wwPDB X-ray Structure Validation Summary Report (i)

May 29, 2020 – 08:26 am BST

PDB ID	:	5IPN
Title	:	SigmaS-transcription initiation complex with 4-nt nascent RNA
Authors	:	Liu, B.; Zuo, Y.; Steitz, T.A.
Deposited on	:	2016-03-09
Resolution	:	4.61 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 following versions of software and data (see references (1)) were used in the production of this report:

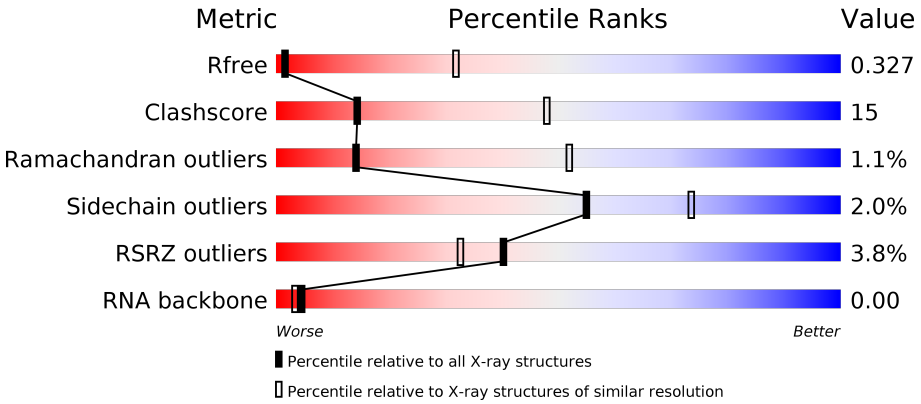
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 4.61 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.



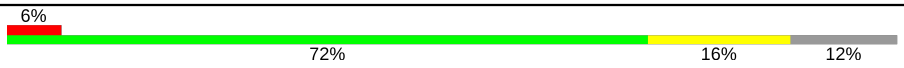
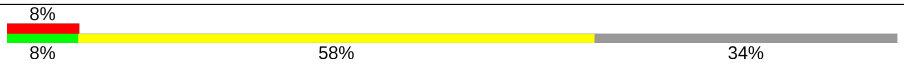
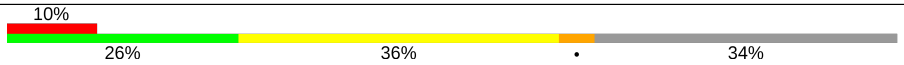
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1062 (5.40-3.80)
Clashscore	141614	1130 (5.40-3.80)
Ramachandran outliers	138981	1074 (5.40-3.80)
Sidechain outliers	138945	1055 (5.40-3.80)
RSRZ outliers	127900	1114 (5.54-3.70)
RNA backbone	3102	1064 (6.22-3.00)

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 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 <=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	242	6% 79% 16% 5%
1	B	242	4% 79% 15% 6%
2	C	1342	3% 76% 22% .
3	D	1407	3% 67% 27% . .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	90	
5	F	336	
6	1	50	
7	2	50	
8	3	4	

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	ZN	D	1502	-	-	X	-

2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 29027 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	230	Total	C	N	O	S	0	0	0
			1787	1112	317	352	6			
1	B	228	Total	C	N	O	S	0	0	0
			1767	1100	312	349	6			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	ALA	-	expression tag	UNP P0A7Z4
A	-5	HIS	-	expression tag	UNP P0A7Z4
A	-4	HIS	-	expression tag	UNP P0A7Z4
A	-3	HIS	-	expression tag	UNP P0A7Z4
A	-2	HIS	-	expression tag	UNP P0A7Z4
A	-1	HIS	-	expression tag	UNP P0A7Z4
A	0	HIS	-	expression tag	UNP P0A7Z4
B	-6	ALA	-	expression tag	UNP P0A7Z4
B	-5	HIS	-	expression tag	UNP P0A7Z4
B	-4	HIS	-	expression tag	UNP P0A7Z4
B	-3	HIS	-	expression tag	UNP P0A7Z4
B	-2	HIS	-	expression tag	UNP P0A7Z4
B	-1	HIS	-	expression tag	UNP P0A7Z4
B	0	HIS	-	expression tag	UNP P0A7Z4

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1340	Total	C	N	O	S	0	0	0
			10570	6631	1841	2055	43			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1362	Total	C	N	O	S	0	0	0
			10568	6633	1887	1998	50			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	79	Total	C	N	O	S	0	0	0
			627	382	118	126	1			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	277	Total	C	N	O	S	0	0	0
			2253	1411	415	423	4			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	2	GLY	SER	conflict	UNP P13445
F	33	GLU	GLN	conflict	UNP P13445
F	329	LEU	ARG	conflict	UNP P13445
F	331	HIS	-	expression tag	UNP P13445
F	332	HIS	-	expression tag	UNP P13445
F	333	HIS	-	expression tag	UNP P13445
F	334	HIS	-	expression tag	UNP P13445
F	335	HIS	-	expression tag	UNP P13445
F	336	HIS	-	expression tag	UNP P13445

- Molecule 6 is a DNA chain called synthetic non-template strand DNA (50-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	1	33	Total	C	N	O	P	0	0	0
			680	323	124	200	33			

- Molecule 7 is a DNA chain called synthetic template strand DNA (50-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	2	33	Total	C	N	O	P	0	0	0
			675	322	125	196	32			

- Molecule 8 is a RNA chain called Nascent RNA 4-mer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	3	4	Total	C	N	O	P	0	0	0
			97	39	17	35	6			

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

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

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

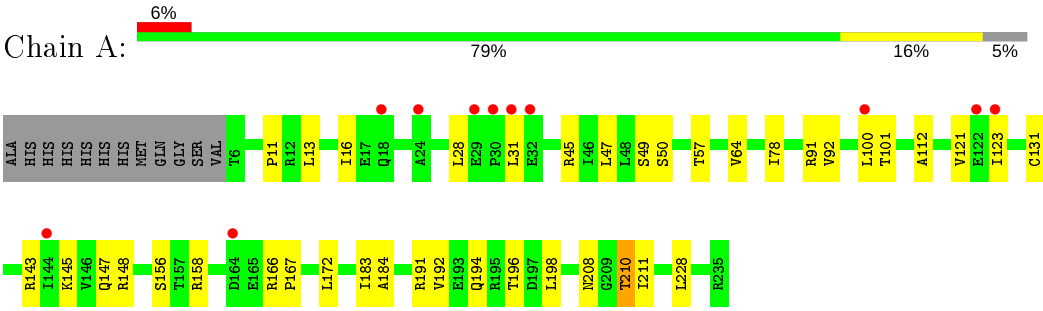
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	D	1	Total	Mg	0	0
			1	1		

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.

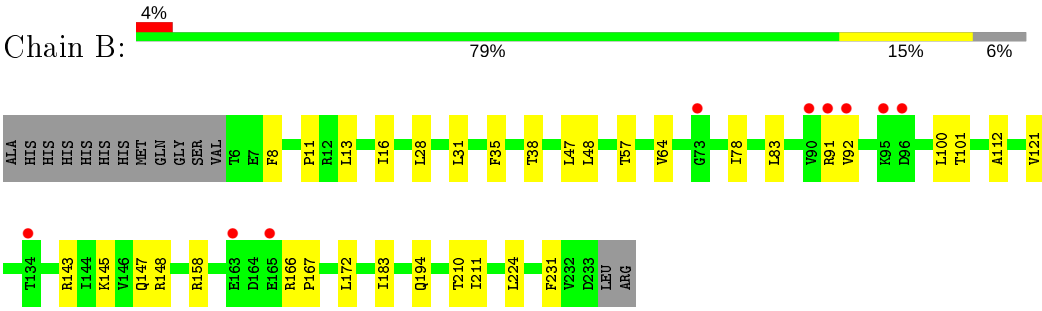
- Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



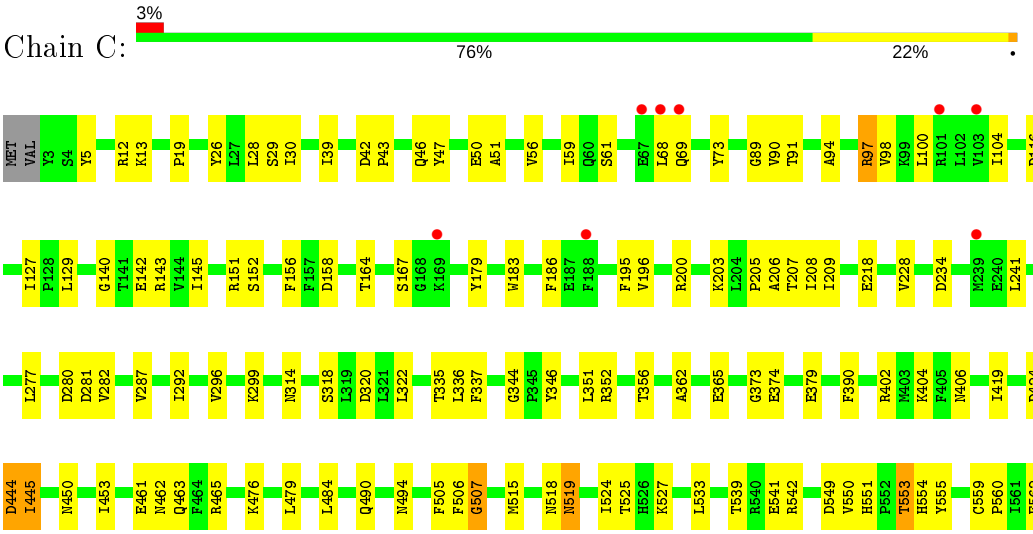
- Molecule 1: DNA-directed RNA polymerase subunit alpha

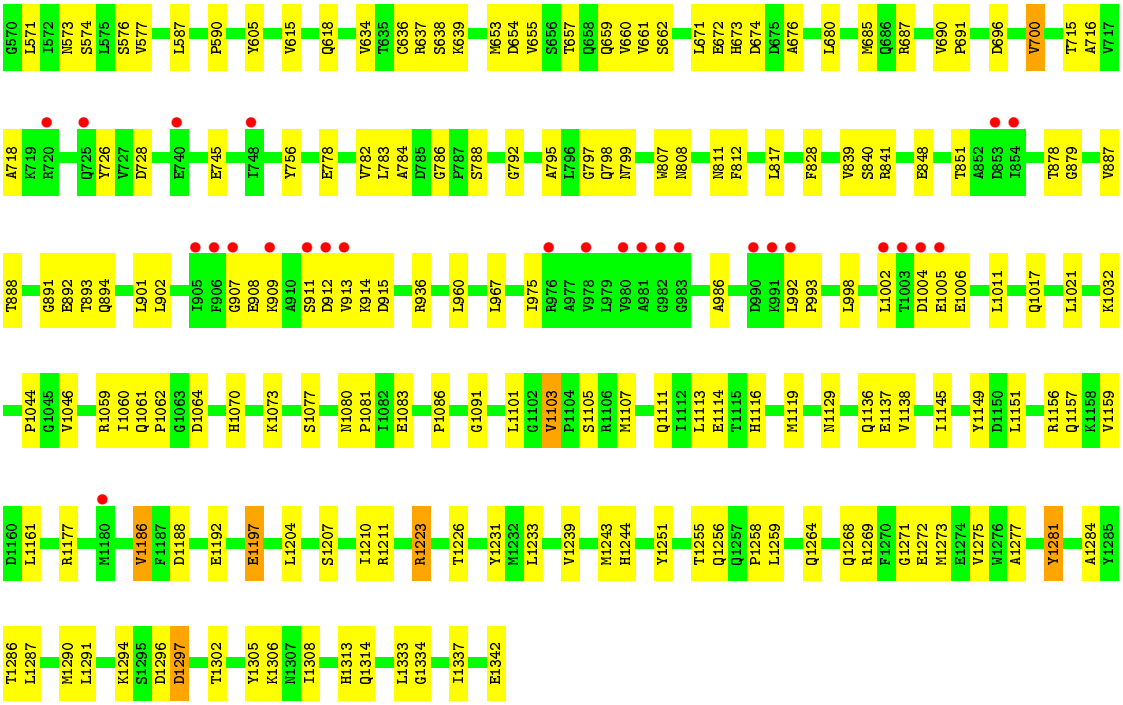
Chain B:



- Molecule 2: DNA-directed RNA polymerase subunit beta

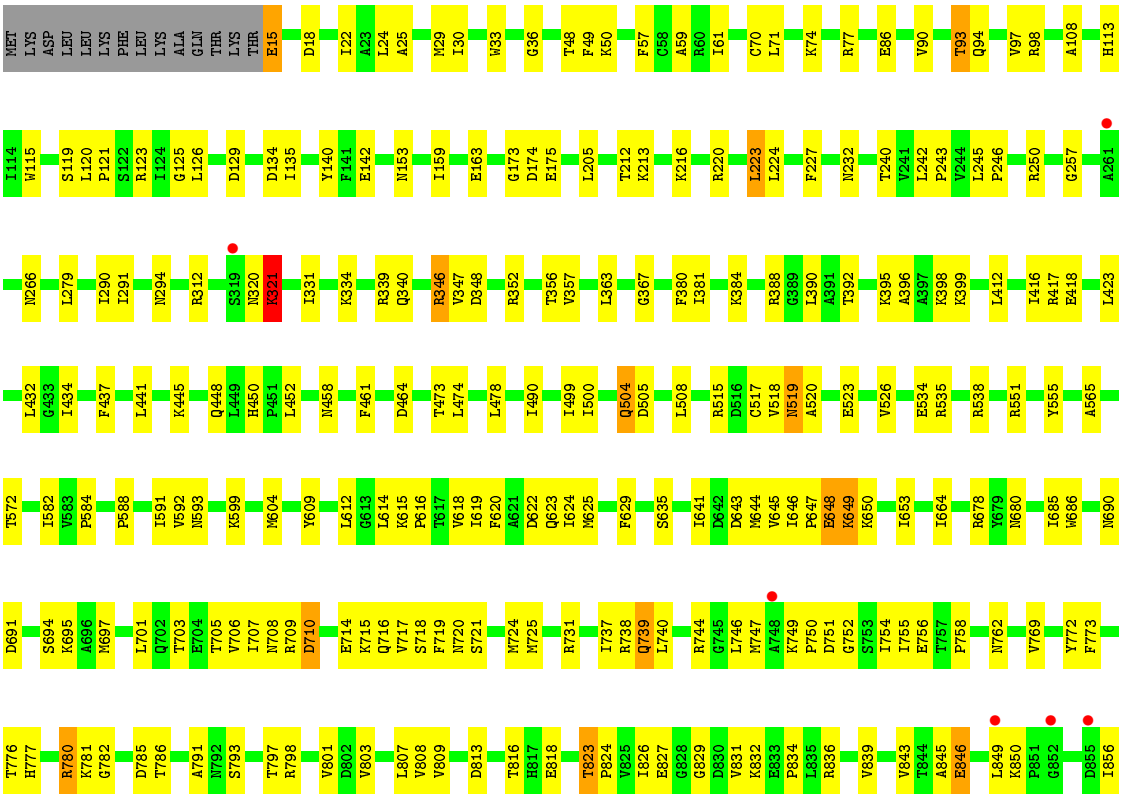
Chain C:

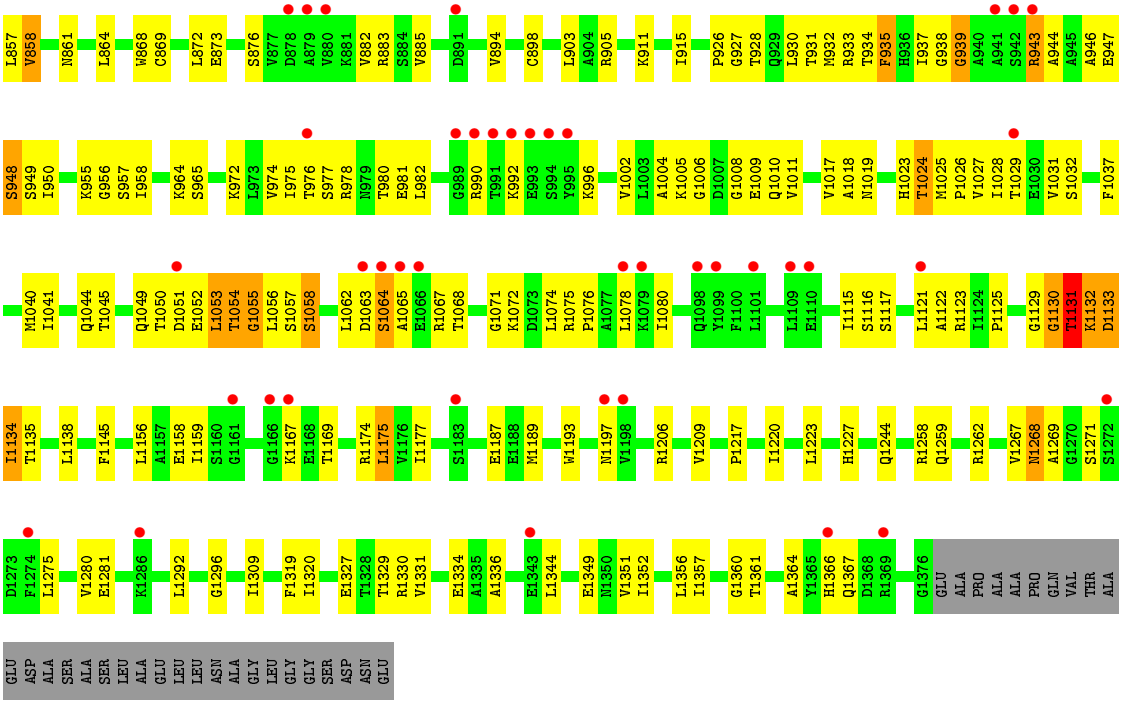




• Molecule 3: DNA-directed RNA polymerase subunit beta'

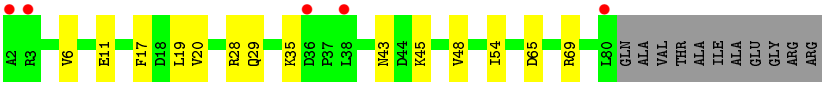
Chain D:





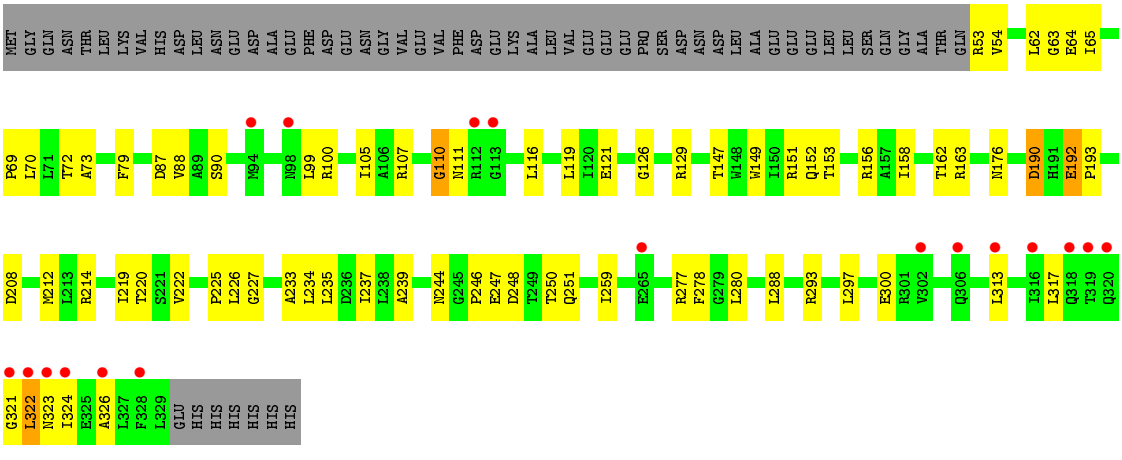
• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:



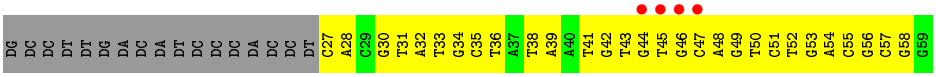
• Molecule 5: RNA polymerase sigma factor RpoS

Chain F:

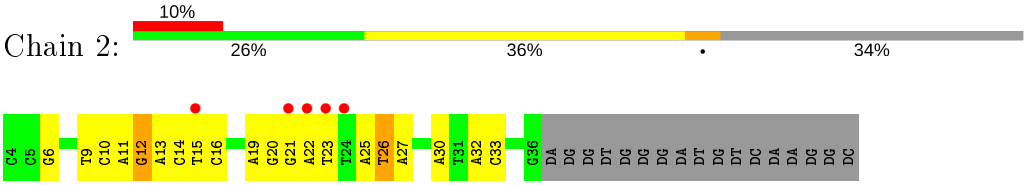


• Molecule 6: synthetic non-template strand DNA (50-MER)

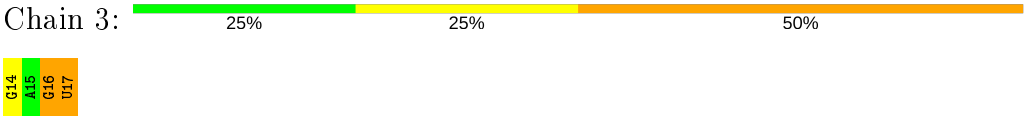
Chain 1:



• Molecule 7: synthetic template strand DNA (50-MER)



• Molecule 8: Nascent RNA 4-mer



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	132.87Å 152.17Å 229.12Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	126.76 – 4.61 50.04 – 4.61	Depositor EDS
% Data completeness (in resolution range)	98.8 (126.76-4.61) 98.9 (50.04-4.61)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.26 (at 4.64Å)	Xtriage
Refinement program	REFMAC 5.8.0123	Depositor
R, R_{free}	0.280 , 0.345 0.266 , 0.327	Depositor DCC
R_{free} test set	1233 reflections (4.74%)	wwPDB-VP
Wilson B-factor (Å^2)	192.7	Xtriage
Anisotropy	0.811	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 280.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.38$, $< L^2 > = 0.21$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	29027	wwPDB-VP
Average B, all atoms (Å^2)	321.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.98% 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: GTP, 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	0.51	0/1809	0.69	0/2450
1	B	0.47	0/1789	0.67	0/2425
2	C	0.60	6/10736 (0.1%)	0.76	6/14480 (0.0%)
3	D	0.70	12/10729 (0.1%)	0.86	8/14487 (0.1%)
4	E	0.51	0/629	0.69	0/847
5	F	0.52	0/2282	0.78	0/3076
6	1	0.33	0/762	0.59	0/1175
7	2	0.53	2/756 (0.3%)	0.66	0/1163
8	3	0.34	0/72	0.64	0/110
All	All	0.61	20/29564 (0.1%)	0.78	14/40213 (0.0%)

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	1131	THR	C-N	27.85	1.98	1.34
3	D	15	GLU	CG-CD	19.01	1.80	1.51
3	D	739	GLN	C-N	15.48	1.69	1.34
3	D	780	ARG	NE-CZ	15.25	1.52	1.33
3	D	780	ARG	CD-NE	13.89	1.70	1.46

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	1131	THR	C-N-CA	-35.47	33.02	121.70
3	D	1131	THR	CA-C-N	-26.62	58.64	117.20
3	D	1131	THR	O-C-N	-22.28	87.05	122.70
3	D	780	ARG	CD-NE-CZ	12.93	141.71	123.60
3	D	346	ARG	NE-CZ-NH2	-9.26	115.67	120.30

There are no chirality outliers.

There are no planarity outliers.

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	1787	0	1813	31	0
1	B	1767	0	1789	26	0
2	C	10570	0	10579	260	1
3	D	10568	0	10778	517	1
4	E	627	0	634	10	0
5	F	2253	0	2298	73	2
6	1	680	0	373	66	1
7	2	675	0	374	34	0
8	3	97	0	44	7	0
9	D	2	0	0	2	0
10	D	1	0	0	0	0
All	All	29027	0	28682	891	3

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 15.

The worst 5 of 891 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:750:PRO:HA	3:D:781:LYS:CD	1.25	1.65
3:D:785:ASP:HB3	3:D:935:PHE:CZ	1.24	1.62
3:D:780:ARG:NE	3:D:780:ARG:CD	1.70	1.55
3:D:15:GLU:CG	3:D:15:GLU:CD	1.80	1.49
3:D:739:GLN:C	3:D:740:LEU:N	1.69	1.43

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:F:69:PRO:CB	5:F:300:GLU:OE2[3_454]	2.01	0.19
2:C:379:GLU:N	5:F:293:ARG:NH1[3_454]	2.12	0.08

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:212:THR:OG1	6:1:27:DC:OP1[3_454]	2.19	0.01

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	228/242 (94%)	216 (95%)	10 (4%)	2 (1%)	17	56
1	B	226/242 (93%)	213 (94%)	13 (6%)	0	100	100
2	C	1332/1342 (99%)	1244 (93%)	78 (6%)	10 (1%)	19	60
3	D	1360/1407 (97%)	1249 (92%)	90 (7%)	21 (2%)	10	46
4	E	77/90 (86%)	74 (96%)	3 (4%)	0	100	100
5	F	275/336 (82%)	257 (94%)	14 (5%)	4 (2%)	10	46
All	All	3498/3659 (96%)	3253 (93%)	208 (6%)	37 (1%)	14	52

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	756	TYR
2	C	908	GLU
3	D	174	ASP
3	D	519	ASN
3	D	1024	THR

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	198/208 (95%)	195 (98%)	3 (2%)	65 80
1	B	196/208 (94%)	192 (98%)	4 (2%)	55 73
2	C	1155/1157 (100%)	1144 (99%)	11 (1%)	76 86
3	D	1135/1168 (97%)	1101 (97%)	34 (3%)	41 63
4	E	67/74 (90%)	64 (96%)	3 (4%)	27 53
5	F	240/292 (82%)	236 (98%)	4 (2%)	60 78
All	All	2991/3107 (96%)	2932 (98%)	59 (2%)	55 73

5 of 59 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	321	LYS
3	D	538	ARG
4	E	45	LYS
3	D	384	LYS
3	D	399	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	458	ASN
3	D	690	ASN
3	D	1244	GLN
3	D	545	HIS
3	D	708	ASN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	3	2/4 (50%)	2 (100%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	3	16	G
8	3	17	U

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 3 ligands modelled in this entry, 3 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

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	C	3
3	D	2
7	2	1

The worst 5 of 6 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	336:LEU	C	337:PHE	N	2.47
1	C	911:SER	C	912:ASP	N	2.42
1	C	891:GLY	C	892:GLU	N	2.39
1	2	22:DA	O3'	23:DT	P	2.27
1	D	1131:THR	C	1132:LYS	N	1.98

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	230/242 (95%)	0.18	14 (6%) 21 18	205, 326, 406, 454	0
1	B	228/242 (94%)	-0.01	9 (3%) 39 32	204, 345, 436, 567	0
2	C	1340/1342 (99%)	-0.13	35 (2%) 56 46	153, 275, 464, 598	0
3	D	1362/1407 (96%)	-0.03	47 (3%) 44 36	148, 303, 466, 628	0
4	E	79/90 (87%)	0.36	5 (6%) 20 16	279, 369, 570, 678	0
5	F	277/336 (82%)	0.36	18 (6%) 18 15	225, 392, 525, 615	0
6	1	33/50 (66%)	0.06	4 (12%) 4 5	294, 392, 503, 560	0
7	2	33/50 (66%)	0.37	5 (15%) 2 3	243, 382, 500, 581	0
8	3	3/4 (75%)	0.73	0 100 100	283, 283, 336, 352	0
All	All	3585/3763 (95%)	-0.01	137 (3%) 40 32	148, 310, 478, 678	0

The worst 5 of 137 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	1004	ASP	10.1
2	C	1003	THR	7.7
5	F	323	ASN	6.8
5	F	318	GLN	6.0
3	D	991	THR	5.8

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, 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	B-factors(Å^2)	Q<0.9
9	ZN	D	1502	1/1	0.93	0.12	287,287,287,287	0
9	ZN	D	1501	1/1	0.97	0.13	339,339,339,339	0
10	MG	D	1503	1/1	0.98	0.15	251,251,251,251	0

6.5 Other polymers

There are no such residues in this entry.