wwPDB X-ray Structure Validation Summary Report (i)

Feb 25, 2024 – 12:54 PM EST

PDB ID	:	5IPL
Title	:	SigmaS-transcription initiation complex with 4-nt nascent RNA
Authors	:	Liu, B.; Zuo, Y.; Steitz, T.A.
Deposited on	:	2016-03-09
Resolution	:	3.60 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:

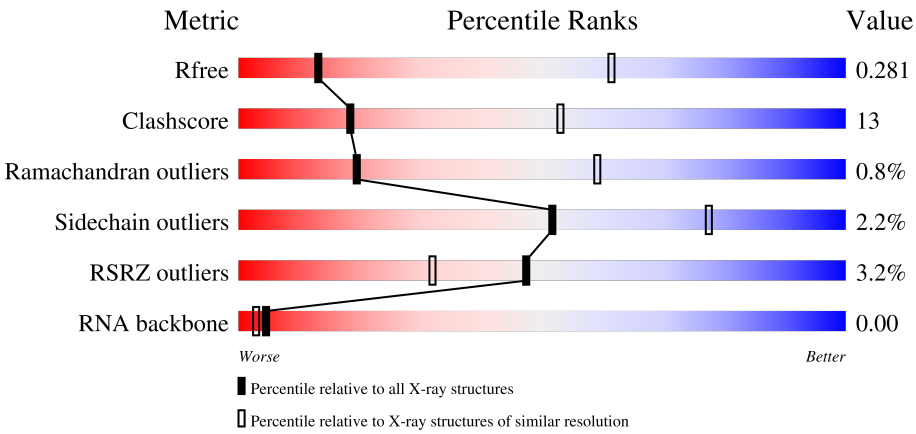
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.60 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	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)
RNA backbone	3102	1017 (4.20-3.00)

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	242	.% 74% 20% 5%
1	B	242	8% 76% 18% 6%
2	C	1342	2% 75% 23% •
3	D	1407	3% 61% 21% • 17%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	E	90	2% 71% 16% 12%
5	F	336	2% 52% 29% 18%
6	1	50	2% 10% 54% 34%
7	2	50	26% 40% 34%
8	3	4	75% 25%

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
10	ZN	D	1502	-	-	X	-
11	MG	D	1504	-	-	-	X
9	DPO	C	1401	-	X	-	-

2 Entry composition (i)

There are 11 unique types of molecules in this entry. The entry contains 27632 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	230	1787	1112	317	352	6	0	0	0
1	B	228	1767	1100	312	349	6	0	0	0

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
			Total	C	N	O	S			
2	C	1340	10570	6631	1841	2055	43	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	1173	9163	5760	1644	1712	47	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	79	627	382	118	126	1	0	0	0

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

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

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 nontemplate strand DNA (50-MER).

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

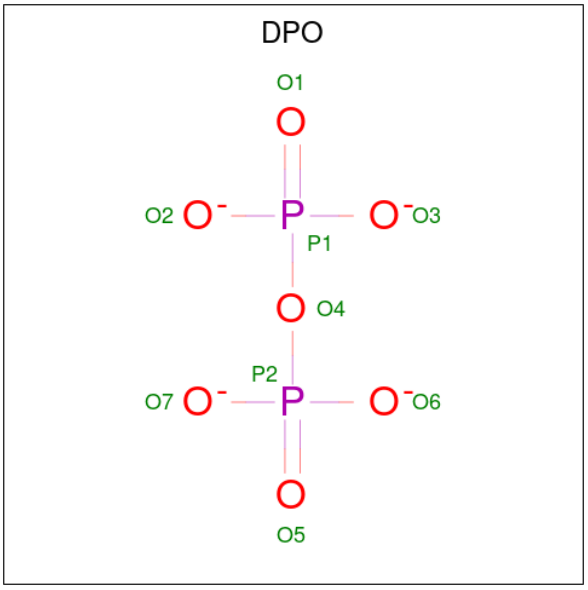
- Molecule 7 is a DNA chain called synthetic template strand DNA (50-MER).

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

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

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

- Molecule 9 is DIPHOSPHATE (three-letter code: DPO) (formula: O_7P_2).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	P		
9	C	1	9	7	2	0	0

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

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

- Molecule 11 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

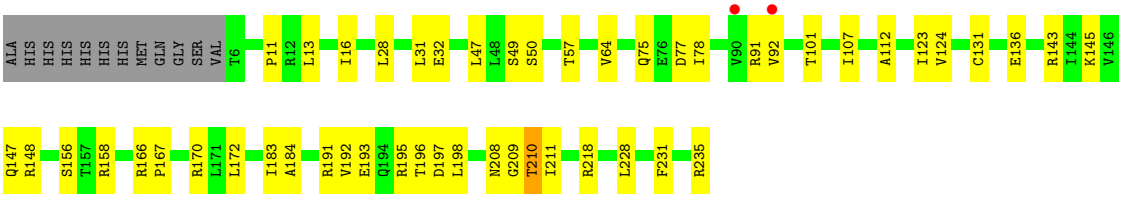
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
11	D	2	2	2	0	0

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.

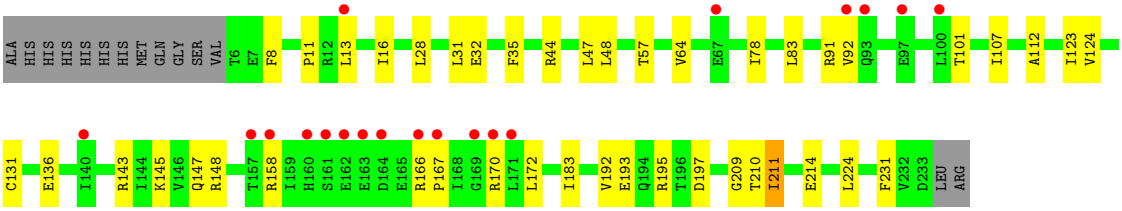
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



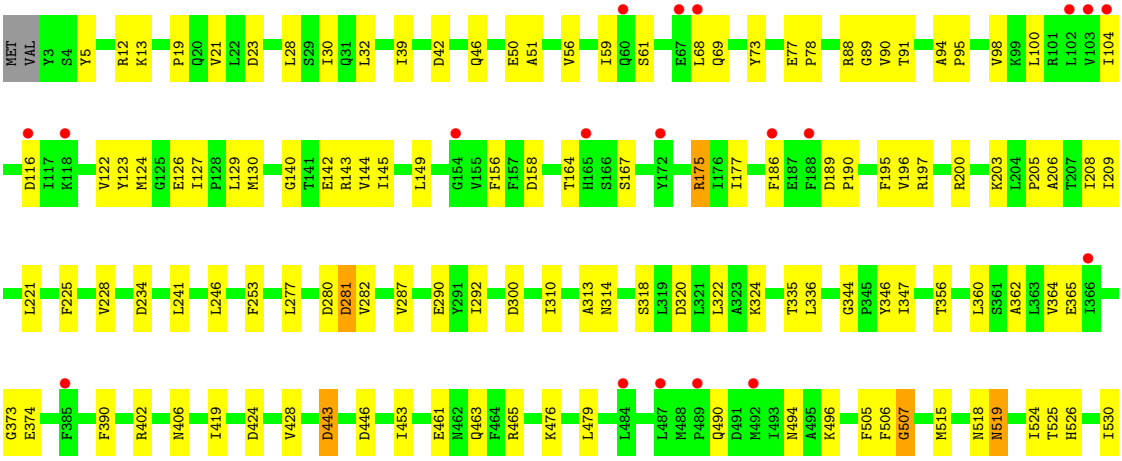
• Molecule 1: DNA-directed RNA polymerase subunit alpha

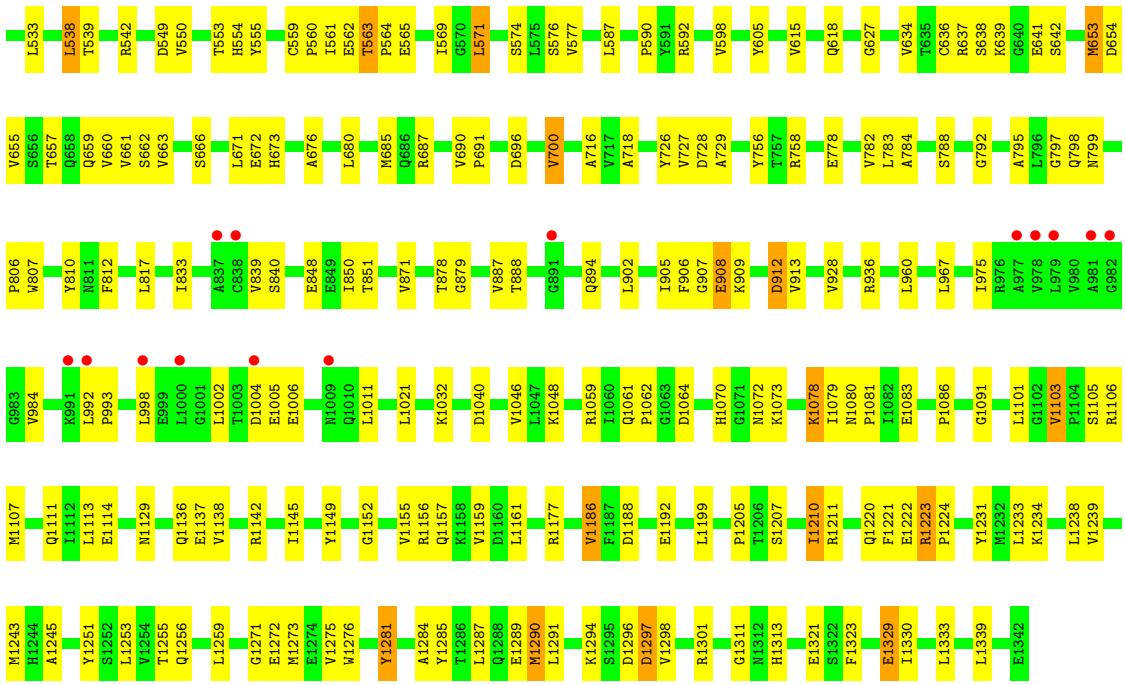
Chain B:



• Molecule 2: DNA-directed RNA polymerase subunit beta

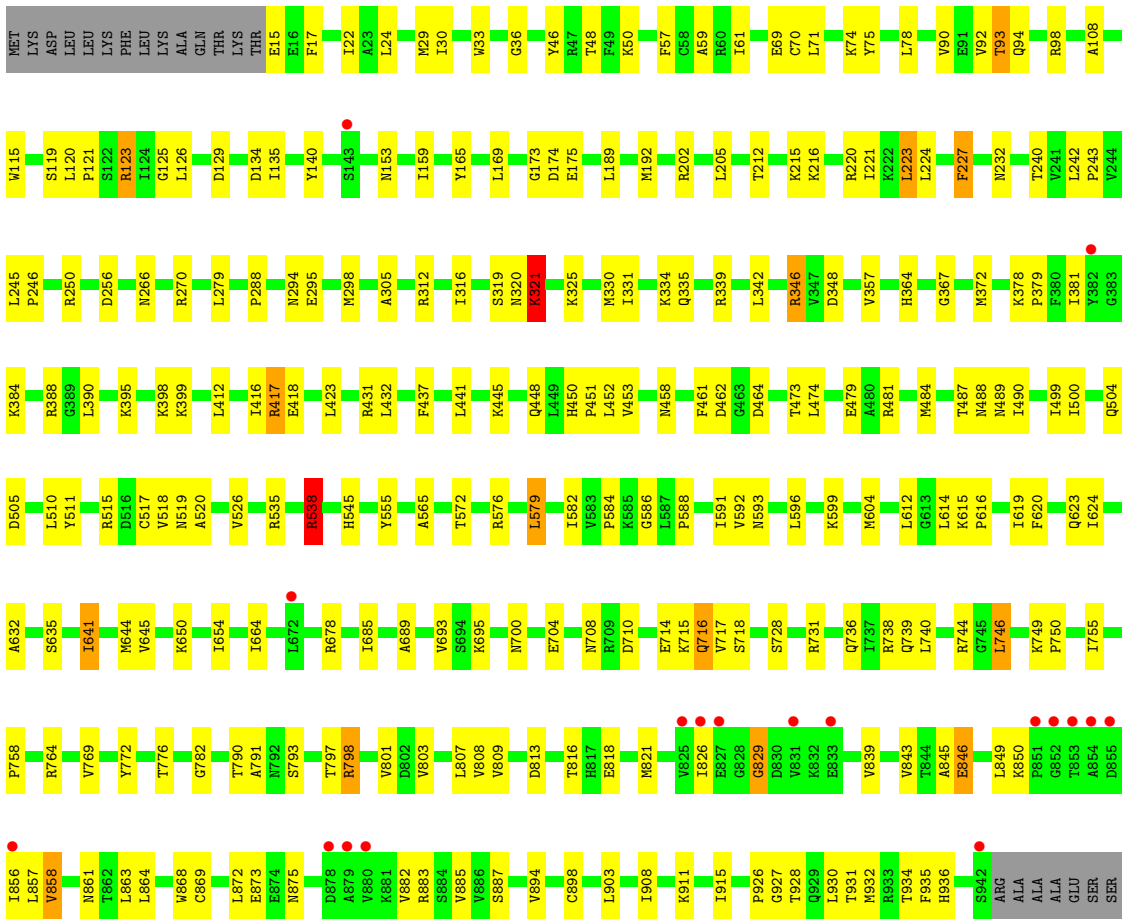
Chain C:

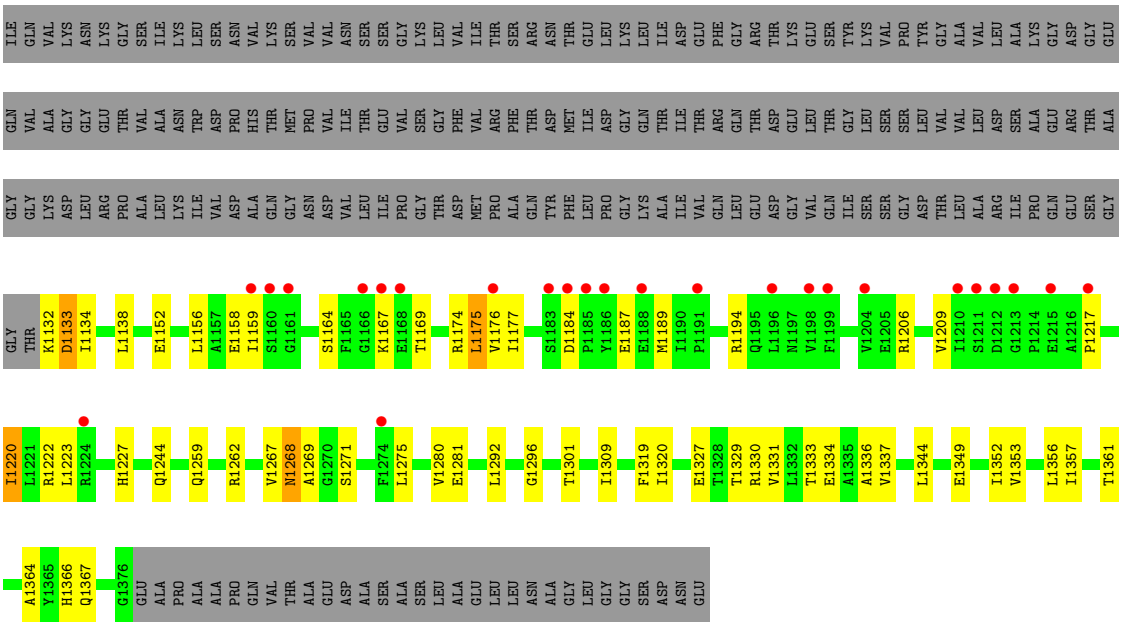




- Molecule 3: DNA-directed RNA polymerase subunit beta'

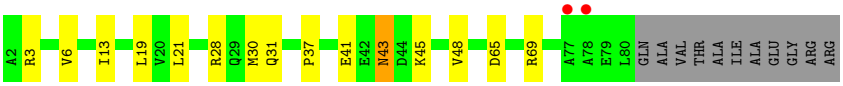
Chain D: 3% 61% 21% • 17%





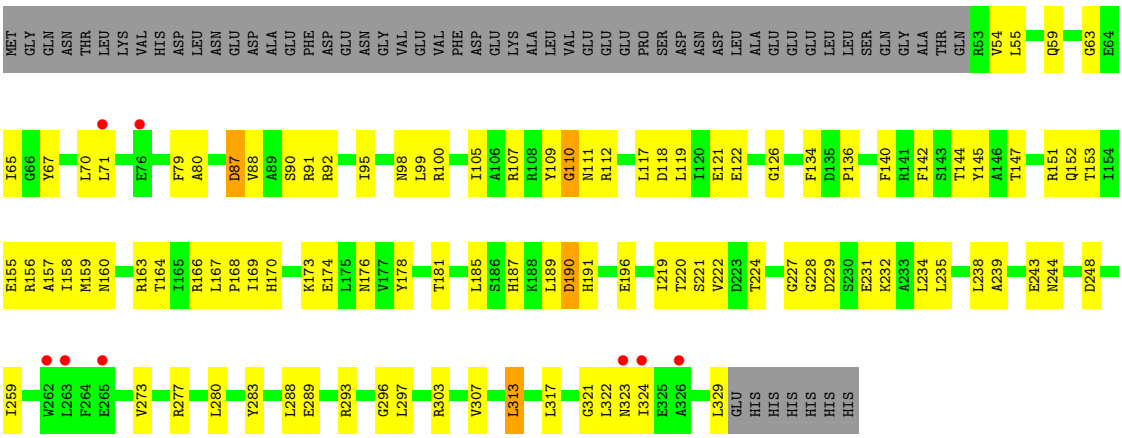
- Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:



- Molecule 5: RNA polymerase sigma factor RpoS

Chain F:

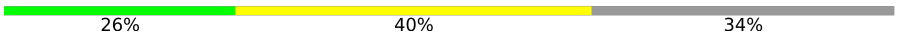


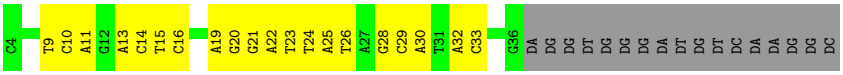
- Molecule 6: synthetic nontemplate strand DNA (50-MER)

Chain 1:

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

Chain 2:





- Molecule 8: nascent RNA 4-mer

Chain 3:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	131.71Å 152.67Å 226.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	126.63 – 3.60 49.73 – 3.60	Depositor EDS
% Data completeness (in resolution range)	99.6 (126.63-3.60) 99.7 (49.73-3.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.14 (at 3.57Å)	Xtriage
Refinement program	REFMAC 5.8.0123	Depositor
R, R_{free}	0.247 , 0.293 0.241 , 0.281	Depositor DCC
R_{free} test set	2560 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å^2)	143.0	Xtriage
Anisotropy	0.464	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 141.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.23$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	27632	wwPDB-VP
Average B, all atoms (Å^2)	206.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.12% 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: DPO, MG, ZN, GTP

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	0.54	0/1809	0.72	0/2450
1	B	0.52	0/1789	0.70	0/2425
2	C	0.64	9/10739 (0.1%)	0.79	4/14489 (0.0%)
3	D	0.59	2/9305 (0.0%)	0.80	6/12556 (0.0%)
4	E	0.50	0/629	0.70	0/847
5	F	0.59	0/2282	0.84	2/3076 (0.1%)
6	1	0.37	0/762	0.65	1/1175 (0.1%)
7	2	0.50	0/757	0.70	0/1167
8	3	0.78	1/71 (1.4%)	0.86	0/106
All	All	0.59	12/28143 (0.0%)	0.78	13/38291 (0.0%)

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	565	GLU	CD-OE1	11.18	1.38	1.25
3	D	346	ARG	CD-NE	10.32	1.64	1.46
2	C	565	GLU	CD-OE2	8.43	1.34	1.25
2	C	1272	GLU	CD-OE2	8.18	1.34	1.25
2	C	1192	GLU	CB-CG	7.58	1.66	1.52

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	538	ARG	NE-CZ-NH1	8.24	124.42	120.30
3	D	346	ARG	NE-CZ-NH2	7.10	123.85	120.30
3	D	346	ARG	NE-CZ-NH1	-6.77	116.92	120.30
6	1	38	DT	O5'-P-OP1	-6.66	99.71	105.70
3	D	417	ARG	NE-CZ-NH1	6.39	123.50	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	42	0
1	B	1767	0	1789	34	0
2	C	10570	0	10582	263	0
3	D	9163	0	9357	257	0
4	E	627	0	634	14	0
5	F	2253	0	2298	135	0
6	1	680	0	373	69	0
7	2	675	0	373	44	0
8	3	97	0	45	6	0
9	C	9	0	0	0	0
10	D	2	0	0	2	0
11	D	2	0	0	0	0
All	All	27632	0	27264	728	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 13.

The worst 5 of 728 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:1133:ASP:O	3:D:1244:GLN:NE2	1.66	1.24
5:F:110:GLY:HA2	5:F:119:LEU:HD11	1.33	1.07
5:F:231:GLU:HG3	5:F:232:LYS:H	1.16	1.06
2:C:967:LEU:HD21	2:C:1021:LEU:HD13	1.44	1.00
6:1:50:DT:H3'	6:1:51:DC:H5"	1.47	0.95

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	228/242 (94%)	216 (95%)	10 (4%)	2 (1%)	17	57
1	B	226/242 (93%)	212 (94%)	13 (6%)	1 (0%)	34	71
2	C	1338/1342 (100%)	1244 (93%)	85 (6%)	9 (1%)	22	61
3	D	1169/1407 (83%)	1095 (94%)	64 (6%)	10 (1%)	17	57
4	E	77/90 (86%)	74 (96%)	3 (4%)	0	100	100
5	F	275/336 (82%)	256 (93%)	13 (5%)	6 (2%)	6	39
All	All	3313/3659 (90%)	3097 (94%)	188 (6%)	28 (1%)	19	59

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	908	GLU
3	D	174	ASP
3	D	519	ASN
3	D	710	ASP
3	D	1268	ASN

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	84
1	B	196/208 (94%)	192 (98%)	4 (2%)	55	79
2	C	1155/1157 (100%)	1134 (98%)	21 (2%)	59	81

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	D	980/1168 (84%)	950 (97%)	30 (3%)	40	71
4	E	67/74 (90%)	65 (97%)	2 (3%)	41	71
5	F	240/292 (82%)	238 (99%)	2 (1%)	81	91
All	All	2836/3107 (91%)	2774 (98%)	62 (2%)	52	77

5 of 62 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	93	THR
3	D	1227	HIS
3	D	321	LYS
3	D	1189	MET
4	E	45	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1279	GLN
4	E	43	ASN
5	F	251	GLN
3	D	153	ASN
2	C	1313	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	3	1/4 (25%)	1 (100%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
8	3	16	G

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)

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 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	$\#\|Z\| > 2$
9	DPO	C	1401	11	6,8,8	2.59	4 (66%)	13,13,13	2.89	7 (53%)

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
9	DPO	C	1401	11	-	0/6/6/6	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
9	C	1401	DPO	P2-O5	-4.12	1.37	1.50
9	C	1401	DPO	P1-O2	-3.15	1.42	1.54
9	C	1401	DPO	P1-O3	-2.78	1.44	1.54
9	C	1401	DPO	P2-O6	2.27	1.63	1.54

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
9	C	1401	DPO	P2-O4-P1	-5.13	115.21	132.83
9	C	1401	DPO	O2-P1-O4	4.47	119.63	104.64
9	C	1401	DPO	O6-P2-O5	3.88	125.86	110.68
9	C	1401	DPO	O4-P1-O1	-3.28	92.99	111.19
9	C	1401	DPO	O7-P2-O5	-3.20	98.14	110.68

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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
8	3	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	3	16:G	O3'	17:U	P	3.02

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	230/242 (95%)	-0.10	2 (0%)	84	73	158, 205, 278, 328	0
1	B	228/242 (94%)	0.32	19 (8%)	11	7	161, 241, 296, 359	0
2	C	1340/1342 (99%)	-0.06	33 (2%)	57	41	115, 181, 287, 402	0
3	D	1173/1407 (83%)	0.03	43 (3%)	41	27	124, 189, 286, 340	0
4	E	79/90 (87%)	-0.25	2 (2%)	57	41	166, 227, 324, 373	0
5	F	277/336 (82%)	0.17	8 (2%)	51	35	161, 225, 276, 362	0
6	1	33/50 (66%)	-0.38	1 (3%)	50	34	190, 238, 348, 380	0
7	2	33/50 (66%)	-0.30	0	100	100	129, 239, 314, 386	0
8	3	3/4 (75%)	1.57	0	100	100	128, 128, 146, 150	0
All	All	3396/3763 (90%)	0.00	108 (3%)	47	32	115, 198, 290, 402	0

The worst 5 of 108 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	323	ASN	9.6
3	D	855	ASP	6.5
3	D	853	THR	5.9
3	D	852	GLY	5.8
3	D	1185	PRO	5.6

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
11	MG	D	1504	1/1	0.72	0.46	129,129,129,129	0
9	DPO	C	1401	9/9	0.92	0.42	147,158,172,173	0
11	MG	D	1503	1/1	0.98	0.33	101,101,101,101	0
10	ZN	D	1501	1/1	0.98	0.15	200,200,200,200	0
10	ZN	D	1502	1/1	0.99	0.16	211,211,211,211	0

6.5 Other polymers (i)

There are no such residues in this entry.