wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Oct 9, 2018 – 11:52 AM EDT

PDB ID	:	6GH5
EMDB ID:	:	EMD-0001
Title	:	Cryo-EM structure of bacterial RNA polymerase-sigma54 holoenzyme transcription open complex
Authors	:	Glyde, R.; Ye, F.Z.; Zhang, X.D.
Deposited on	:	2018-05-04
Resolution	:	3.40 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

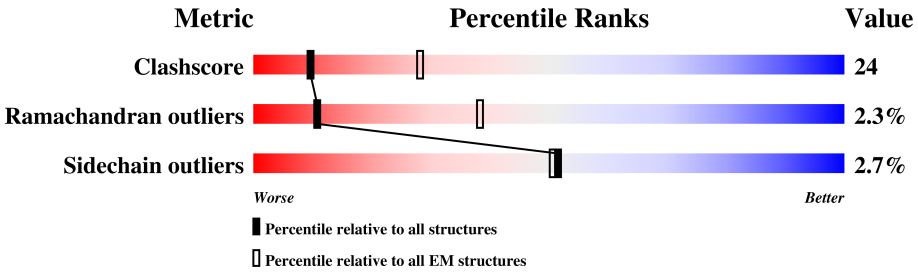
MolProbity	:	4.02b-467
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031633

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.40 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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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%

Mol	Chain	Length	Quality of chain
1	A	329	56% 30% 8% • 6%
1	B	329	41% 25% 5% • 29%
2	C	1342	54% 30% 13% •
3	D	1407	48% 31% 13% • •
4	E	91	47% 24% 10% • 18%
5	M	497	55% 10% • 32%
6	F	63	14% 46% 13% 27%
7	G	63	29% 44% 27%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 28316 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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					AltConf	Trace
1	A	309	Total	C	N	O	S	0	0
			2302	1441	400	454	7		
1	B	235	Total	C	N	O	S	0	0
			1733	1085	301	341	6		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
2	C	1341	Total	C	N	O	S	0	0
			10034	6289	1746	1961	38		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
3	D	1345	Total	C	N	O	S	0	0
			9790	6144	1746	1858	42		

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

Mol	Chain	Residues	Atoms					AltConf	Trace
4	E	75	Total	C	N	O	S	0	0
			565	347	110	107	1		

- Molecule 5 is a protein called RNA polymerase sigma-54 factor,RNA polymerase sigma-54 factor,RNA polymerase sigma-54 factor,RNA polymerase sigma-54 factor,RNA polymerase sigma-54 factor.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	M	338	Total	C	N	O	S	0	0
			2002	1243	356	400	3		

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	-19	MET	-	initiating methionine	UNP A0A0J4U551
M	-18	GLY	-	expression tag	UNP A0A0J4U551
M	-17	SER	-	expression tag	UNP A0A0J4U551
M	-16	SER	-	expression tag	UNP A0A0J4U551
M	-15	HIS	-	expression tag	UNP A0A0J4U551
M	-14	HIS	-	expression tag	UNP A0A0J4U551
M	-13	HIS	-	expression tag	UNP A0A0J4U551
M	-12	HIS	-	expression tag	UNP A0A0J4U551
M	-11	HIS	-	expression tag	UNP A0A0J4U551
M	-10	HIS	-	expression tag	UNP A0A0J4U551
M	-9	SER	-	expression tag	UNP A0A0J4U551
M	-8	SER	-	expression tag	UNP A0A0J4U551
M	-7	GLY	-	expression tag	UNP A0A0J4U551
M	-6	LEU	-	expression tag	UNP A0A0J4U551
M	-5	VAL	-	expression tag	UNP A0A0J4U551
M	-4	PRO	-	expression tag	UNP A0A0J4U551
M	-3	ARG	-	expression tag	UNP A0A0J4U551
M	-2	GLY	-	expression tag	UNP A0A0J4U551
M	-1	SER	-	expression tag	UNP A0A0J4U551
M	0	HIS	-	expression tag	UNP A0A0J4U551
M	336	ALA	ARG	engineered mutation	UNP A0A0J4U551

- Molecule 6 is a DNA chain called nifH promoter template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	46	Total	C	N	O	P	0	0
			944	445	182	271	46		

- Molecule 7 is a DNA chain called nifH promoter non-template DNA.

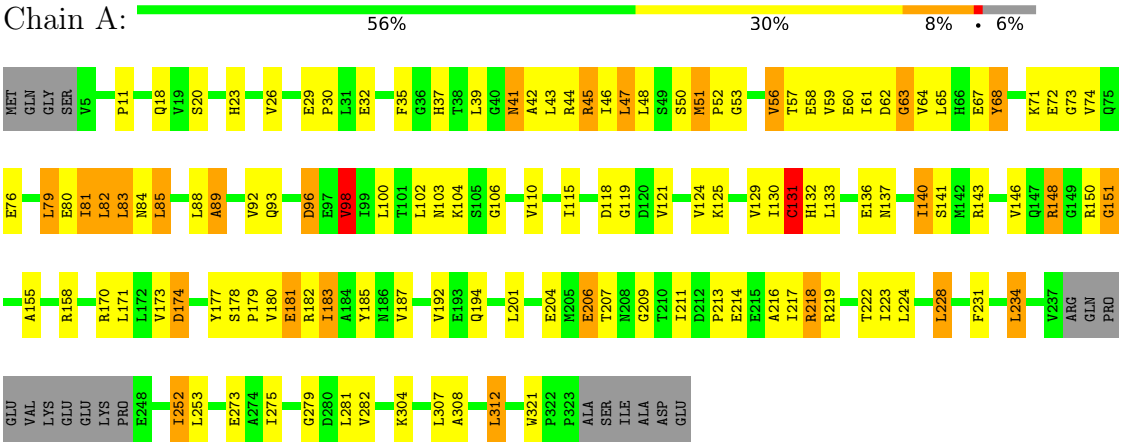
Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	46	Total	C	N	O	P	0	0
			946	448	173	279	46		

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. 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. 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.

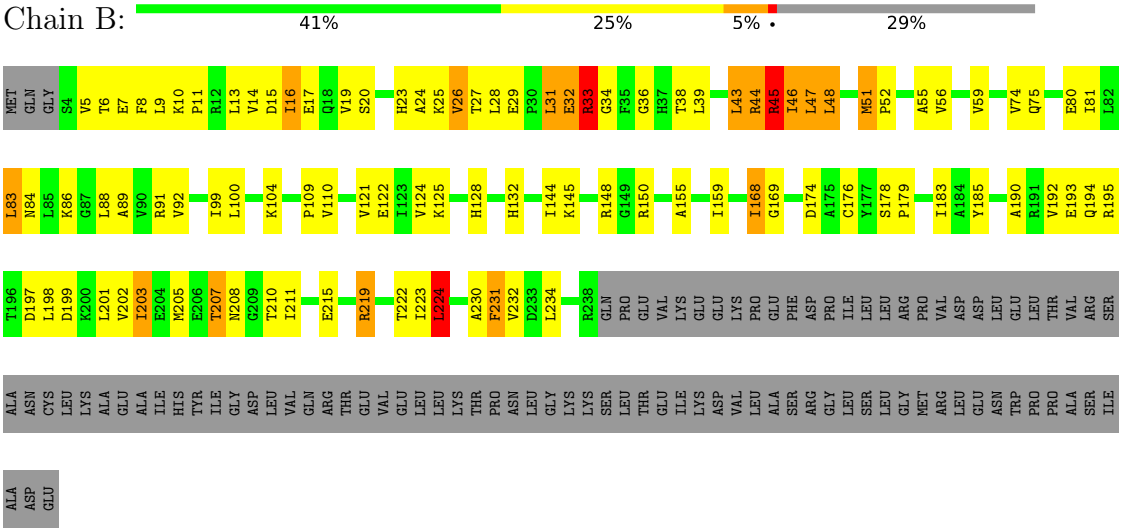
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



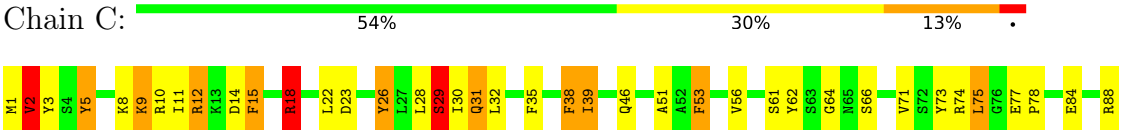
• Molecule 1: DNA-directed RNA polymerase subunit alpha

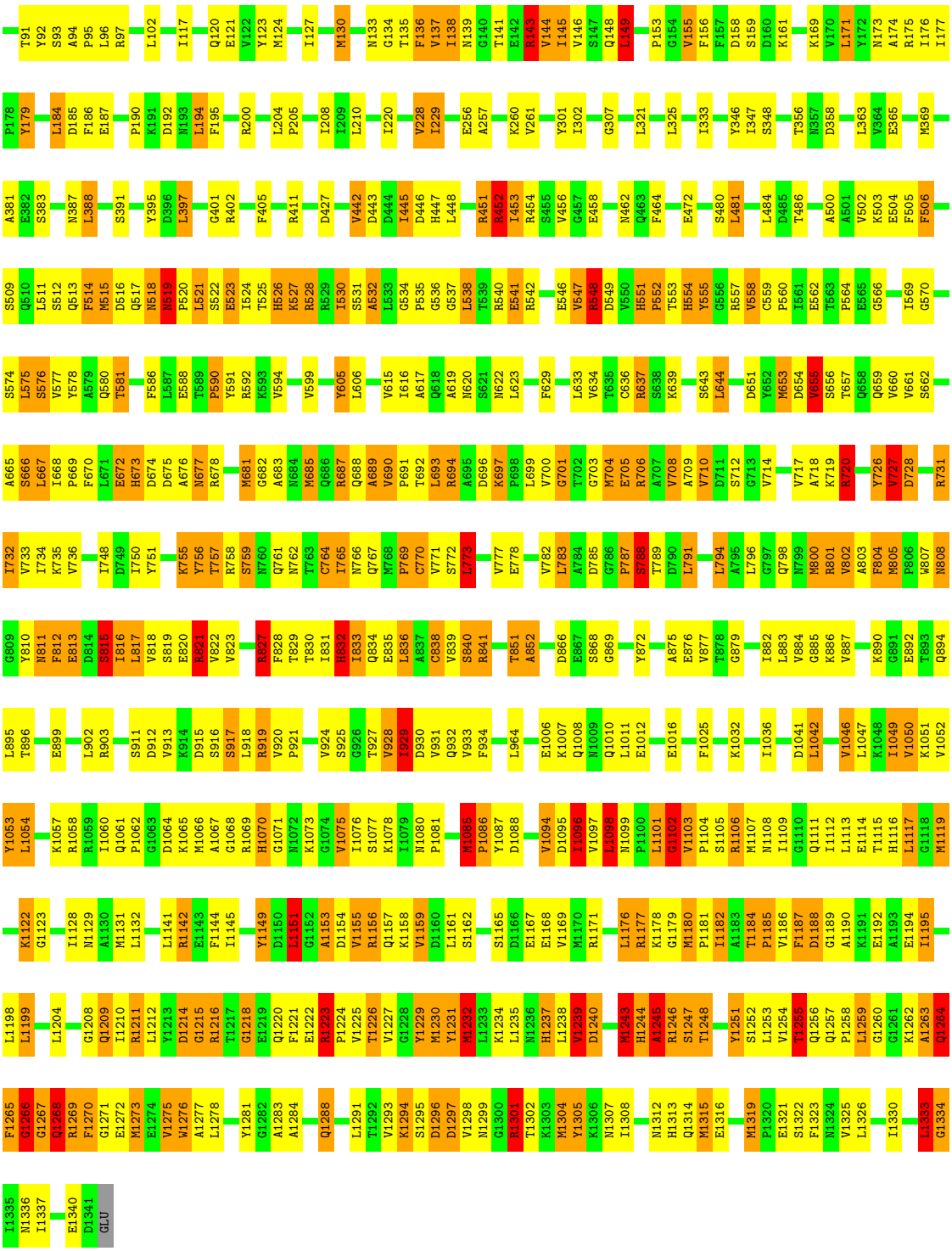
Chain B:



• Molecule 2: DNA-directed RNA polymerase subunit beta

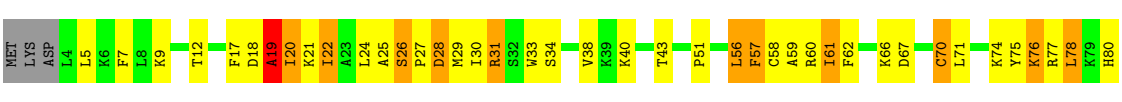
Chain C:

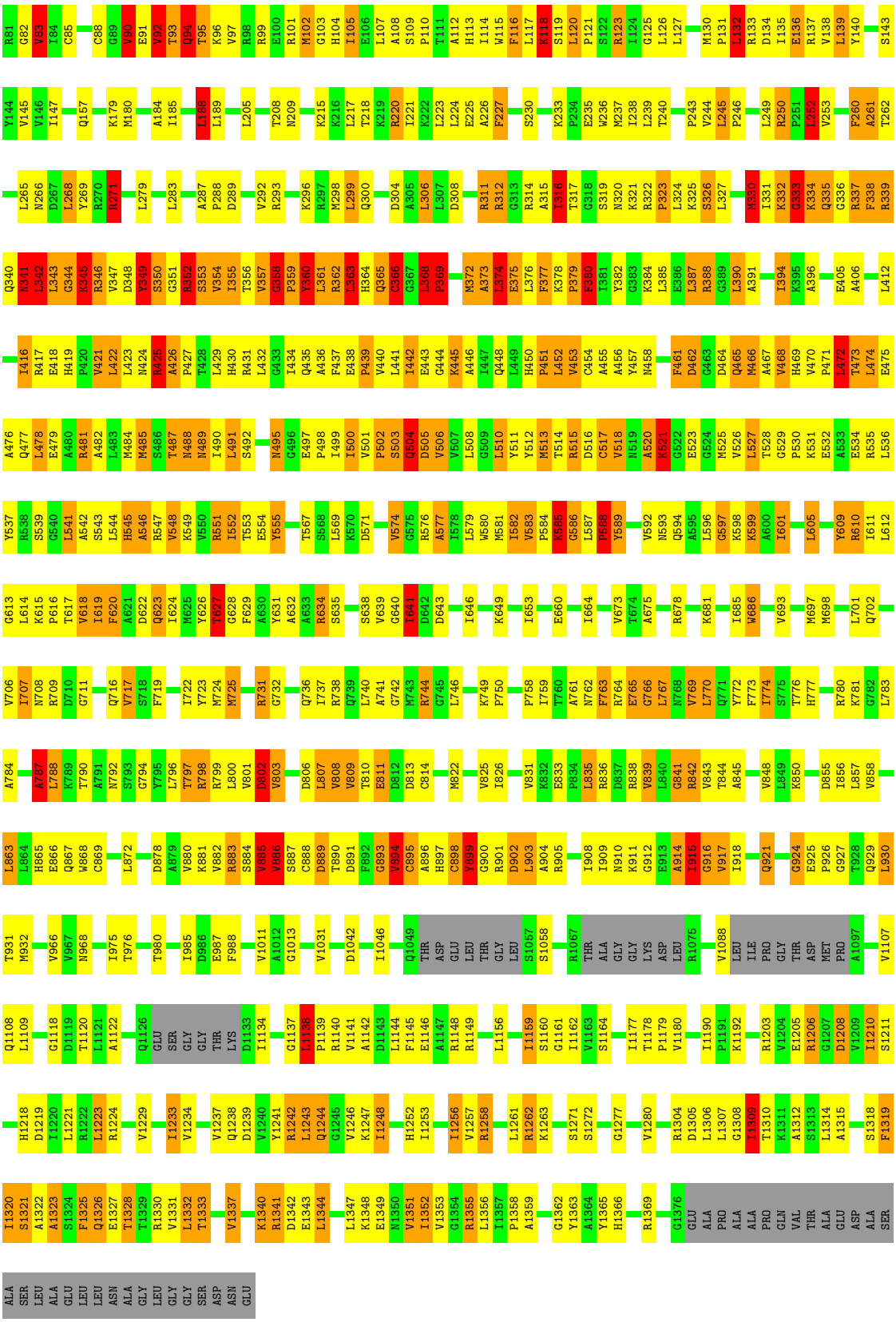




• Molecule 3: DNA-directed RNA polymerase subunit beta'

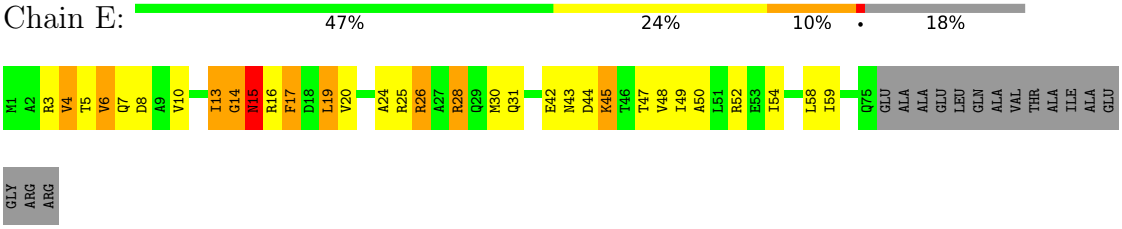
Chain D:





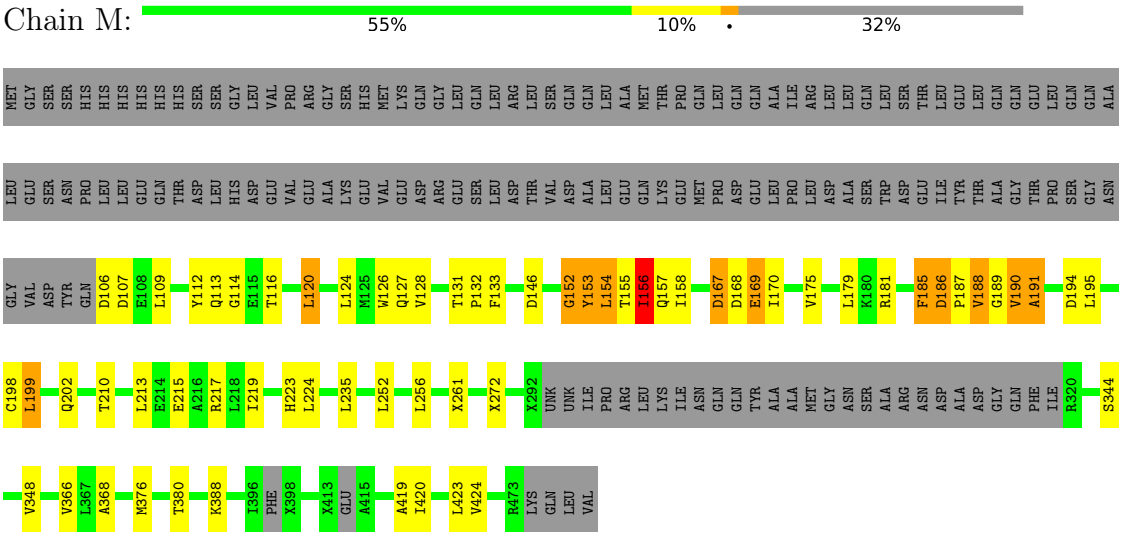
• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E:



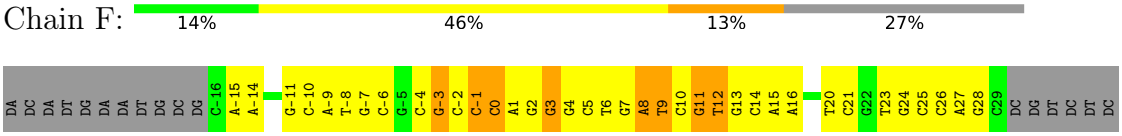
• Molecule 5: RNA polymerase sigma-54 factor,RNA polymerase sigma-54 factor,RNA polymerase sigma-54 factor,RNA polymerase sigma-54 factor,RNA polymerase sigma-54 factor

Chain M:



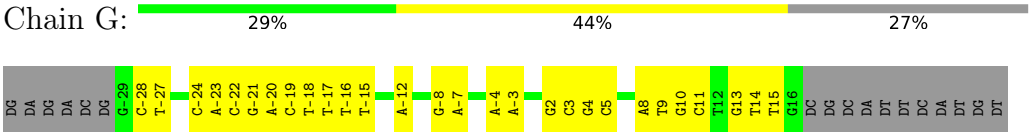
• Molecule 6: nifH promoter template DNA

Chain F:



• Molecule 7: nifH promoter non-template DNA

Chain G:



4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	79678	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	45	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

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	#$\|Z\|$ >2	RMSZ	#$\|Z\|$ >2
1	A	1.46	24/2331 (1.0%)	1.65	48/3177 (1.5%)
1	B	1.15	8/1752 (0.5%)	1.41	29/2385 (1.2%)
2	C	1.84	276/10187 (2.7%)	1.96	384/13822 (2.8%)
3	D	1.84	272/9923 (2.7%)	1.98	394/13482 (2.9%)
4	E	1.21	3/567 (0.5%)	1.72	15/767 (2.0%)
5	M	0.93	11/1764 (0.6%)	1.02	8/2445 (0.3%)
6	F	0.70	1/1060 (0.1%)	1.31	17/1633 (1.0%)
7	G	0.54	0/1060	0.84	1/1635 (0.1%)
All	All	1.66	595/28644 (2.1%)	1.81	896/39346 (2.3%)

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
1	A	0	6
2	C	0	28
3	D	0	28
4	E	0	1
5	M	0	1
All	All	0	64

The worst 5 of 595 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	462	ASP	CB-CG	18.37	1.90	1.51
3	D	99	ARG	CZ-NH1	17.98	1.56	1.33
2	C	1216	ARG	CZ-NH1	17.86	1.56	1.33
2	C	555	TYR	CE2-CZ	16.61	1.60	1.38
3	D	333	GLY	N-CA	16.11	1.70	1.46

The worst 5 of 896 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	1216	ARG	NE-CZ-NH2	-24.92	107.84	120.30
2	C	1301	ARG	NE-CZ-NH1	23.30	131.95	120.30
2	C	1301	ARG	NE-CZ-NH2	-23.28	108.66	120.30
2	C	149	LEU	CB-CG-CD1	-23.16	71.63	111.00
3	D	422	LEU	CB-CG-CD2	-22.53	72.70	111.00

There are no chirality outliers.

5 of 64 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	131	CYS	Mainchain
1	A	151	GLY	Mainchain
1	A	47	LEU	Mainchain
1	A	63	GLY	Peptide
1	A	93	GLN	Peptide

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	2302	0	2266	94	0
1	B	1733	0	1720	80	0
2	C	10034	0	9668	444	0
3	D	9790	0	9539	583	0
4	E	565	0	563	27	0
5	M	2002	0	1412	62	0
6	F	944	0	513	99	0
7	G	946	0	518	42	0
All	All	28316	0	26199	1305	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 24.

The worst 5 of 1305 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)
7:G:14:DT:H2"	7:G:15:DT:C7	1.31	1.55

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:C:1066:MET:CE	2:C:1232:MET:HE2	1.31	1.53
2:C:1260:GLY:CA	3:D:346:ARG:HH12	1.20	1.52
3:D:333:GLY:CA	3:D:333:GLY:N	1.70	1.51
3:D:316:ILE:CD1	3:D:317:THR:H	1.23	1.50

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 PDB entries followed by that with respect to all EM entries.

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	305/329 (93%)	269 (88%)	32 (10%)	4 (1%)	13 50
1	B	233/329 (71%)	209 (90%)	22 (9%)	2 (1%)	19 58
2	C	1339/1342 (100%)	1183 (88%)	128 (10%)	28 (2%)	8 40
3	D	1335/1407 (95%)	1127 (84%)	169 (13%)	39 (3%)	5 33
4	E	73/91 (80%)	68 (93%)	3 (4%)	2 (3%)	5 34
5	M	283/497 (57%)	236 (83%)	40 (14%)	7 (2%)	6 36
All	All	3568/3995 (89%)	3092 (87%)	394 (11%)	82 (2%)	11 38

5 of 82 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	252	ILE
2	C	29	SER
2	C	347	ILE
2	C	541	GLU
2	C	1155	VAL

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 PDB entries followed by that with respect to all EM entries.

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	244/286 (85%)	241 (99%)	3 (1%)	74	88
1	B	183/286 (64%)	178 (97%)	5 (3%)	48	77
2	C	1029/1157 (89%)	1003 (98%)	26 (2%)	50	79
3	D	961/1168 (82%)	925 (96%)	36 (4%)	37	70
4	E	56/75 (75%)	55 (98%)	1 (2%)	62	83
5	M	113/393 (29%)	113 (100%)	0	100	100
All	All	2586/3365 (77%)	2515 (97%)	71 (3%)	51	77

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	1243	MET
3	D	93	THR
3	D	987	GLU
2	C	1319	MET
3	D	28	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	688	GLN
2	C	1136	GLN
3	D	907	HIS
2	C	799	ASN
1	B	23	HIS

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)

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.