wwPDB X-ray Structure Validation Summary Report (i)

Jul 18, 2016 – 04:24 PM EDT

PDB ID	:	5E18
Title	:	T. thermophilus transcription initiation complex having a YYY discriminator sequence and a nontemplate-strand length corresponding to TSS selection at position 8 (RPo-CCC-8)
Authors	:	Zhang, Y.; Ebright, R.H.
Deposited on	:	2015-09-29
Resolution	:	3.30 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

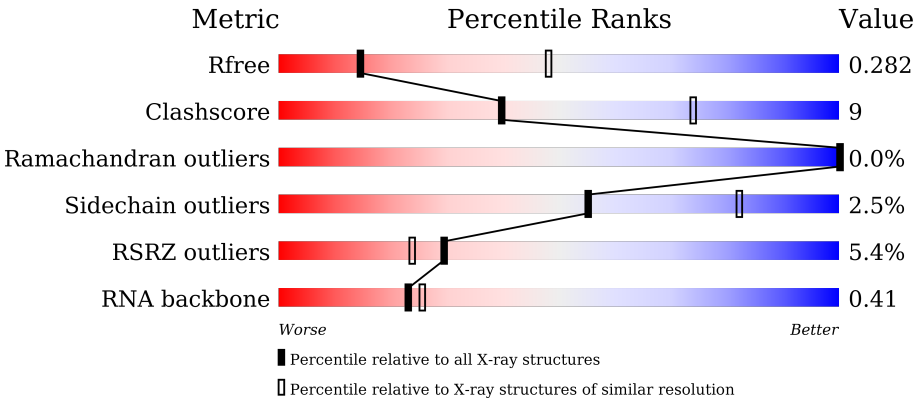
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027790
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027790

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 3.30 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.



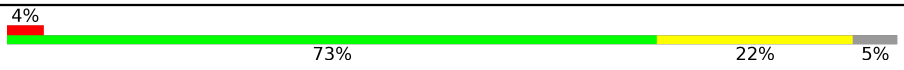
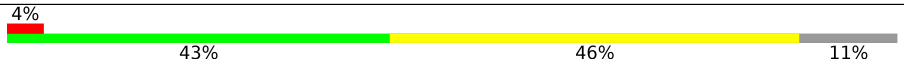
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2060 (3.40-3.20)
Clashscore	102246	1058 (3.38-3.22)
Ramachandran outliers	100387	1038 (3.38-3.22)
Sidechain outliers	100360	1037 (3.38-3.22)
RSRZ outliers	91569	2070 (3.40-3.20)
RNA backbone	2183	1005 (3.82-2.78)

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. 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	4% 61% 12% 27%
1	B	315	% 53% 17% • 29%
2	C	1119	4% 76% 23% ••
3	D	1524	6% 74% 23% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	99	
5	F	443	
6	G	21	
7	I	7	
8	H	28	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
9	MG	D	2004	-	-	-	X

2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 28645 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
1	A	231	Total	C	N	O	S	0	1	0
			1814	1158	316	338	2			
1	B	223	Total	C	N	O	S	0	0	0
			1758	1124	305	327	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1112	Total	C	N	O	S	0	3	0
			8792	5562	1570	1636	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1486	Total	C	N	O	S	0	2	0
			11751	7450	2070	2195	36			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	94	Total	C	N	O	S	0	0	0
			758	483	132	139	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	335	Total	C	N	O	S	0	0	0
			2718	1713	497	504	4			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-19	MET	-	initiating methionine	UNP Q5SKW1
F	-18	GLY	-	expression tag	UNP Q5SKW1
F	-17	SER	-	expression tag	UNP Q5SKW1
F	-16	SER	-	expression tag	UNP Q5SKW1
F	-15	HIS	-	expression tag	UNP Q5SKW1
F	-14	HIS	-	expression tag	UNP Q5SKW1
F	-13	HIS	-	expression tag	UNP Q5SKW1
F	-12	HIS	-	expression tag	UNP Q5SKW1
F	-11	HIS	-	expression tag	UNP Q5SKW1
F	-10	HIS	-	expression tag	UNP Q5SKW1
F	-9	SER	-	expression tag	UNP Q5SKW1
F	-8	SER	-	expression tag	UNP Q5SKW1
F	-7	GLY	-	expression tag	UNP Q5SKW1
F	-6	LEU	-	expression tag	UNP Q5SKW1
F	-5	VAL	-	expression tag	UNP Q5SKW1
F	-4	PRO	-	expression tag	UNP Q5SKW1
F	-3	ARG	-	expression tag	UNP Q5SKW1
F	-2	GLY	-	expression tag	UNP Q5SKW1
F	-1	SER	-	expression tag	UNP Q5SKW1
F	0	HIS	-	expression tag	UNP Q5SKW1

- Molecule 6 is a DNA chain called DNA (5'-D(*CP*CP*T*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*GP*TP*CP*GP*AP*GP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	18	Total	C	N	O	P	0	0	0
			372	176	73	106	17			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	I	7	Total	C	N	O	P	0	0	0
			142	65	24	47	6			

- Molecule 8 is a DNA chain called DNA (28-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	25	Total	C	N	O	P	0	0	0
			508	243	93	148	24			

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	B	1	Total Mg 1 1	0	0
9	D	2	Total Mg 2 2	0	0
9	F	1	Total Mg 1 1	0	0

- Molecule 10 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

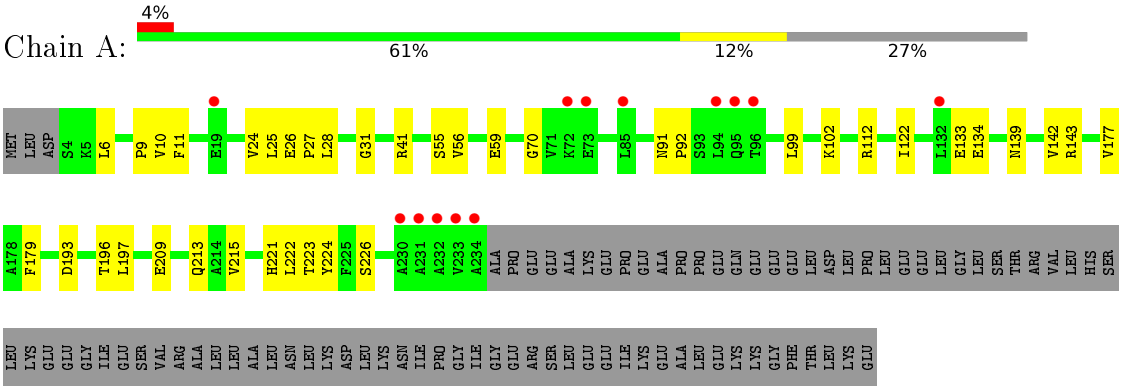
- Molecule 11 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
11	A	1	Total O 1 1	0	0
11	B	3	Total O 3 3	0	0
11	C	8	Total O 8 8	0	0
11	D	10	Total O 10 10	0	0
11	E	1	Total O 1 1	0	0
11	F	2	Total O 2 2	0	0
11	H	1	Total O 1 1	0	0

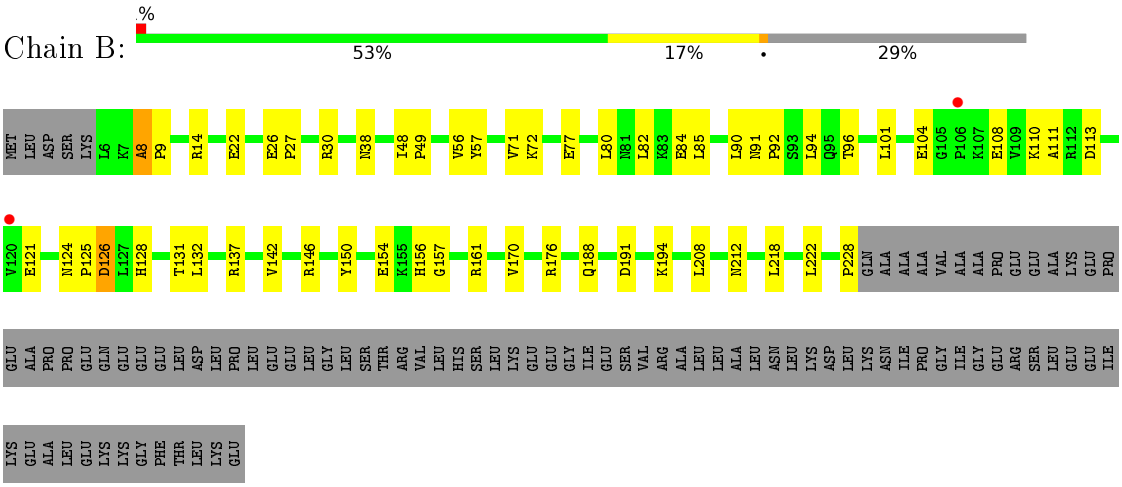
3 Residue-property plots (i)

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

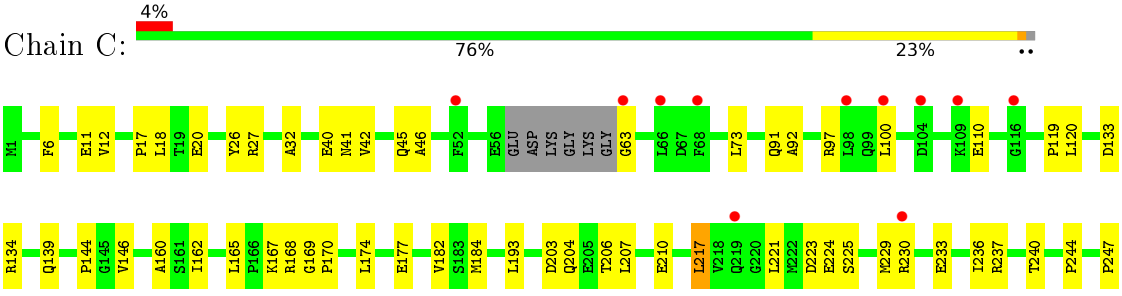
- Molecule 1: DNA-directed RNA polymerase subunit alpha

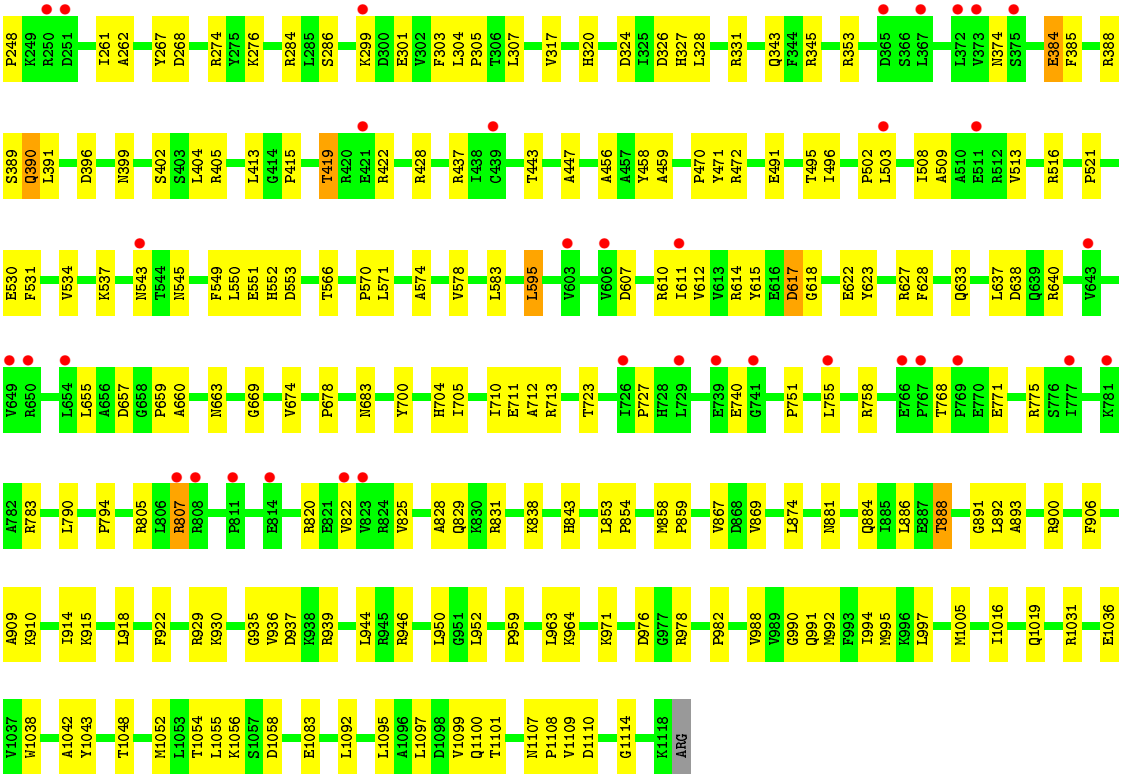


- Molecule 1: DNA-directed RNA polymerase subunit alpha



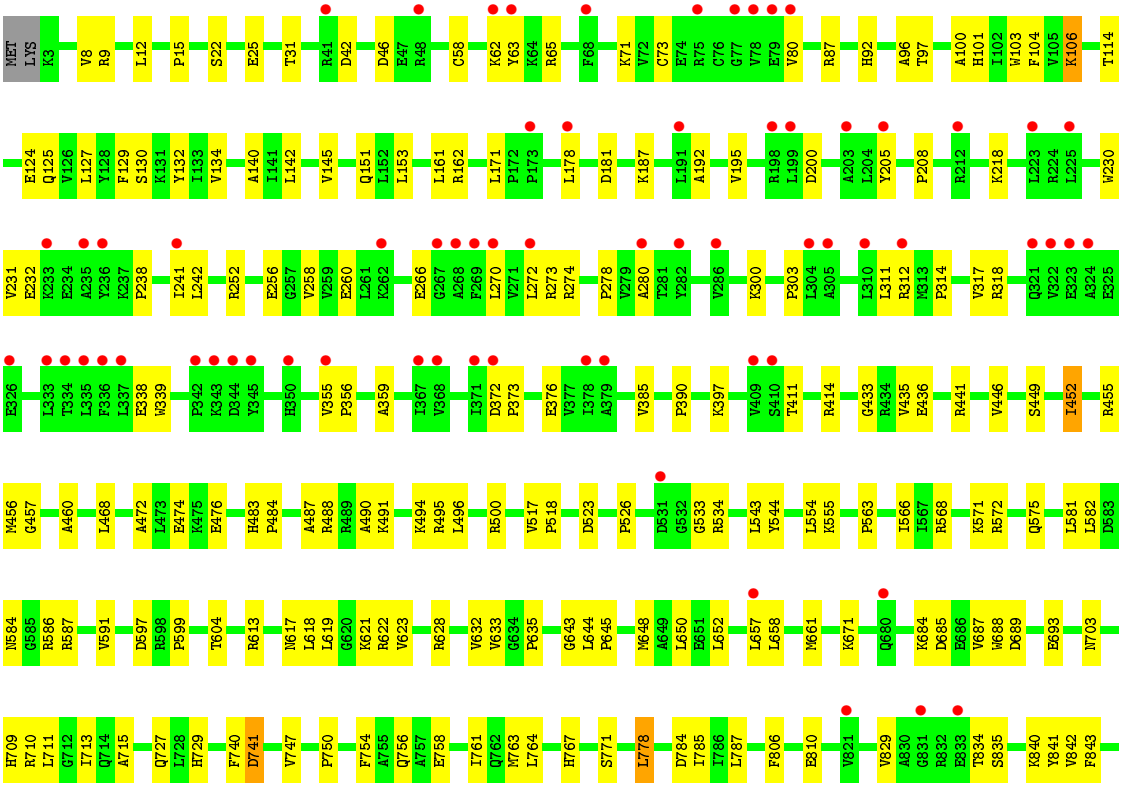
- Molecule 2: DNA-directed RNA polymerase subunit beta

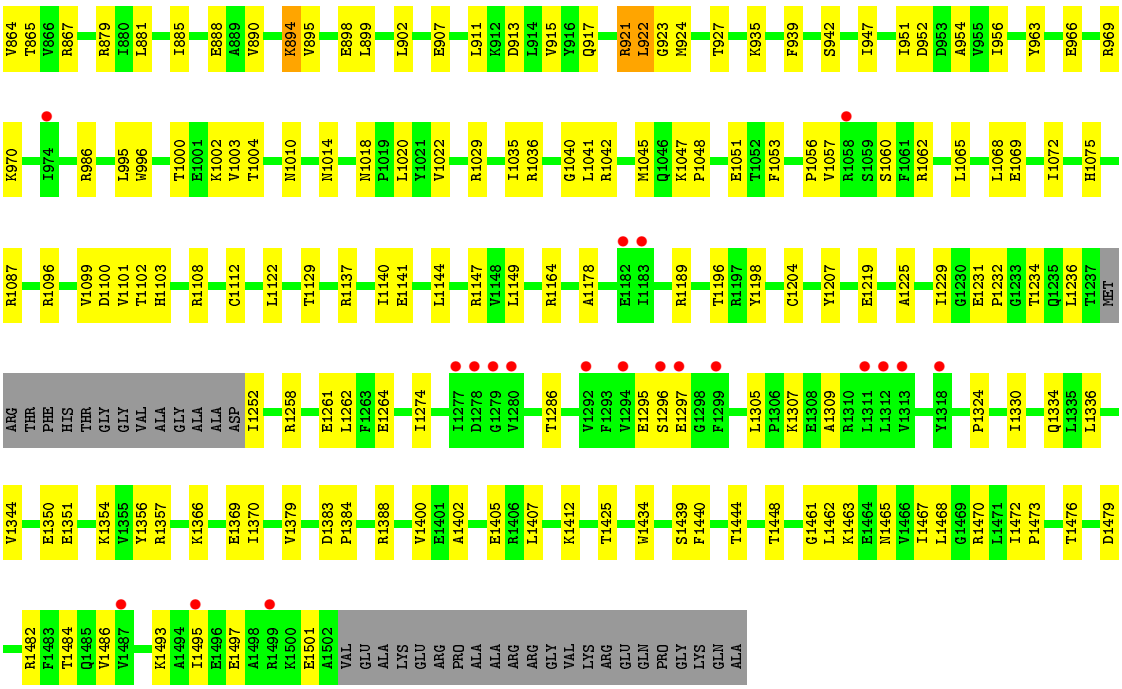




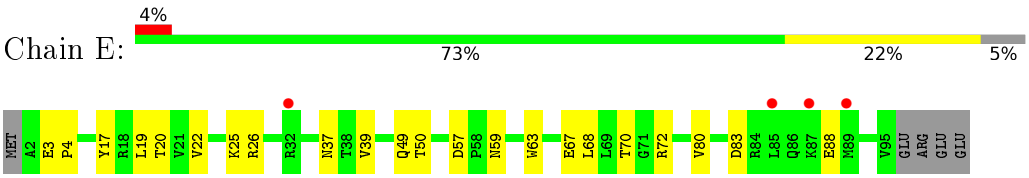
• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D:

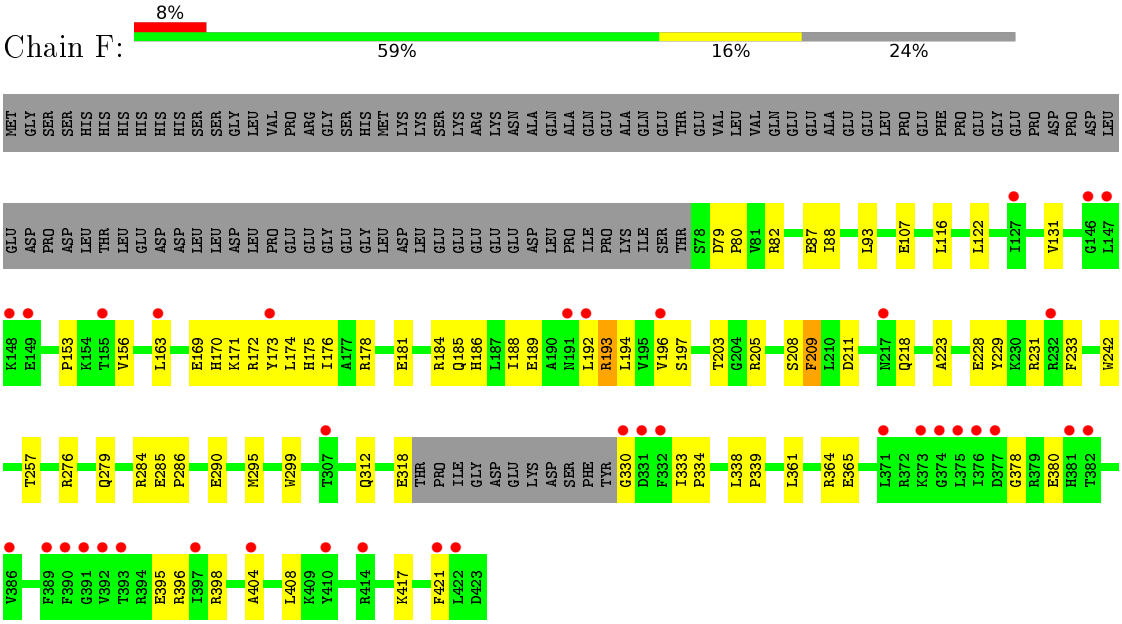




• Molecule 4: DNA-directed RNA polymerase subunit omega



• Molecule 5: RNA polymerase sigma factor SigA

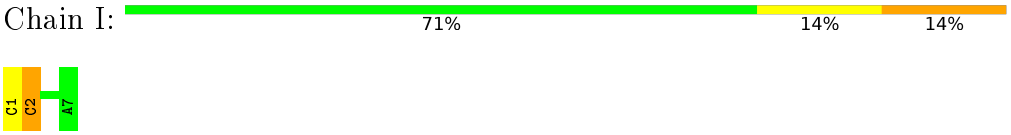


• Molecule 6: DNA (5'-D(*CP*CP*T*GP*CP*AP*TP*CP*CP*GP*TP*GP*AP*GP*TP*CP*GP*AP*GP*GP*G)-3')

Chain G:

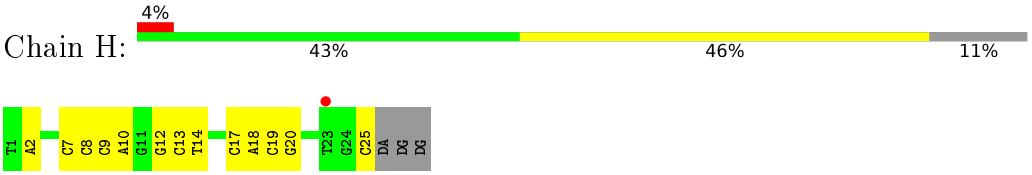
- Molecule 7: RNA (5'-R(*CP*CP*CP*UP*CP*GP*A)-3')

Chain I:



- Molecule 8: DNA (28-MER)

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	185.92Å 103.24Å 298.40Å 90.00° 97.93° 90.00°	Depositor
Resolution (Å)	38.34 – 3.30 39.66 – 3.28	Depositor EDS
% Data completeness (in resolution range)	90.3 (38.34-3.30) 90.4 (39.66-3.28)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	1.10 (at 3.25Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.238 , 0.284 0.238 , 0.282	Depositor DCC
R_{free} test set	3826 reflections (4.96%)	DCC
Wilson B-factor (Å^2)	101.6	Xtriage
Anisotropy	0.584	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 50.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	28645	wwPDB-VP
Average B, all atoms (Å^2)	115.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.25% 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

Bond lengths and bond angles in the following residue types are not validated in this section: 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	A	0.24	0/1849	0.43	0/2515
1	B	0.24	0/1790	0.47	0/2435
2	C	0.25	0/8969	0.44	0/12129
3	D	0.26	0/11963	0.44	0/16173
4	E	0.24	0/772	0.41	0/1040
5	F	0.29	0/2759	0.44	0/3709
6	G	0.53	0/418	0.80	0/645
7	I	0.26	0/157	0.72	0/242
8	H	0.64	0/569	0.91	0/876
All	All	0.27	0/29246	0.47	0/39764

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	A	1814	0	1869	23	0
1	B	1758	0	1808	40	0
2	C	8792	0	8902	164	0
3	D	11751	0	11994	223	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	758	0	770	16	0
5	F	2718	0	2803	53	0
6	G	372	0	203	12	0
7	I	142	0	78	3	0
8	H	508	0	283	35	0
9	B	1	0	0	0	0
9	D	2	0	0	0	0
9	F	1	0	0	0	0
10	D	2	0	0	0	0
11	A	1	0	0	0	0
11	B	3	0	0	0	0
11	C	8	0	0	0	0
11	D	10	0	0	3	0
11	E	1	0	0	0	0
11	F	2	0	0	0	0
11	H	1	0	0	0	0
All	All	28645	0	28710	498	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 9.

The worst 5 of 498 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)
6:G:20:DG:H1	7:I:2:C:H42	1.18	0.88
8:H:13:DC:O3'	8:H:14:DT:H72	1.74	0.86
5:F:209:PHE:HB2	8:H:9:DC:C4	2.12	0.84
2:C:628:PHE:H	2:C:638:ASP:HB3	1.43	0.83
2:C:950:LEU:HB3	2:C:952:LEU:HD13	1.66	0.77

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	230/315 (73%)	227 (99%)	3 (1%)	0	100	100
1	B	221/315 (70%)	216 (98%)	4 (2%)	1 (0%)	34	71
2	C	1111/1119 (99%)	1076 (97%)	35 (3%)	0	100	100
3	D	1484/1524 (97%)	1433 (97%)	51 (3%)	0	100	100
4	E	92/99 (93%)	89 (97%)	3 (3%)	0	100	100
5	F	331/443 (75%)	327 (99%)	4 (1%)	0	100	100
All	All	3469/3815 (91%)	3368 (97%)	100 (3%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	8	ALA

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	201/273 (74%)	199 (99%)	2 (1%)	82	91
1	B	196/273 (72%)	193 (98%)	3 (2%)	72	88
2	C	939/941 (100%)	913 (97%)	26 (3%)	51	80
3	D	1255/1279 (98%)	1224 (98%)	31 (2%)	55	82
4	E	82/88 (93%)	80 (98%)	2 (2%)	57	83
5	F	291/388 (75%)	281 (97%)	10 (3%)	44	77
All	All	2964/3242 (91%)	2890 (98%)	74 (2%)	55	82

5 of 74 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	145	VAL
3	D	687	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
5	F	205	ARG
3	D	231	VAL
3	D	544	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	669	ASN
3	D	717	GLN
3	D	1172	HIS
2	C	1026	GLN
3	D	66	GLN

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
7	I	6/7 (85%)	1 (16%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
7	I	2	C

There are no RNA pucker outliers to report.

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

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

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, 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	231/315 (73%)	0.11	13 (5%) 28 22	74, 114, 142, 189	0
1	B	223/315 (70%)	-0.10	2 (0%) 85 82	66, 99, 137, 157	0
2	C	1112/1119 (99%)	0.21	47 (4%) 40 33	55, 114, 176, 217	0
3	D	1486/1524 (97%)	0.24	87 (5%) 26 20	50, 97, 174, 205	1 (0%)
4	E	94/99 (94%)	0.21	4 (4%) 39 32	76, 124, 176, 182	0
5	F	335/443 (75%)	0.54	37 (11%) 7 6	88, 138, 224, 236	0
6	G	18/21 (85%)	-0.15	0 100 100	73, 104, 196, 202	0
7	I	7/7 (100%)	-0.39	0 100 100	69, 76, 122, 132	0
8	H	25/28 (89%)	0.06	1 (4%) 42 34	102, 124, 181, 209	0
All	All	3531/3871 (91%)	0.22	191 (5%) 29 24	50, 110, 179, 236	1 (0%)

The worst 5 of 191 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	766	GLU	7.0
5	F	381	HIS	6.8
3	D	409	VAL	6.7
2	C	63	GLY	6.5
5	F	390	PHE	6.3

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. 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, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
9	MG	D	2004	1/1	0.96	0.50	3.88	65,65,65,65	0
10	ZN	D	2001	1/1	0.99	0.14	-0.53	73,73,73,73	0
10	ZN	D	2002	1/1	0.94	0.09	-1.73	163,163,163,163	0
9	MG	F	2001	1/1	0.91	0.04	-3.63	135,135,135,135	0
9	MG	D	2003	1/1	0.90	0.21	-	53,53,53,53	0
9	MG	B	2001	1/1	0.58	0.35	-	118,118,118,118	0

6.5 Other polymers

There are no such residues in this entry.