wwPDB X-ray Structure Validation Summary Report (i)

Feb 15, 2017 – 08:42 am GMT

PDB ID	:	4JK1
Title	:	X-ray crystal structure of Escherichia coli sigma70 holoenzyme in complex with Guanosine tetraphosphate (ppGpp)
Authors	:	Murakami, K.S.
Deposited on	:	2013-03-09
Resolution	:	3.90 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 (i)) were used in the production of this report:

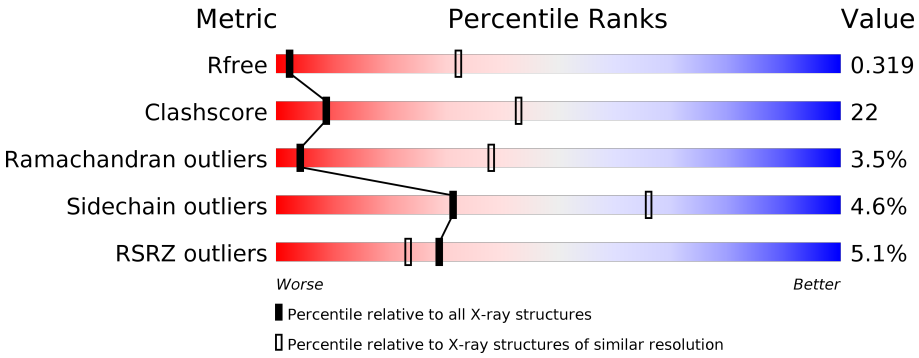
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

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 3.90 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.



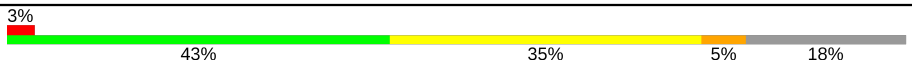
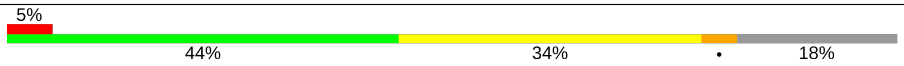
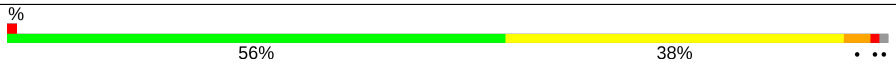
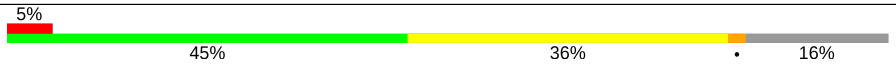
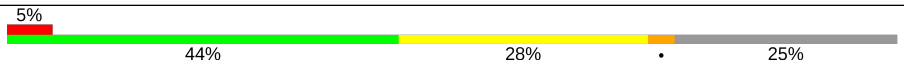
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1007 (4.20-3.60)
Clashscore	112137	1103 (4.20-3.60)
Ramachandran outliers	110173	1062 (4.20-3.60)
Sidechain outliers	110143	1053 (4.20-3.60)
RSRZ outliers	101464	1020 (4.20-3.60)

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	329	4% 64% 31%
1	B	329	4% 39% 26% 33%
1	F	329	5% 48% 20% 30%
1	G	329	4% 39% 26% 34%
2	C	1342	3% 55% 39% 5%
2	H	1342	6% 57% 38% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	D	1407	
3	I	1407	
4	E	91	
4	J	91	
5	X	613	
5	Y	613	

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 56126 atoms, of which 11 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 Escherichia coli RNA polymerase alpha subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	323	Total	C	N	O	S	0	0	0
			2514	1571	443	492	8			
1	B	221	Total	C	N	O	S	0	0	0
			1706	1065	300	335	6			
1	F	229	Total	C	N	O	S	0	0	0
			1775	1106	313	350	6			
1	G	217	Total	C	N	O	S	0	0	0
			1671	1045	293	327	6			

- Molecule 2 is a protein called Escherichia coli RNA polymerase beta subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1335	Total	C	N	O	S	0	0	0
			10523	6601	1836	2043	43			
2	H	1335	Total	C	N	O	S	0	0	0
			10523	6601	1836	2043	43			

- Molecule 3 is a protein called Escherichia coli RNA polymerase beta' subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1160	Total	C	N	O	S	0	0	0
			9060	5695	1621	1697	47			
3	I	1160	Total	C	N	O	S	0	0	0
			9060	5695	1621	1697	47			

- Molecule 4 is a protein called Escherichia coli RNA polymerase omega subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	90	Total	C	N	O	S	0	0	0
			708	430	136	141	1			
4	J	76	Total	C	N	O	S	0	0	0
			605	368	115	121	1			

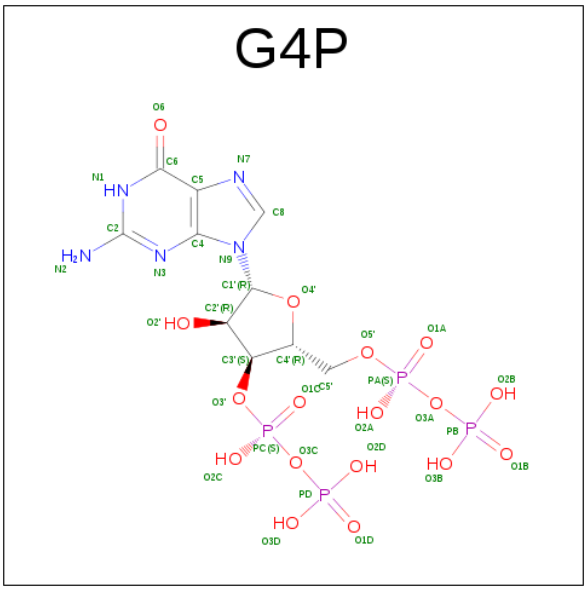
- Molecule 5 is a protein called Escherichia coli RNA polymerase sigma70 subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	X	517	Total	C	N	O	S	0	0	0
			4198	2621	745	806	26			
5	Y	458	Total	C	N	O	S	0	0	0
			3732	2335	671	703	23			

- Molecule 6 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	I	2	Total	Zn	0	0
			2	2		
6	D	2	Total	Zn	0	0
			2	2		

- Molecule 7 is GUANOSINE-5',3'-TETRAPHOSPHATE (three-letter code: G4P) (formula: $C_{10}H_{17}N_5O_{17}P_4$).



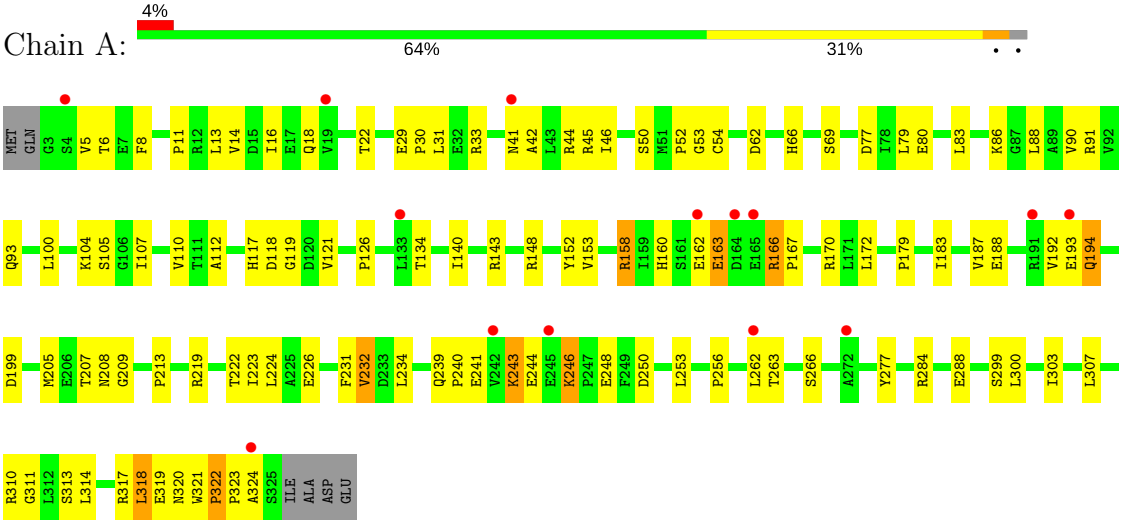
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
7	D	1	Total	C	H	N	O	P	0	0
			47	10	11	5	17	4		

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.

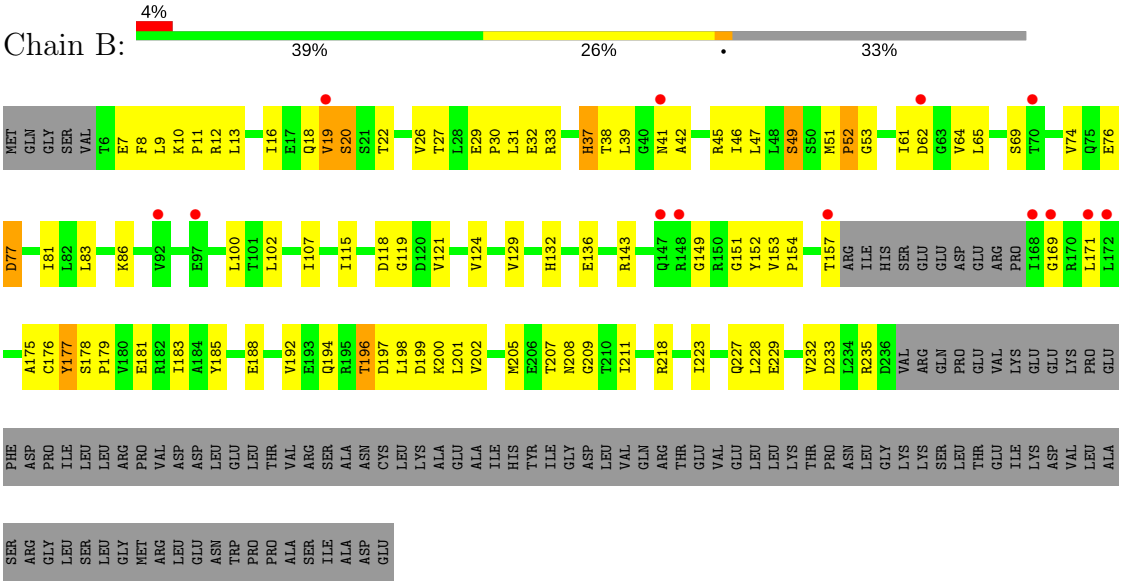
• Molecule 1: Escherichia coli RNA polymerase alpha subunit

Chain A:



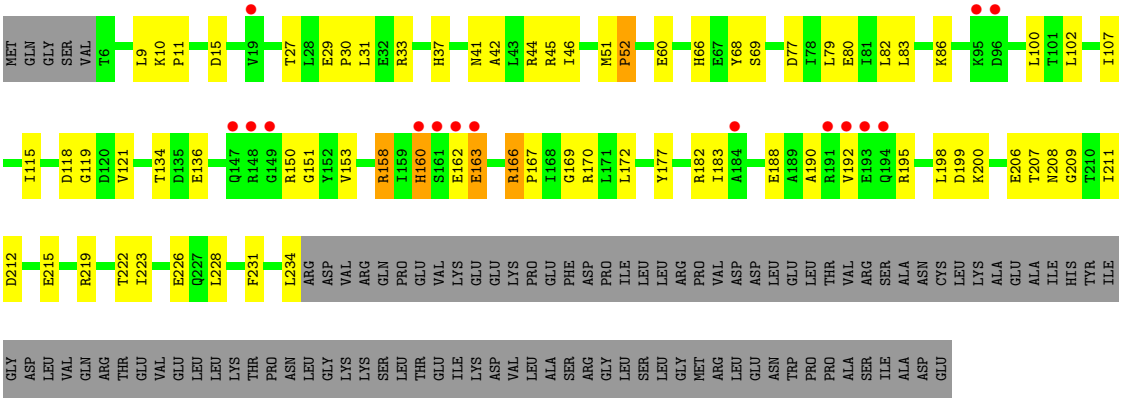
• Molecule 1: Escherichia coli RNA polymerase alpha subunit

Chain B:



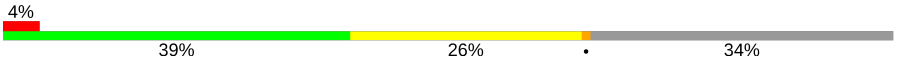
• Molecule 1: Escherichia coli RNA polymerase alpha subunit

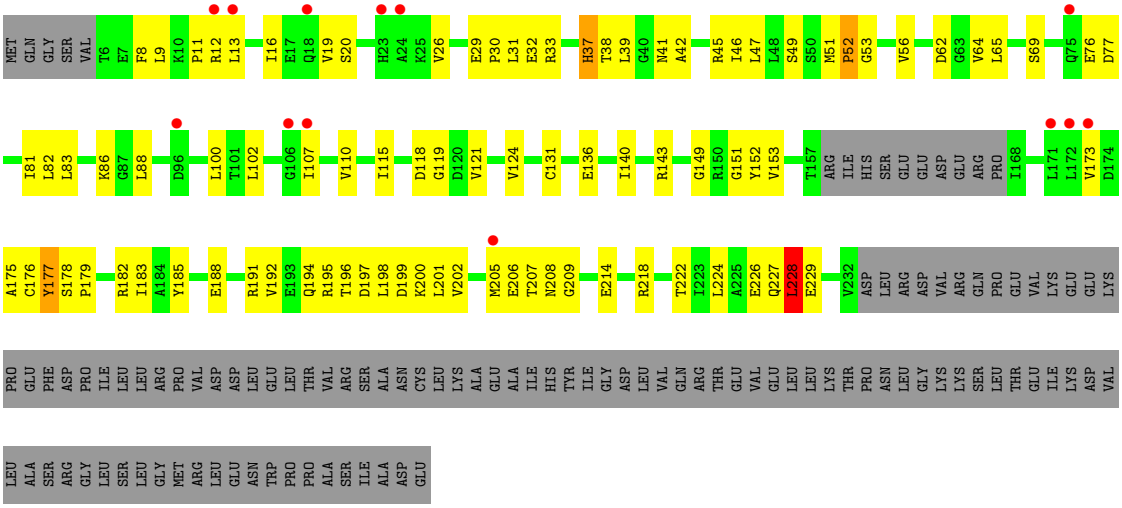
Chain F:



• Molecule 1: Escherichia coli RNA polymerase alpha subunit

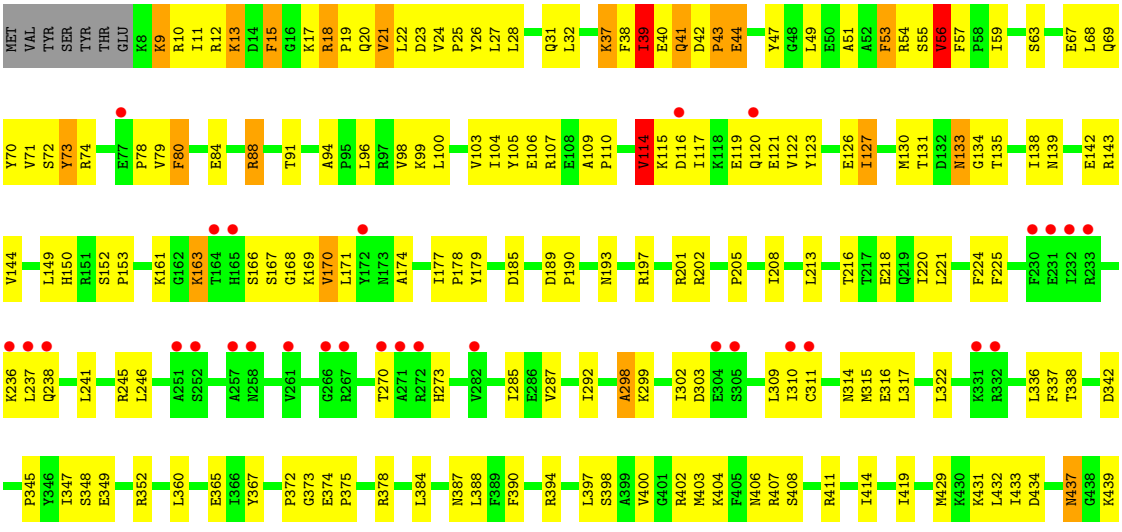
Chain G:

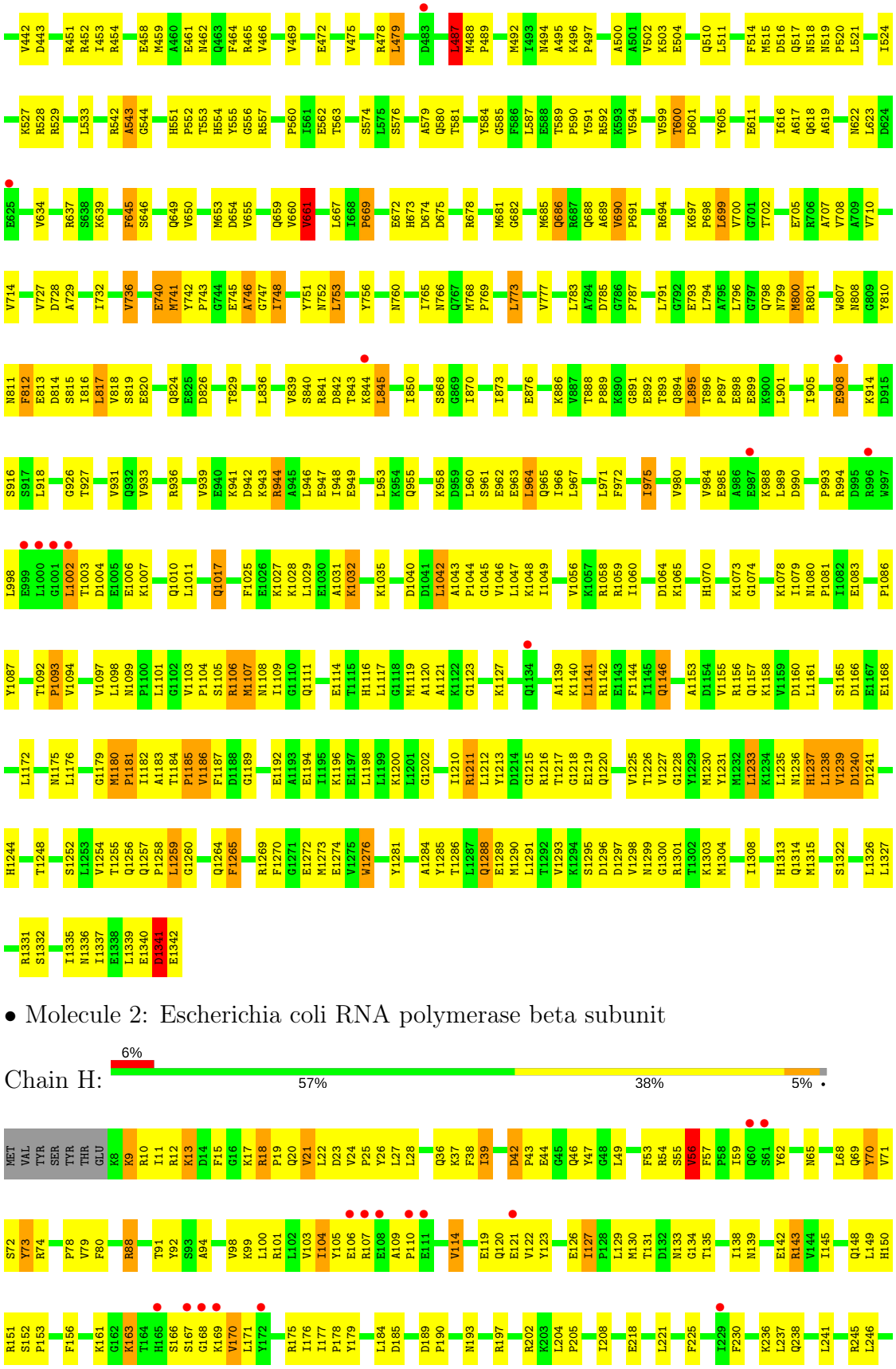




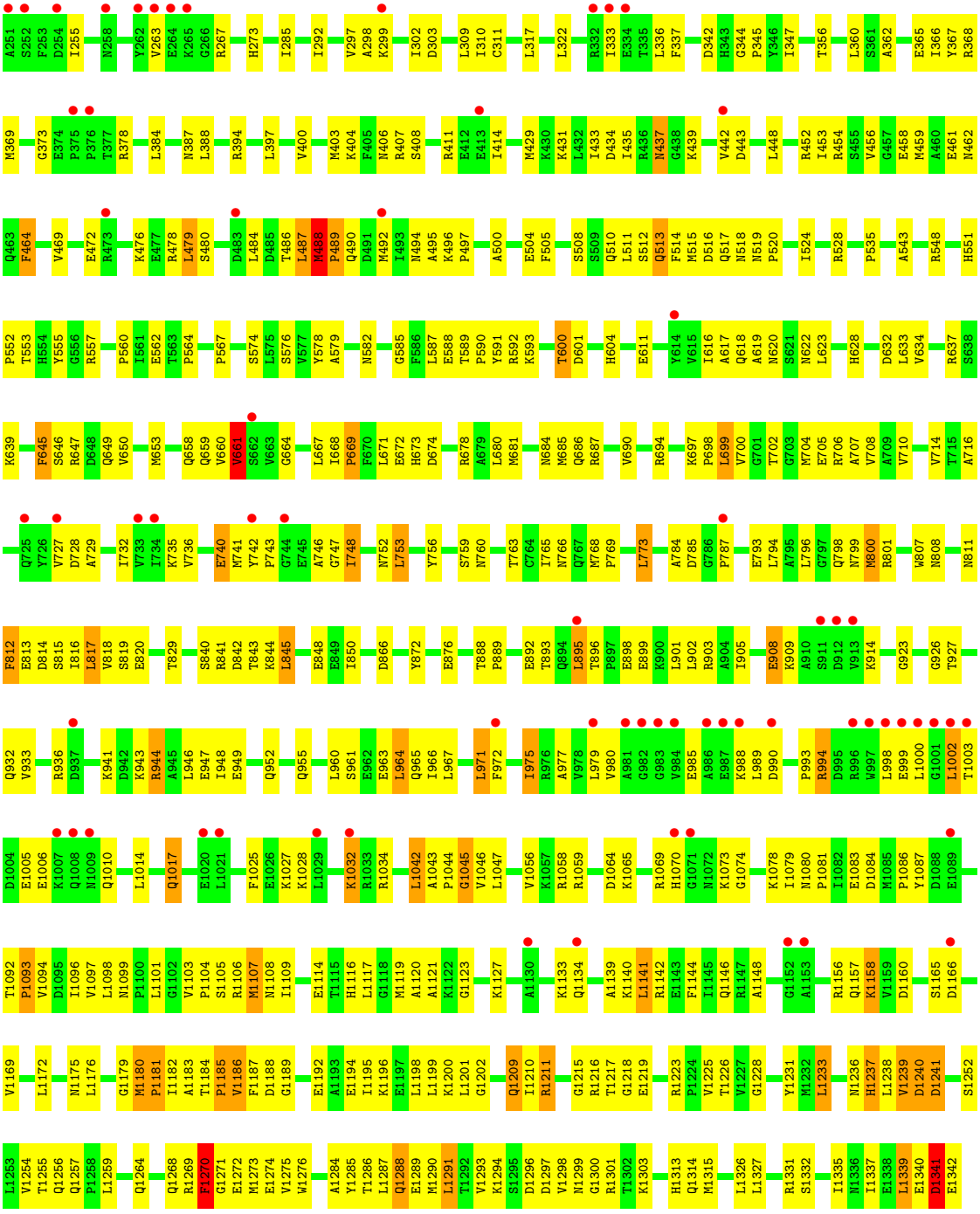
• Molecule 2: Escherichia coli RNA polymerase beta subunit

Chain C:

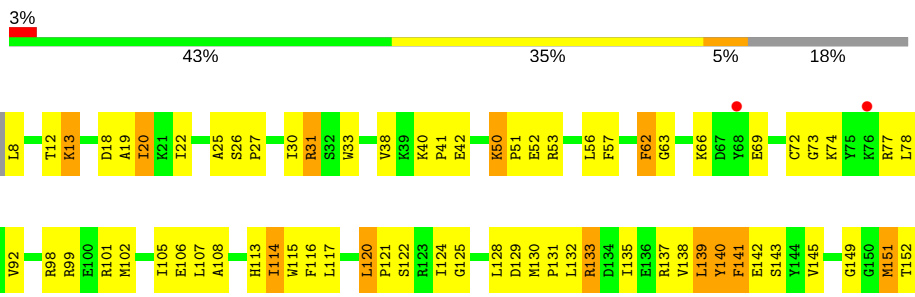


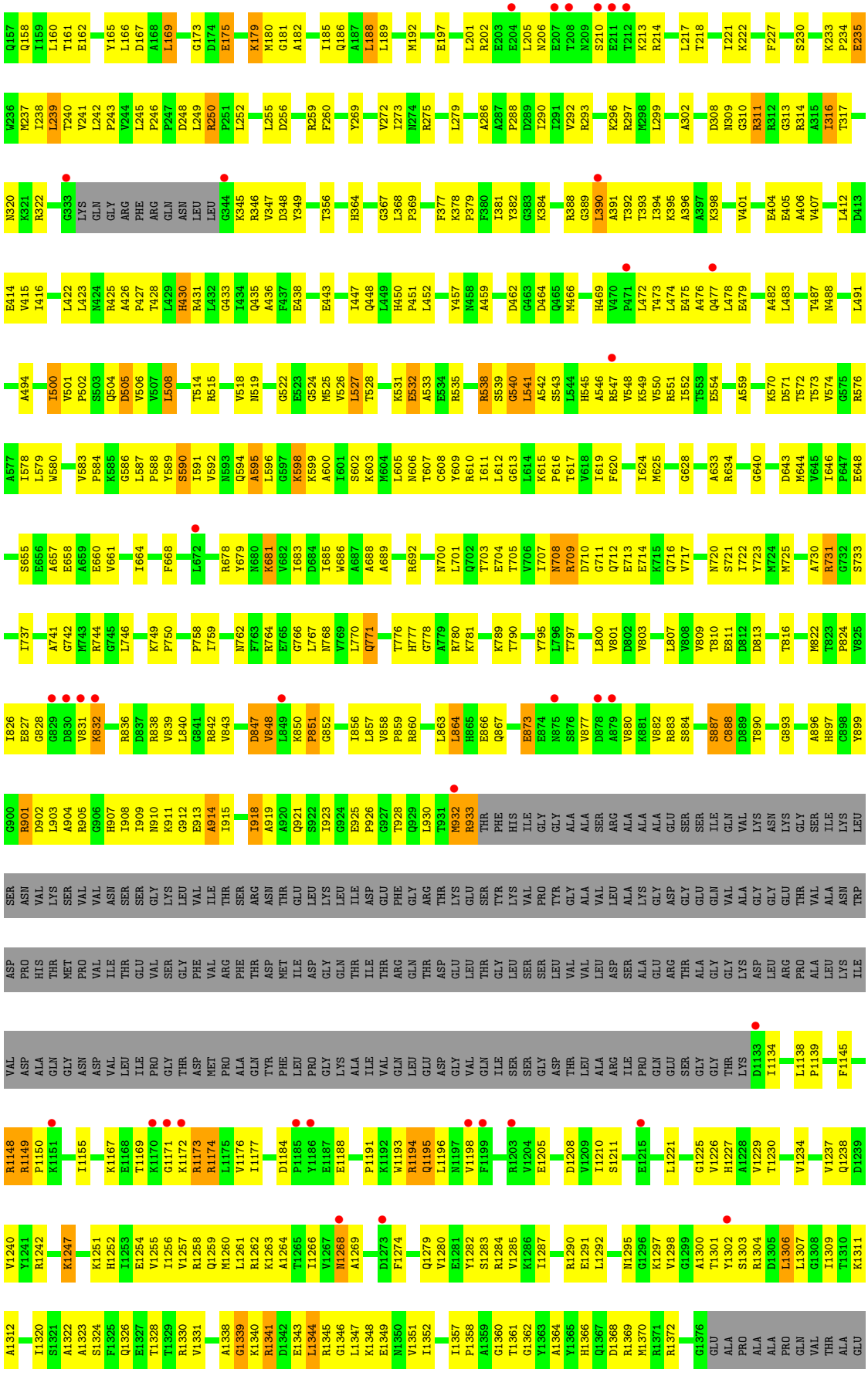


• Molecule 2: Escherichia coli RNA polymerase beta subunit



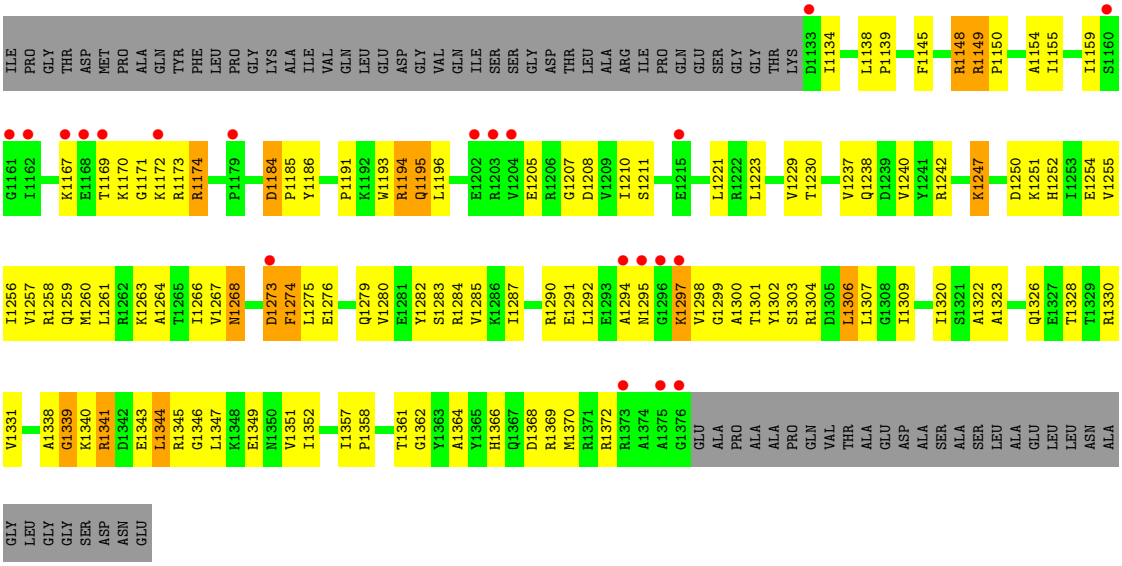
• Molecule 3: Escherichia coli RNA polymerase beta' subunit

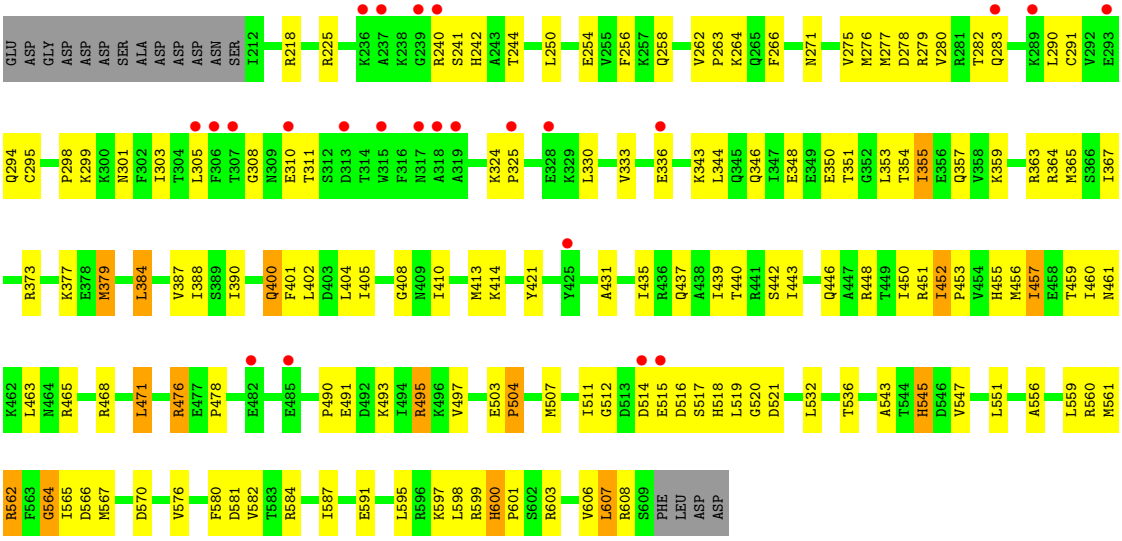




• Molecule 3: Escherichia coli RNA polymerase beta' subunit

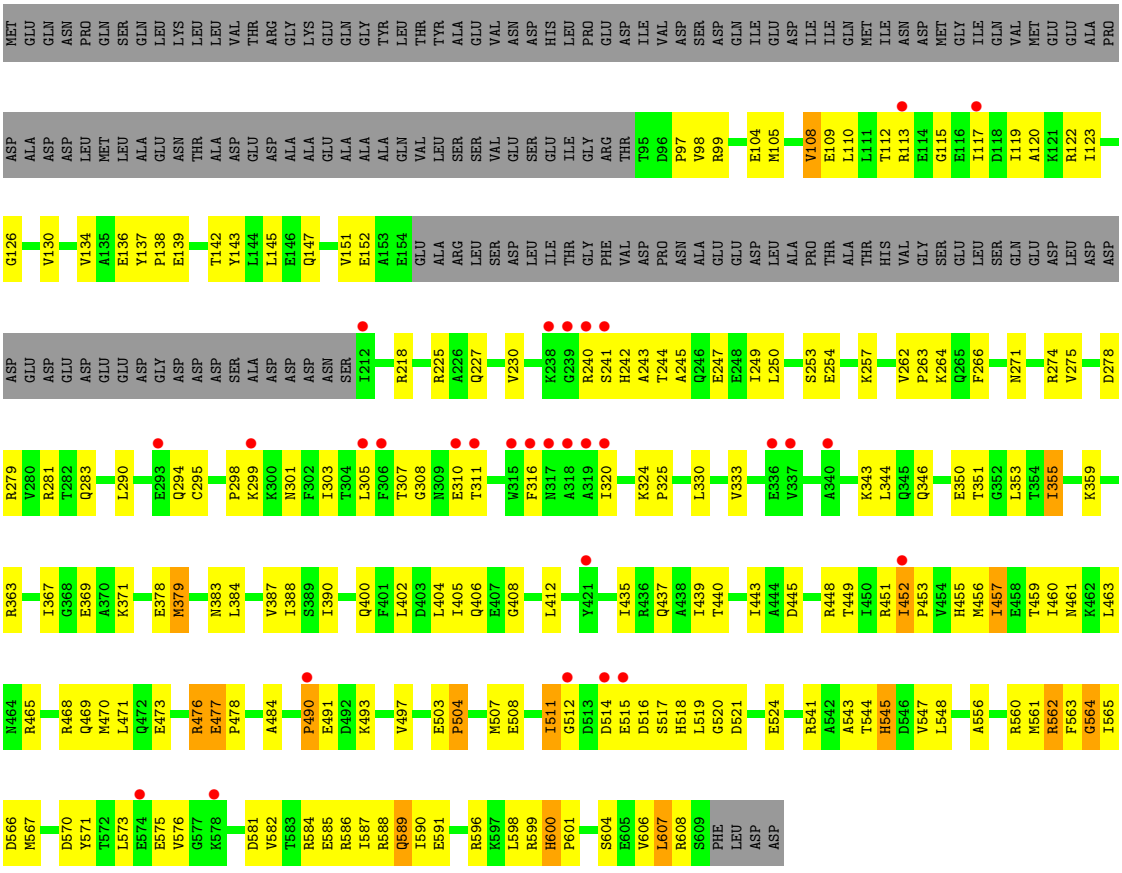
Chain I:





• Molecule 5: Escherichia coli RNA polymerase sigma70 subunit

Chain Y:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	184.57Å 203.82Å 307.67Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.84 – 3.90 29.84 – 3.85	Depositor EDS
% Data completeness (in resolution range)	89.6 (29.84-3.90) 82.4 (29.84-3.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.33 (at 3.86Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.252 , 0.320 0.249 , 0.319	Depositor DCC
R_{free} test set	4724 reflections (4.98%)	DCC
Wilson B-factor (Å^2)	122.2	Xtriage
Anisotropy	0.142	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 37.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.40$, $< L^2 > = 0.22$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	56126	wwPDB-VP
Average B, all atoms (Å^2)	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.71% 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: G4P, ZN

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	0.19	0/2548	0.37	0/3454
1	B	0.19	0/1725	0.38	0/2337
1	F	0.19	0/1797	0.38	0/2436
1	G	0.19	0/1690	0.37	0/2290
2	C	0.20	0/10690	0.37	0/14423
2	H	0.20	0/10690	0.37	0/14423
3	D	0.20	0/9198	0.38	0/12413
3	I	0.20	0/9198	0.38	0/12413
4	E	0.19	0/710	0.38	0/956
4	J	0.19	0/607	0.37	0/817
5	X	0.20	0/4253	0.36	0/5719
5	Y	0.20	0/3783	0.35	0/5083
All	All	0.20	0/56889	0.37	0/76764

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	2514	0	2566	97	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	1706	0	1738	89	0
1	F	1775	0	1800	62	0
1	G	1671	0	1706	84	0
2	C	10523	0	10546	517	0
2	H	10523	0	10546	501	0
3	D	9060	0	9257	530	0
3	I	9060	0	9257	511	0
4	E	708	0	719	40	0
4	J	605	0	612	32	0
5	X	4198	0	4250	169	0
5	Y	3732	0	3809	137	0
6	D	2	0	0	0	0
6	I	2	0	0	0	0
7	D	36	11	11	2	0
All	All	56115	11	56817	2538	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 22.

The worst 5 of 2538 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:I:1173:ARG:HA	3:I:1174:ARG:HB2	1.28	1.14
3:D:1173:ARG:HA	3:D:1174:ARG:HB2	1.29	1.11
3:D:310:GLY:HA3	3:D:311:ARG:HB2	1.23	1.11
2:H:488:MET:HB2	2:H:490:GLN:H	1.14	1.05
2:C:42:ASP:HB3	2:C:43:PRO:HD2	1.38	1.02

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	321/329 (98%)	266 (83%)	41 (13%)	14 (4%)	3 31
1	B	217/329 (66%)	186 (86%)	24 (11%)	7 (3%)	5 40
1	F	227/329 (69%)	196 (86%)	26 (12%)	5 (2%)	8 47
1	G	213/329 (65%)	188 (88%)	21 (10%)	4 (2%)	9 50
2	C	1333/1342 (99%)	1069 (80%)	213 (16%)	51 (4%)	4 35
2	H	1333/1342 (99%)	1070 (80%)	213 (16%)	50 (4%)	4 35
3	D	1154/1407 (82%)	922 (80%)	189 (16%)	43 (4%)	4 36
3	I	1154/1407 (82%)	929 (80%)	183 (16%)	42 (4%)	4 37
4	E	88/91 (97%)	76 (86%)	8 (9%)	4 (4%)	3 31
4	J	74/91 (81%)	64 (86%)	6 (8%)	4 (5%)	2 28
5	X	511/613 (83%)	450 (88%)	45 (9%)	16 (3%)	5 40
5	Y	454/613 (74%)	411 (90%)	32 (7%)	11 (2%)	7 45
All	All	7079/8222 (86%)	5827 (82%)	1001 (14%)	251 (4%)	4 38

5 of 251 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	319	GLU
1	B	20	SER
1	B	52	PRO
2	C	21	VAL
2	C	39	ILE

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	281/286 (98%)	274 (98%)	7 (2%)	53 78
1	B	189/286 (66%)	185 (98%)	4 (2%)	59 82
1	F	197/286 (69%)	192 (98%)	5 (2%)	53 78
1	G	185/286 (65%)	180 (97%)	5 (3%)	50 77

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	C	1150/1157 (99%)	1087 (94%)	63 (6%)	25	62
2	H	1150/1157 (99%)	1092 (95%)	58 (5%)	28	64
3	D	971/1168 (83%)	919 (95%)	52 (5%)	26	62
3	I	971/1168 (83%)	921 (95%)	50 (5%)	28	63
4	E	74/75 (99%)	71 (96%)	3 (4%)	35	69
4	J	65/75 (87%)	65 (100%)	0	100	100
5	X	460/540 (85%)	442 (96%)	18 (4%)	37	69
5	Y	407/540 (75%)	391 (96%)	16 (4%)	37	69
All	All	6100/7024 (87%)	5819 (95%)	281 (5%)	31	66

5 of 281 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	E	8	ASP
2	H	9	LYS
3	I	1247	LYS
5	X	28	ASN
5	X	476	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 84 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1268	ASN
5	X	469	GLN
5	Y	242	HIS
4	E	31	GLN
5	X	54	GLN

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 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	$\#\|Z\| > 2$
7	G4P	D	1503	-	32,38,38	2.37	7 (21%)	32,61,61	1.67	3 (9%)

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
7	G4P	D	1503	-	-	0/23/43/43	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	D	1503	G4P	C2'-C1'	-4.94	1.45	1.53
7	D	1503	G4P	C6-N1	-4.01	1.31	1.36
7	D	1503	G4P	C5-C4	-3.82	1.31	1.40
7	D	1503	G4P	C2'-C3'	-3.58	1.45	1.53
7	D	1503	G4P	C8-N7	2.26	1.38	1.34

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	D	1503	G4P	C4'-O4'-C1'	-6.60	102.75	109.77
7	D	1503	G4P	C4-C5-N7	2.17	111.51	109.41
7	D	1503	G4P	O3C-PC-O3'	5.76	112.34	102.05

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	D	1503	G4P	2	0

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	323/329 (98%)	-0.01	14 (4%)	36	28	1, 73, 178, 299	0
1	B	221/329 (67%)	0.28	13 (5%)	23	17	5, 98, 204, 259	0
1	F	229/329 (69%)	0.30	15 (6%)	19	14	27, 123, 199, 278	0
1	G	217/329 (65%)	0.34	13 (5%)	23	17	34, 118, 187, 236	0
2	C	1335/1342 (99%)	-0.15	41 (3%)	49	39	0, 48, 170, 262	0
2	H	1335/1342 (99%)	0.09	80 (5%)	23	17	0, 88, 203, 293	0
3	D	1160/1407 (82%)	-0.04	41 (3%)	44	35	0, 42, 158, 289	0
3	I	1160/1407 (82%)	0.09	71 (6%)	22	16	0, 60, 192, 317	0
4	E	90/91 (98%)	-0.27	1 (1%)	80	72	1, 52, 121, 158	0
4	J	76/91 (83%)	0.26	5 (6%)	19	14	10, 89, 165, 211	0
5	X	517/613 (84%)	0.10	42 (8%)	13	10	0, 101, 226, 326	0
5	Y	458/613 (74%)	0.14	30 (6%)	19	14	1, 109, 234, 357	0
All	All	7121/8222 (86%)	0.04	366 (5%)	29	23	0, 73, 197, 357	0

The worst 5 of 366 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	I	10	ALA	12.7
5	Y	239	GLY	9.3
3	I	212	THR	9.0
3	I	11	GLN	8.3
3	I	1161	GLY	7.6

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 (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, 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
7	G4P	D	1503	36/36	0.83	0.20	-0.66	31,56,93,118	0
6	ZN	D	1502	1/1	0.99	0.18	-0.70	8,8,8,8	0
6	ZN	I	1502	1/1	0.98	0.15	-0.77	49,49,49,49	0
6	ZN	I	1501	1/1	0.97	0.07	-1.48	60,60,60,60	0
6	ZN	D	1501	1/1	0.98	0.08	-1.51	54,54,54,54	0

6.5 Other polymers (i)

There are no such residues in this entry.