	76 (6%) 21 17	25, 66, 203, 265	0
4	E	81/110 (73%)	0.16	3 (3%) 41 34	51, 77, 133, 142	0
5	G	19/19 (100%)	-0.03	0 100 100	38, 65, 133, 140	0
6	H	23/27 (85%)	0.85	3 (13%) 3 4	74, 134, 200, 210	0
7	I	6/6 (100%)	0.71	0 100 100	38, 45, 67, 93	0
8	F	174/177 (98%)	0.75	20 (11%) 4 5	66, 123, 159, 174	0
All	All	3136/3527 (88%)	0.08	123 (3%) 39 32	24, 67, 151, 265	0

The worst 5 of 123 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	771	ASN	14.6
3	D	814	THR	12.1
3	D	767	HIS	10.7
3	D	761	GLN	8.4
3	D	830	GLU	7.7

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 monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.