



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 08:40 PM GMT

PDB ID : 4XLS
Title : Crystal structure of T. aquaticus transcription initiation complex with CarD containing upstream fork promoter.
Authors : Bae, B.; Darst, S.A.
Deposited on : 2015-01-13
Resolution : 4.01 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : unknown
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026982
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-20026982

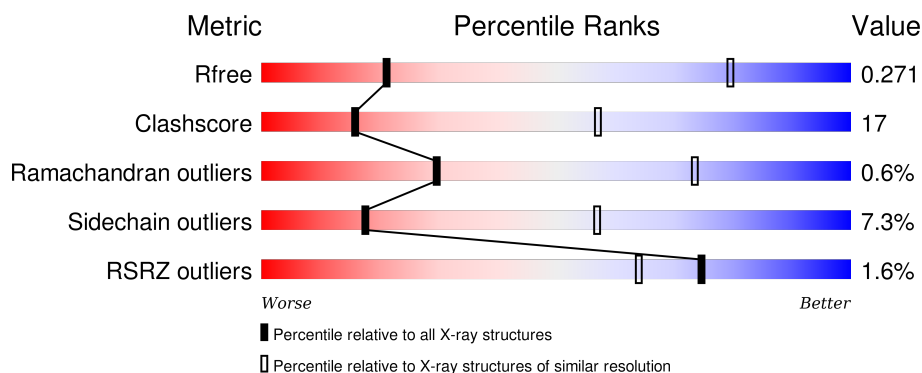
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.01 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	1015 (4.48-3.56)
Clashscore	102246	1054 (4.44-3.60)
Ramachandran outliers	100387	1006 (4.42-3.60)
Sidechain outliers	100360	1016 (4.46-3.58)
RSRZ outliers	91569	1018 (4.48-3.56)

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 $\leq 5\%$. 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	A	314	
1	B	314	
1	G	314	
1	H	314	
2	C	1119	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	1119	
3	D	1524	
3	J	1524	
4	E	99	
4	K	99	
5	F	347	
5	L	347	
6	M	164	
6	N	164	
7	O	30	
7	R	30	
8	P	24	
8	S	24	

2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 58966 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	227	Total	C	N	O	S	0	0	0
			1770	1130	303	334	3			
1	B	227	Total	C	N	O	S	0	0	0
			1770	1130	303	334	3			
1	G	227	Total	C	N	O	S	0	0	0
			1770	1130	303	334	3			
1	H	227	Total	C	N	O	S	0	0	0
			1770	1130	303	334	3			

- 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	0	0
			8739	5531	1553	1632	23			
2	I	1112	Total	C	N	O	S	0	0	0
			8739	5531	1553	1632	23			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1490	Total	C	N	O	S	0	0	0
			11761	7439	2088	2196	38			
3	J	1367	Total	C	N	O	S	0	0	0
			10779	6810	1923	2010	36			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	93	Total	C	N	O	S	0	0	0
			768	490	136	138	4			
4	K	93	Total	C	N	O	S	0	0	0
			768	490	136	138	4			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	345	Total	C	N	O	S	0	0	0
			2787	1758	502	523	4			
5	L	345	Total	C	N	O	S	0	0	0
			2787	1758	502	523	4			

- Molecule 6 is a protein called CarD-like transcriptional regulator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	M	162	Total	C	N	O	S	0	0	0
			1274	807	234	231	2			
6	N	162	Total	C	N	O	S	0	0	0
			1274	807	234	231	2			

- Molecule 7 is a DNA chain called DNA (30-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	O	30	Total	C	N	O	P	0	0	0
			613	296	109	179	29			
7	R	30	Total	C	N	O	P	0	0	0
			613	296	109	179	29			

- Molecule 8 is a DNA chain called DNA (5'-D(P*GP*CP*AP*CP*AP*AP*TP*TP*TP*AP*AP*CP*AP*CP*TP*TP*TP*TP*GP*TP*CP*AP*AP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	P	24	Total	C	N	O	P	0	0	0
			489	235	86	144	24			
8	S	24	Total	C	N	O	P	0	0	0
			489	235	86	144	24			

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

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

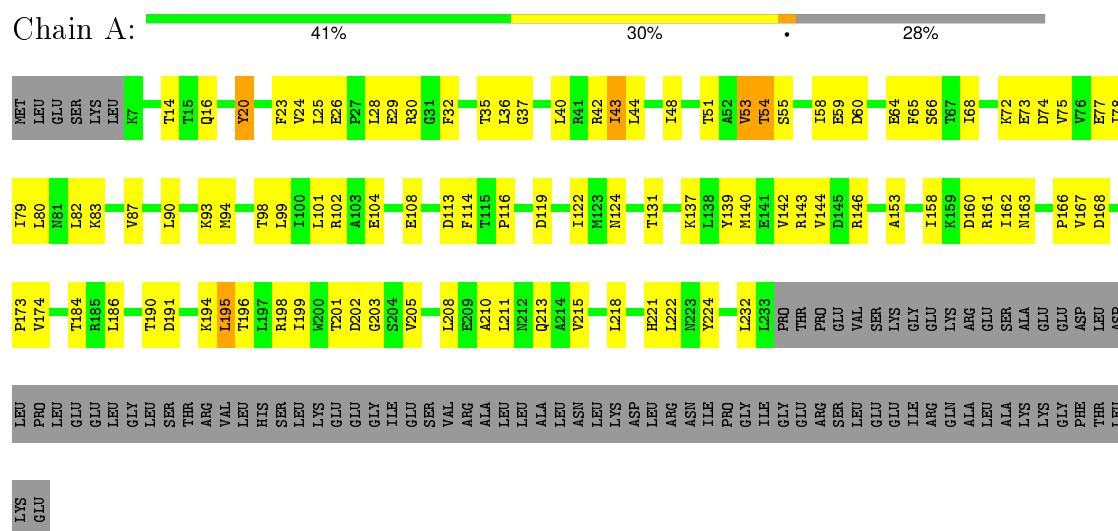
- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	J	1	Total 1	Mg 1	0	0
10	D	1	Total 1	Mg 1	0	0

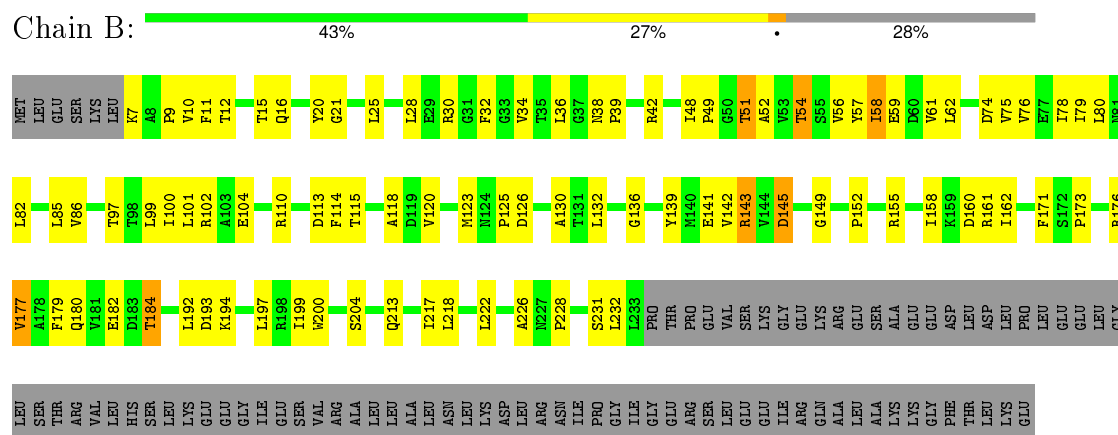
3 Residue-property plots

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

• Molecule 1: DNA-directed RNA polymerase subunit alpha

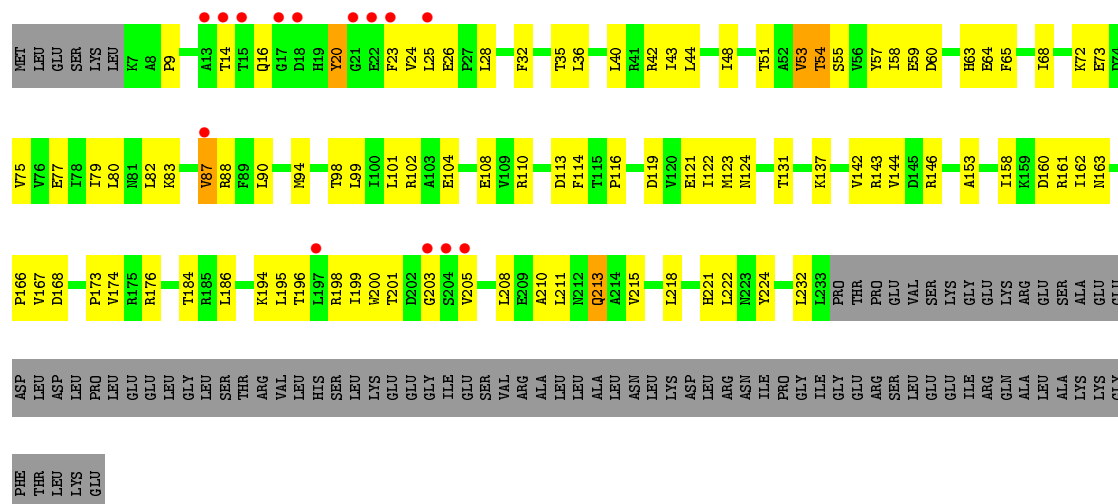


• Molecule 1: DNA-directed RNA polymerase subunit alpha

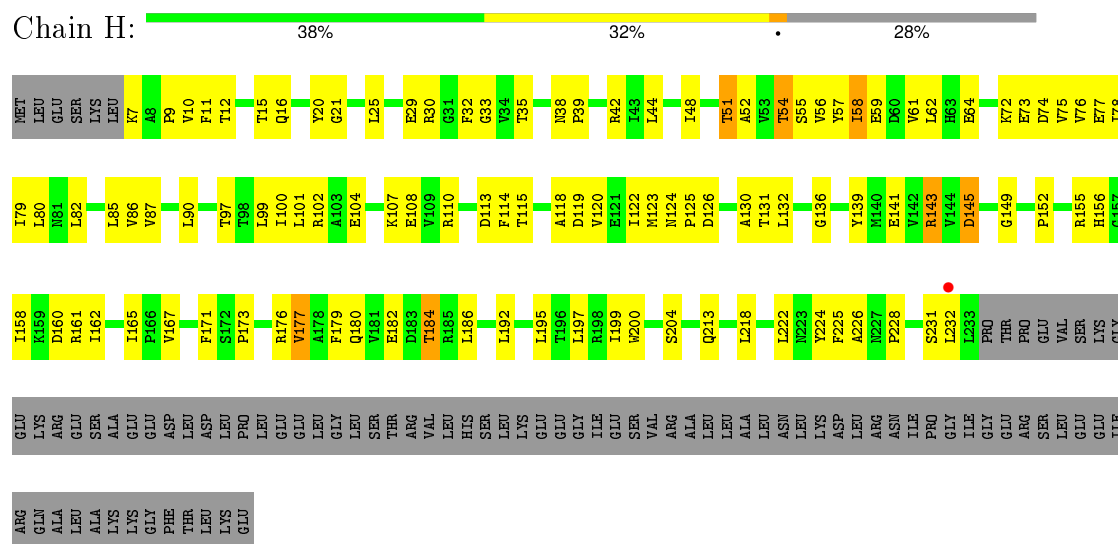


• Molecule 1: DNA-directed RNA polymerase subunit alpha

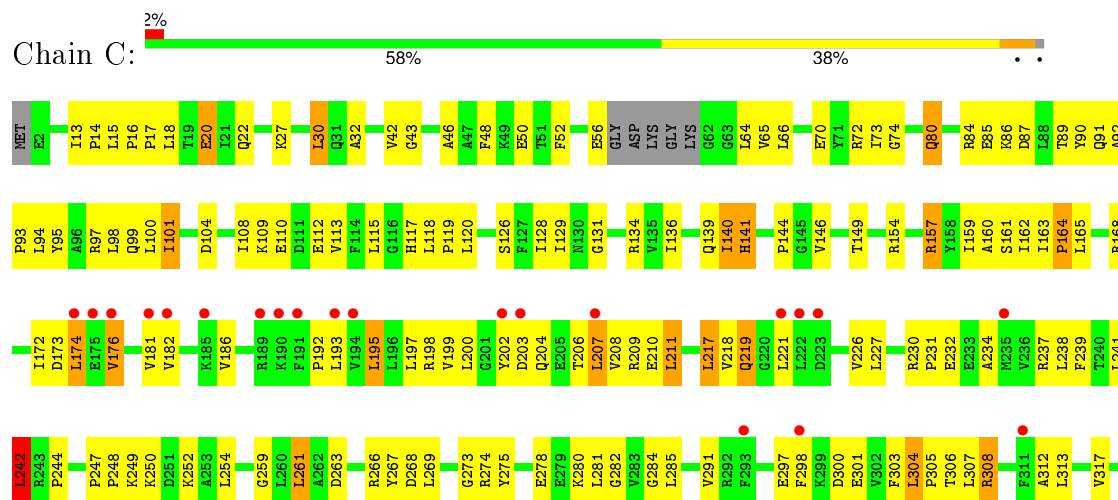


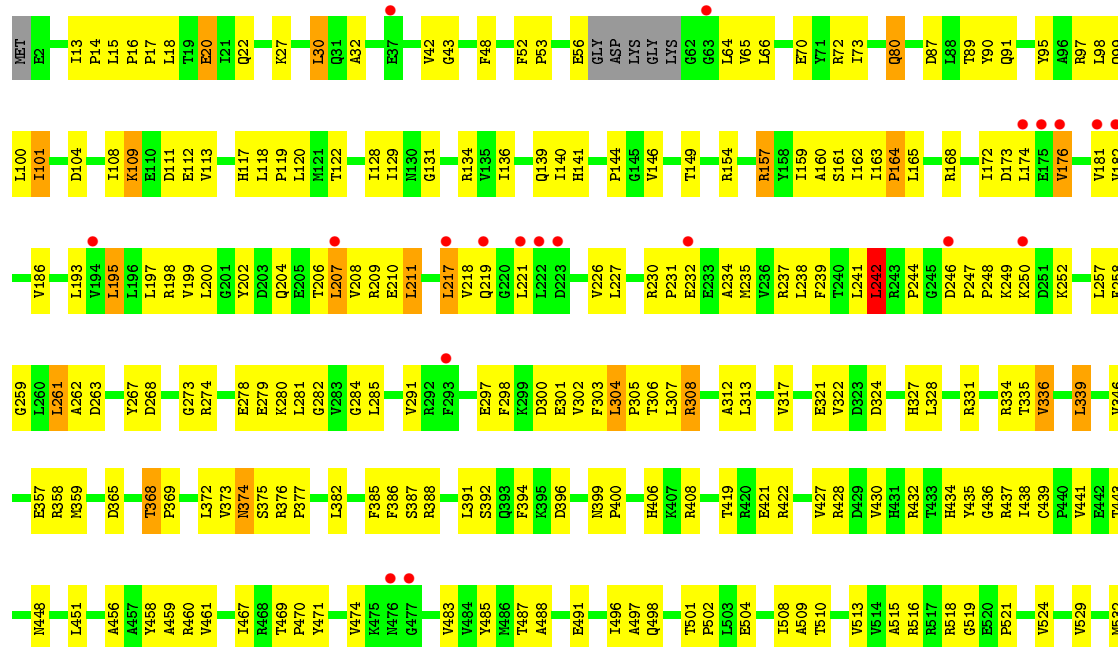


• Molecule 1: DNA-directed RNA polymerase subunit alpha



• Molecule 2: DNA-directed RNA polymerase subunit beta



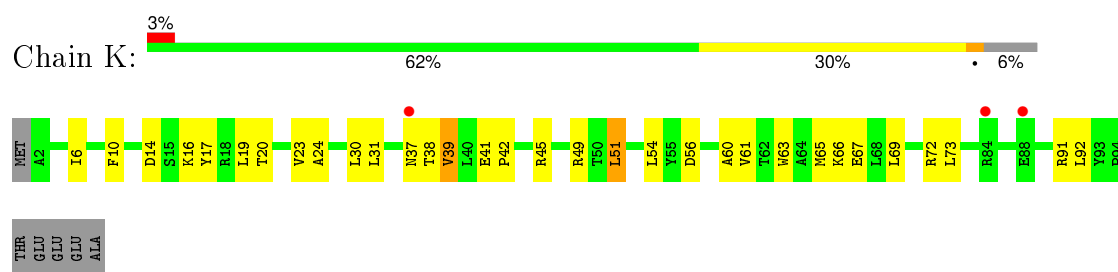




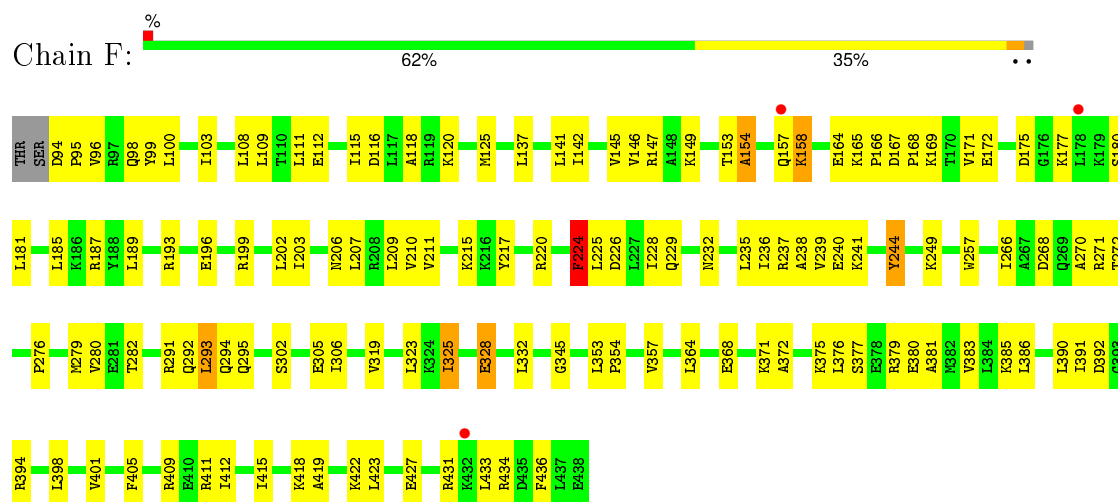




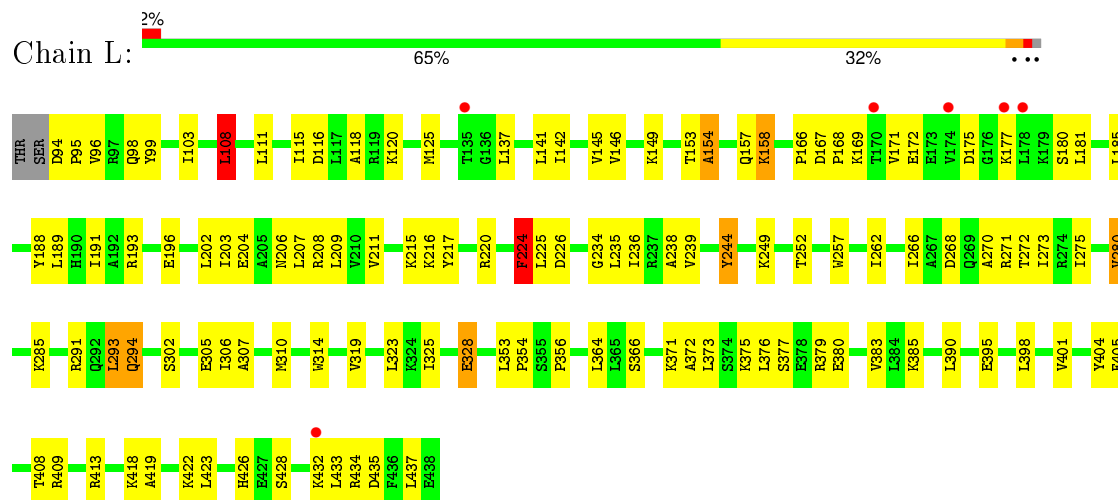
WORLDWIDE
PDB
PROTEIN DATA BANK



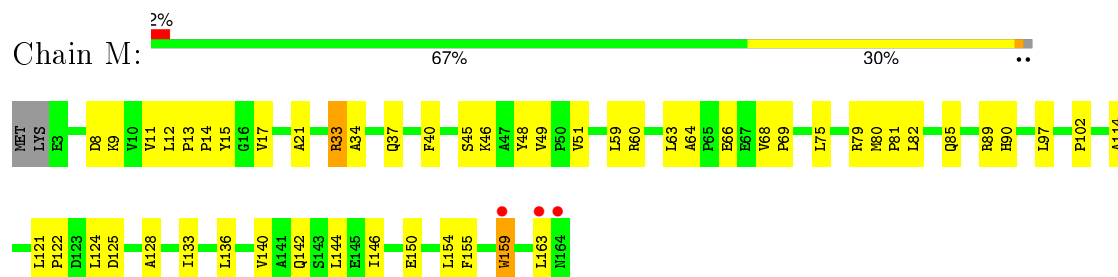
- Molecule 5: RNA polymerase sigma factor SigA



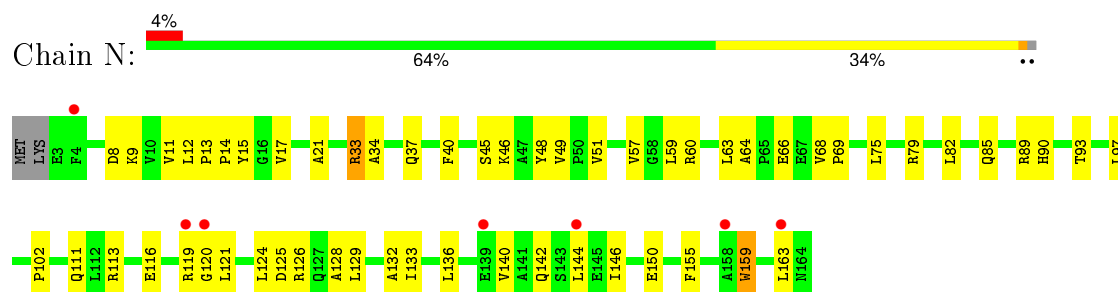
- Molecule 5: RNA polymerase sigma factor SigA



- Molecule 6: CarD-like transcriptional regulator



- Molecule 6: CarD-like transcriptional regulator



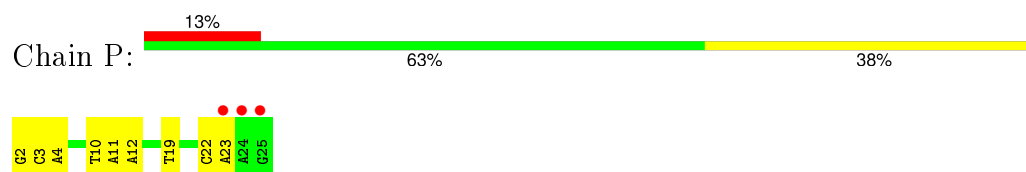
- Molecule 7: DNA (30-MER)



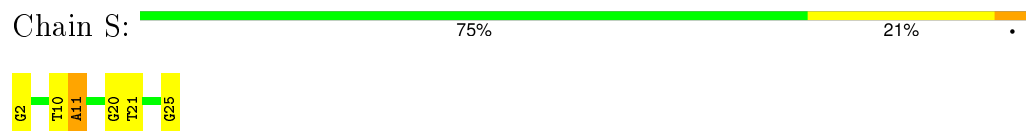
- Molecule 7: DNA (30-MER)



- Molecule 8: DNA (5'-D(P*GP*CP*AP*CP*AP*AP*TP*TP*TP*AP*AP*CP*AP*CP*TP*T
P*TP*TP*GP*TP*CP*AP*AP*G)-3')



- Molecule 8: DNA (5'-D(P*GP*CP*AP*CP*AP*AP*TP*TP*TP*AP*AP*CP*AP*CP*TP*T
P*TP*TP*GP*TP*CP*AP*AP*G)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	293.15Å 293.15Å 539.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.79 – 4.01 49.86 – 4.01	Depositor EDS
% Data completeness (in resolution range)	75.7 (49.79-4.01) 75.7 (49.86-4.01)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.79 (at 4.00Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: DEV_1839)	Depositor
R, R_{free}	0.232 , 0.272 0.228 , 0.271	Depositor DCC
R_{free} test set	7465 reflections (5.04%)	DCC
Wilson B-factor (Å ²)	146.9	Xtriage
Anisotropy	0.030	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 191.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.36$, $\langle L^2 \rangle = 0.20$	Xtriage
Outliers	1 of 148260 reflections (0.001%)	Xtriage
F_o, F_c correlation	0.87	EDS
Total number of atoms	58966	wwPDB-VP
Average B, all atoms (Å ²)	173.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.80% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: 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	0.27	0/1804	0.52	0/2455
1	B	0.27	0/1804	0.54	0/2455
1	G	0.27	0/1804	0.52	0/2455
1	H	0.28	0/1804	0.55	0/2455
2	C	0.26	0/8905	0.53	1/12040 (0.0%)
2	I	0.26	0/8905	0.53	1/12040 (0.0%)
3	D	0.27	0/11963	0.52	0/16165
3	J	0.27	0/10959	0.51	0/14802
4	E	0.25	0/783	0.52	0/1054
4	K	0.25	0/783	0.53	0/1054
5	F	0.27	0/2829	0.52	0/3804
5	L	0.27	0/2829	0.53	0/3804
6	M	0.26	0/1302	0.51	0/1765
6	N	0.26	0/1302	0.49	0/1765
7	O	0.46	0/687	1.08	0/1059
7	R	0.46	0/687	1.11	1/1059 (0.1%)
8	P	0.45	0/547	1.13	3/841 (0.4%)
8	S	0.48	0/547	1.18	2/841 (0.2%)
All	All	0.28	0/60244	0.56	8/81913 (0.0%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	S	25	DG	P-O5'-C5'	-6.14	111.08	120.90
2	I	242	LEU	CA-CB-CG	5.95	128.99	115.30
2	C	242	LEU	CA-CB-CG	5.89	128.85	115.30
7	R	28	DA	O4'-C1'-N9	5.80	112.06	108.00
8	P	19	DT	C5-C4-O4	-5.26	121.22	124.90

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	1770	0	1799	75	0
1	B	1770	0	1799	61	0
1	G	1770	0	1799	78	0
1	H	1770	0	1799	74	0
2	C	8739	0	8841	357	0
2	I	8739	0	8841	363	0
3	D	11761	0	11976	449	0
3	J	10779	0	10993	397	0
4	E	768	0	784	27	0
4	K	768	0	784	27	0
5	F	2787	0	2866	86	0
5	L	2787	0	2866	79	0
6	M	1274	0	1288	34	0
6	N	1274	0	1288	39	0
7	O	613	0	343	14	0
7	R	613	0	343	9	0
8	P	489	0	273	7	0
8	S	489	0	273	5	0
9	D	2	0	0	0	0
9	J	2	0	0	0	0
10	D	1	0	0	0	0
10	J	1	0	0	0	0
All	All	58966	0	58955	2002	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 17.

The worst 5 of 2002 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)
5:F:149:LYS:HB3	5:F:193:ARG:HH12	1.35	0.91

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:I:204:GLN:HB2	2:I:227:LEU:HD21	1.54	0.90
5:L:149:LYS:HB3	5:L:193:ARG:HH12	1.37	0.88
2:C:502:PRO:HG3	2:C:510:THR:HG22	1.56	0.88
4:K:30:LEU:HD12	4:K:37:ASN:HD21	1.38	0.88

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	225/314 (72%)	198 (88%)	26 (12%)	1 (0%)	39	79
1	B	225/314 (72%)	199 (88%)	25 (11%)	1 (0%)	39	79
1	G	225/314 (72%)	200 (89%)	24 (11%)	1 (0%)	39	79
1	H	225/314 (72%)	199 (88%)	25 (11%)	1 (0%)	39	79
2	C	1108/1119 (99%)	980 (88%)	119 (11%)	9 (1%)	24	69
2	I	1108/1119 (99%)	977 (88%)	123 (11%)	8 (1%)	26	71
3	D	1486/1524 (98%)	1333 (90%)	143 (10%)	10 (1%)	26	71
3	J	1361/1524 (89%)	1227 (90%)	126 (9%)	8 (1%)	30	73
4	E	91/99 (92%)	79 (87%)	12 (13%)	0	100	100
4	K	91/99 (92%)	80 (88%)	11 (12%)	0	100	100
5	F	343/347 (99%)	309 (90%)	32 (9%)	2 (1%)	30	73
5	L	343/347 (99%)	309 (90%)	31 (9%)	3 (1%)	21	67
6	M	160/164 (98%)	142 (89%)	18 (11%)	0	100	100
6	N	160/164 (98%)	141 (88%)	19 (12%)	0	100	100
All	All	7151/7762 (92%)	6373 (89%)	734 (10%)	44 (1%)	30	73

5 of 44 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	53	VAL
3	D	683	ILE
3	D	1128	VAL
1	G	53	VAL
3	J	681	ARG

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	194/270 (72%)	184 (95%)	10 (5%)	29	68
1	B	194/270 (72%)	176 (91%)	18 (9%)	11	46
1	G	194/270 (72%)	184 (95%)	10 (5%)	29	68
1	H	194/270 (72%)	176 (91%)	18 (9%)	11	46
2	C	931/936 (100%)	864 (93%)	67 (7%)	18	57
2	I	931/936 (100%)	863 (93%)	68 (7%)	17	57
3	D	1252/1281 (98%)	1139 (91%)	113 (9%)	12	47
3	J	1150/1281 (90%)	1062 (92%)	88 (8%)	16	55
4	E	83/88 (94%)	81 (98%)	2 (2%)	57	82
4	K	83/88 (94%)	79 (95%)	4 (5%)	31	69
5	F	296/299 (99%)	279 (94%)	17 (6%)	25	65
5	L	296/299 (99%)	275 (93%)	21 (7%)	18	58
6	M	131/133 (98%)	127 (97%)	4 (3%)	47	78
6	N	131/133 (98%)	127 (97%)	4 (3%)	47	78
All	All	6060/6554 (92%)	5616 (93%)	444 (7%)	17	57

5 of 444 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	1456	LYS
1	H	114	PHE
3	J	1499	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	D	1497	GLU
5	F	328	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 107 such sidechains are listed below:

Mol	Chain	Res	Type
5	F	263	ASN
1	H	38	ASN
5	L	200	GLN
5	F	269	GLN
1	G	163	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

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 [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, 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	227/314 (72%)	-0.15	0 100 100	120, 182, 228, 247	0
1	B	227/314 (72%)	-0.29	0 100 100	83, 155, 210, 239	0
1	G	227/314 (72%)	0.33	14 (6%) 24 16	127, 193, 236, 261	0
1	H	227/314 (72%)	-0.14	1 (0%) 93 90	93, 165, 210, 245	0
2	C	1112/1119 (99%)	-0.09	24 (2%) 65 54	70, 174, 245, 303	0
2	I	1112/1119 (99%)	-0.04	25 (2%) 65 54	82, 182, 250, 314	0
3	D	1490/1524 (97%)	-0.17	8 (0%) 91 88	66, 147, 206, 277	0
3	J	1367/1524 (89%)	-0.15	14 (1%) 84 77	70, 153, 211, 274	0
4	E	93/99 (93%)	-0.12	0 100 100	93, 158, 217, 254	0
4	K	93/99 (93%)	0.06	3 (3%) 51 39	106, 166, 222, 247	0
5	F	345/347 (99%)	-0.10	3 (0%) 85 80	108, 179, 259, 323	0
5	L	345/347 (99%)	-0.21	6 (1%) 73 62	119, 184, 260, 323	0
6	M	162/164 (98%)	0.15	3 (1%) 70 59	154, 223, 269, 295	0
6	N	162/164 (98%)	0.47	7 (4%) 39 29	161, 228, 271, 302	0
7	O	30/30 (100%)	0.15	3 (10%) 9 7	139, 200, 269, 277	0
7	R	30/30 (100%)	-0.35	0 100 100	156, 209, 257, 267	0
8	P	24/24 (100%)	0.23	3 (12%) 5 5	162, 229, 257, 280	0
8	S	24/24 (100%)	-0.47	0 100 100	173, 222, 252, 265	0
All	All	7297/7870 (92%)	-0.09	114 (1%) 74 64	66, 168, 242, 323	0

The worst 5 of 114 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	I	221	LEU	5.8
2	I	182	VAL	5.0
1	G	13	ALA	4.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	C	175	GLU	4.6
6	N	119	ARG	4.3

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
9	ZN	J	2001	1/1	0.93	0.20	-0.25	200,200,200,200	0
9	ZN	D	2001	1/1	0.97	0.13	-1.10	96,96,96,96	0
9	ZN	D	2002	1/1	0.94	0.14	-1.11	203,203,203,203	0
9	ZN	J	2002	1/1	0.97	0.06	-1.76	242,242,242,242	0
10	MG	D	2003	1/1	0.95	0.27	-	198,198,198,198	0
10	MG	J	2003	1/1	0.88	0.29	-	219,219,219,219	0

6.5 Other polymers [i](#)

There are no such residues in this entry.