wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:51 PM GMT

PDB ID	:	4JK2
Title	:	X-ray crystal structure of Escherichia coli sigma70 holoenzyme in complex with guanosine pentaphosphate (pppGpp)
Authors	:	Murakami, K.S.
Deposited on	:	2013-03-09
Resolution	:	4.20 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 (1)) were used in the production of this report:

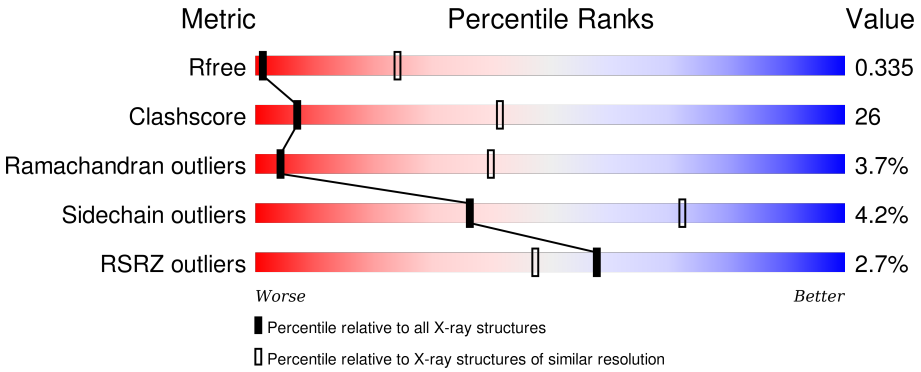
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 4.20 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	1039 (4.80-3.60)
Clashscore	102246	1140 (4.80-3.60)
Ramachandran outliers	100387	1083 (4.80-3.60)
Sidechain outliers	100360	1067 (4.80-3.60)
RSRZ outliers	91569	1042 (4.80-3.60)

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	59% 36%
1	B	329	38% 27% 33%
1	F	329	4% 46% 22% 30%
1	G	329	40% 24% 34%
2	C	1342	53% 41% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	1342	3% 53% 41% 5%
3	D	1407	% 39% 39% 5% 18%
3	I	1407	3% 41% 37% 5% 18%
4	E	91	59% 35%
4	J	91	% 53% 29% 16%
5	X	613	3% 47% 35% 16%
5	Y	613	3% 43% 30% 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
7	0O2	D	1503	-	-	X	-

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 56129 atoms, of which 10 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			

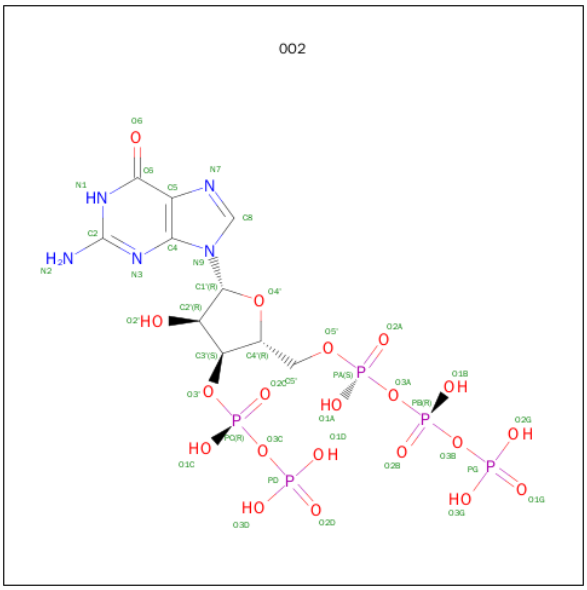
- Molecule 5 is a protein called Escherichia coli RNA polymerase sigma70 subunit.

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'-(TETRAHYDROGEN TRIPHOSPHATE) 3'-(TRIHYDROGEN DIPHOSPHATE) (three-letter code: 0O2) (formula: $C_{10}H_{18}N_5O_{20}P_5$).



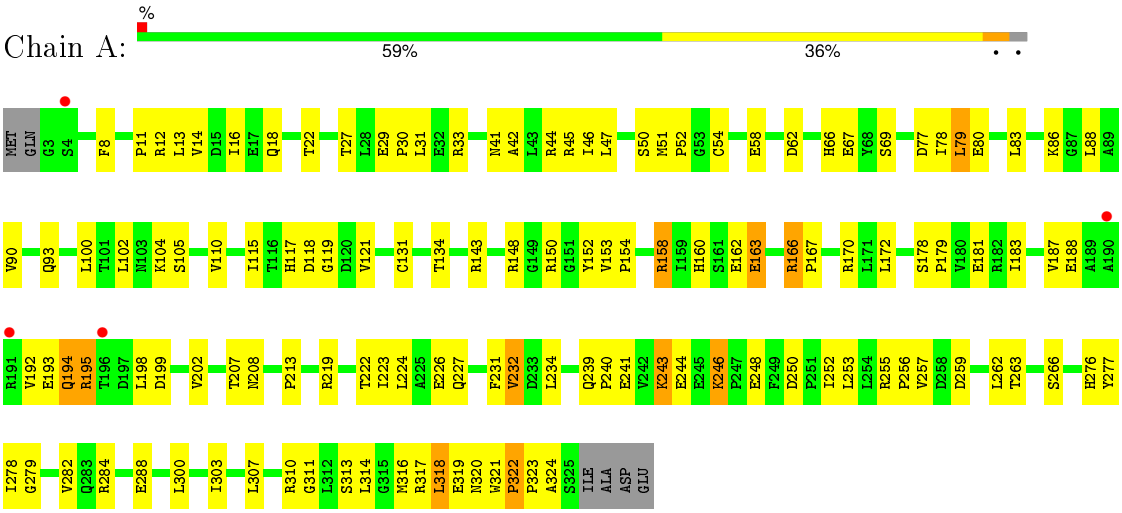
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
7	D	1	Total	C	H	N	O	P	0	0
			50	10	10	5	20	5		

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 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.

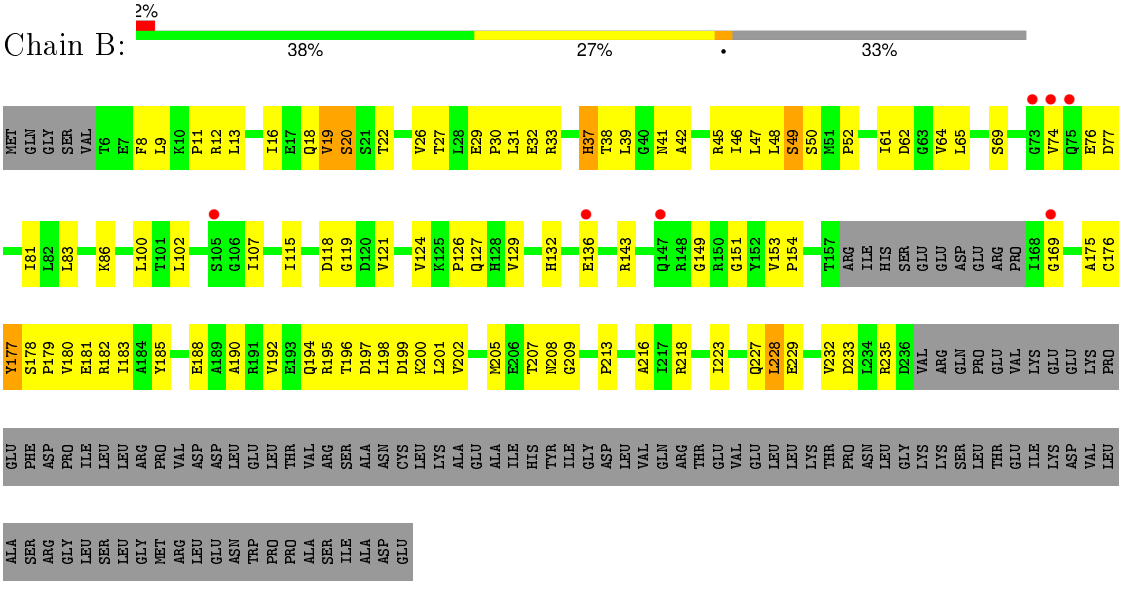
- Molecule 1: Escherichia coli RNA polymerase alpha subunit

Chain A:



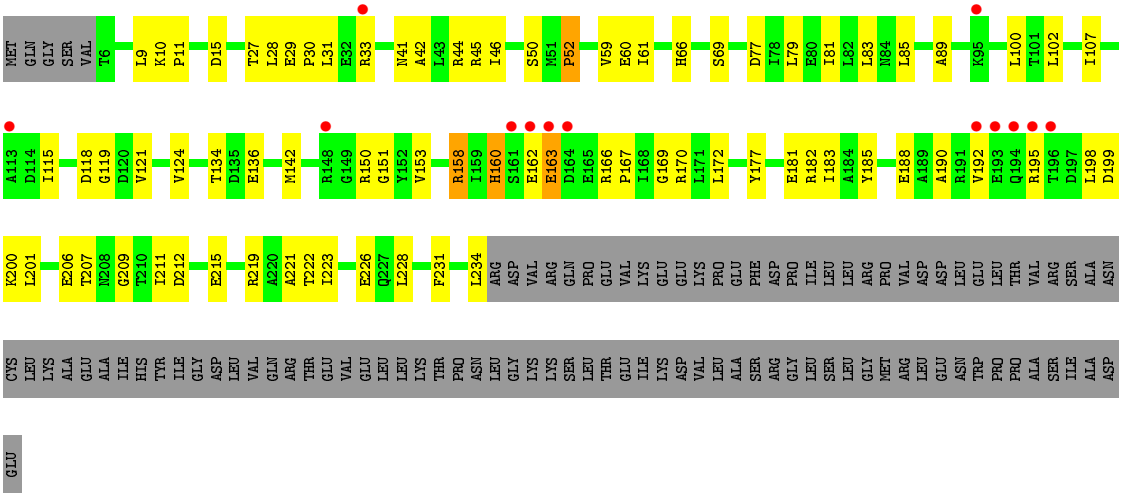
- Molecule 1: Escherichia coli RNA polymerase alpha subunit

Chain B:



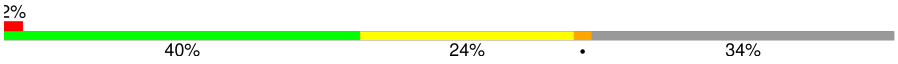
- Molecule 1: Escherichia coli RNA polymerase alpha subunit

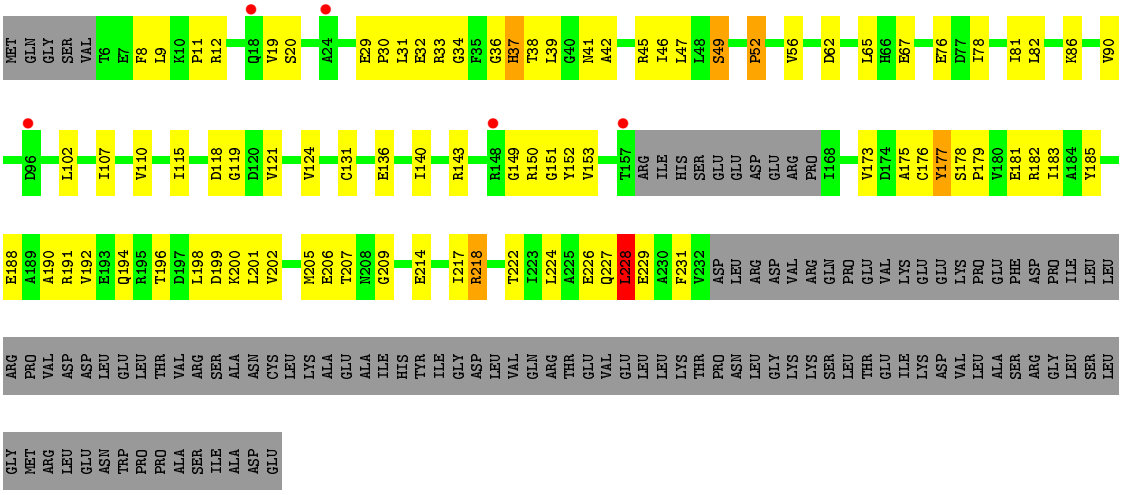
Chain F:



• Molecule 1: Escherichia coli RNA polymerase alpha subunit

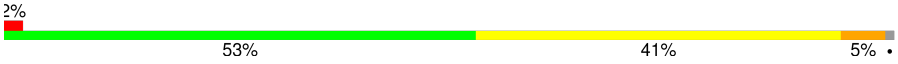
Chain G:

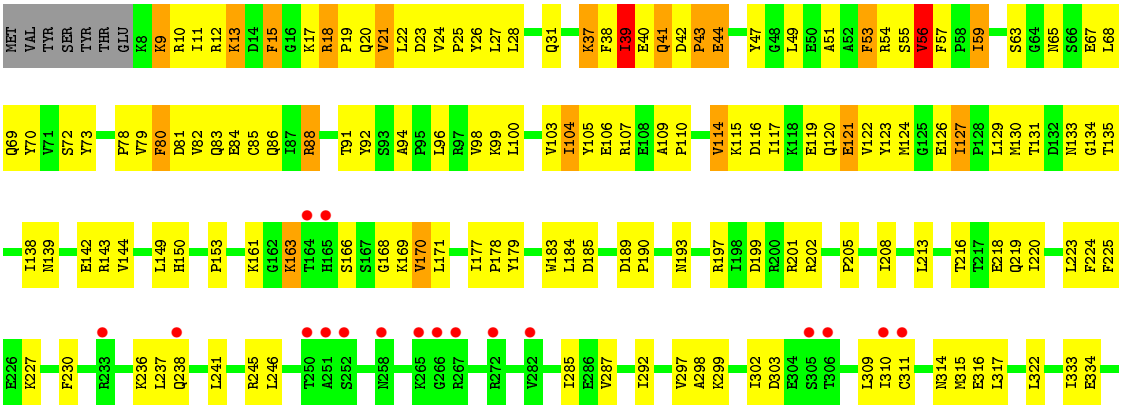


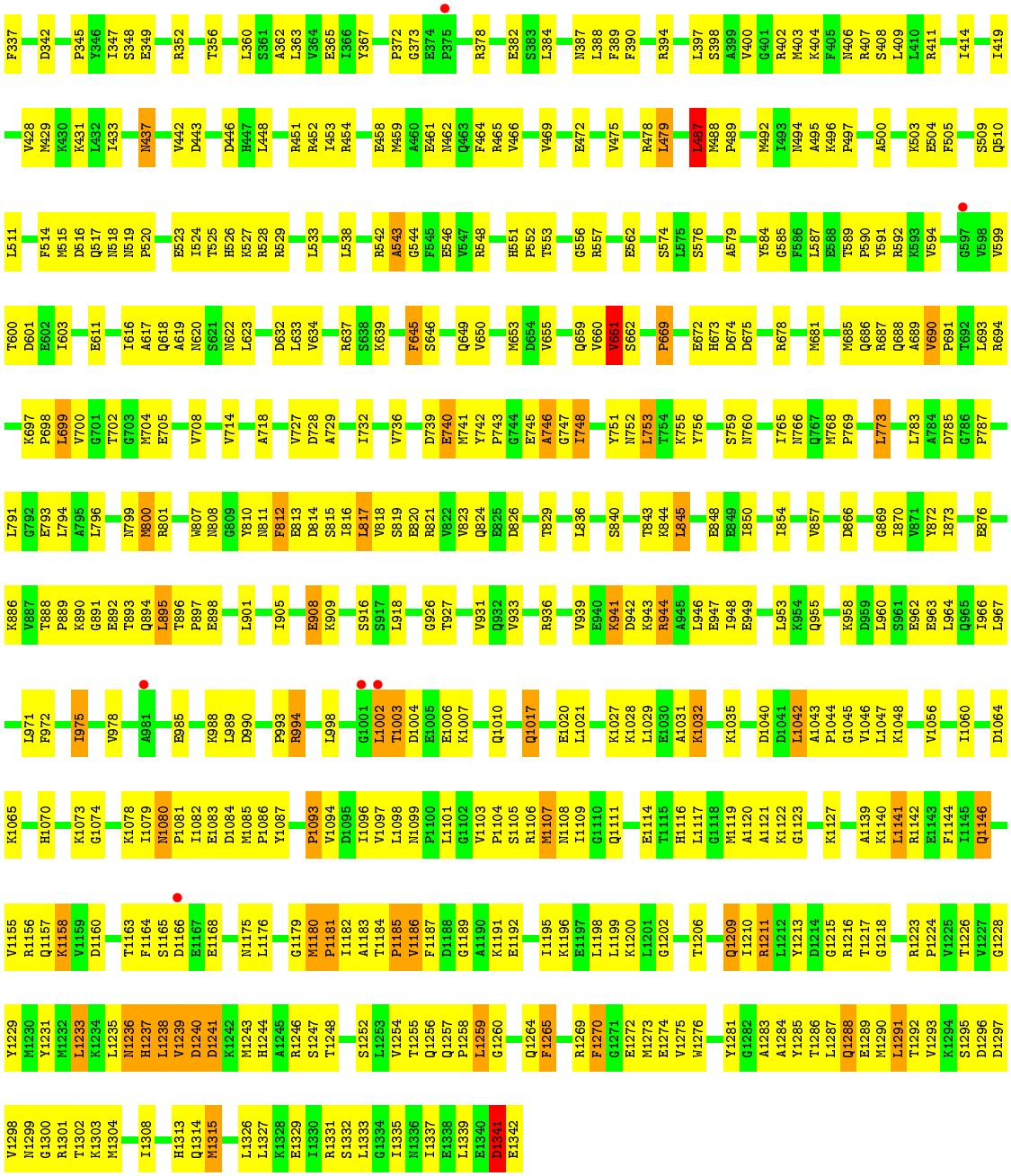


• Molecule 2: Escherichia coli RNA polymerase beta subunit

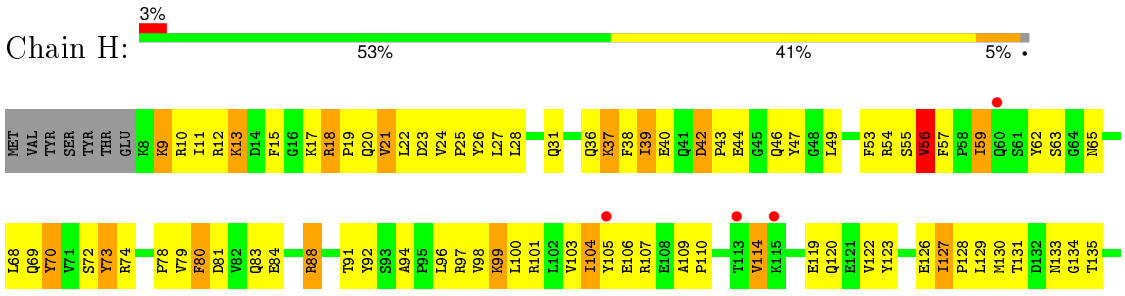
Chain C:

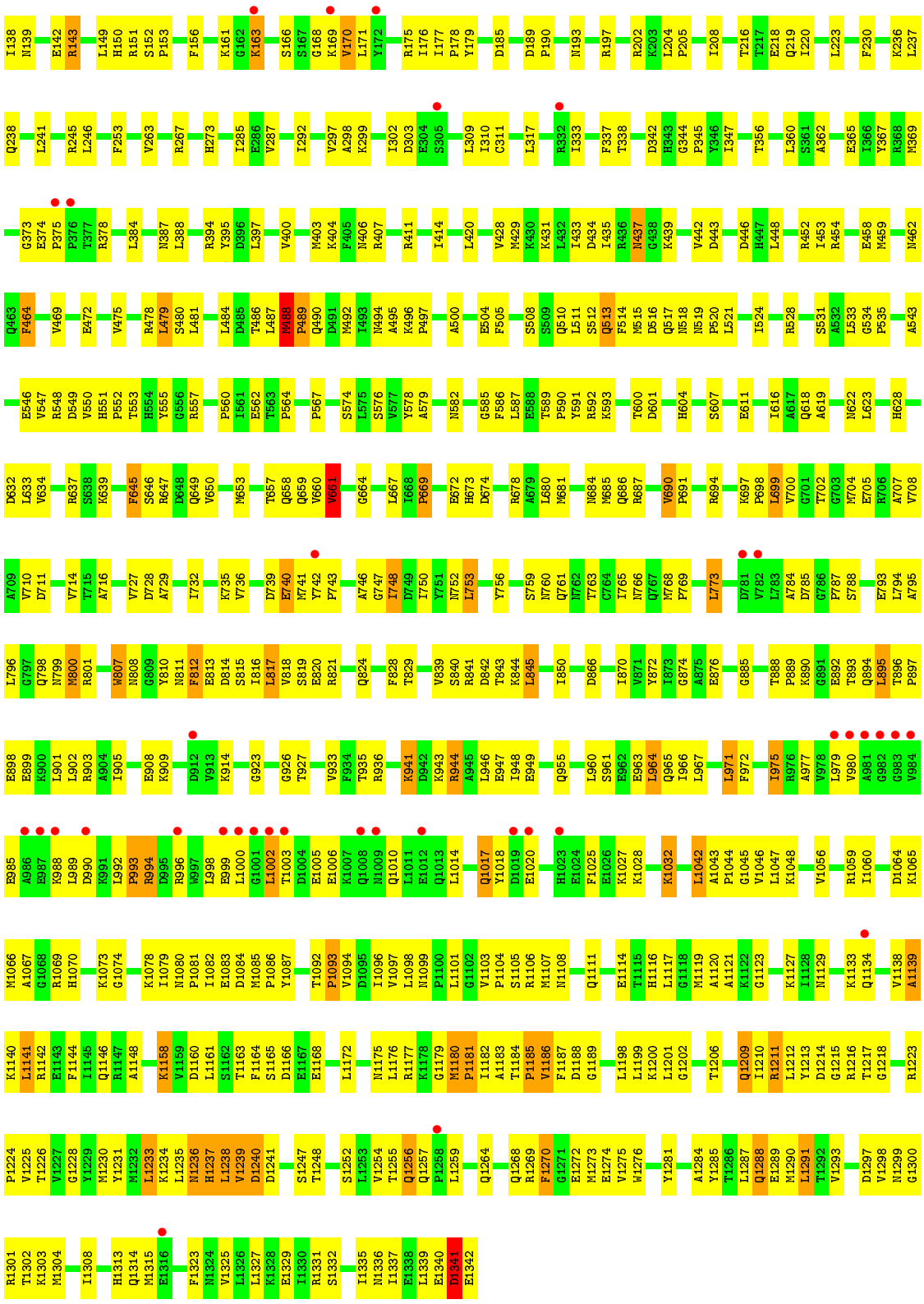




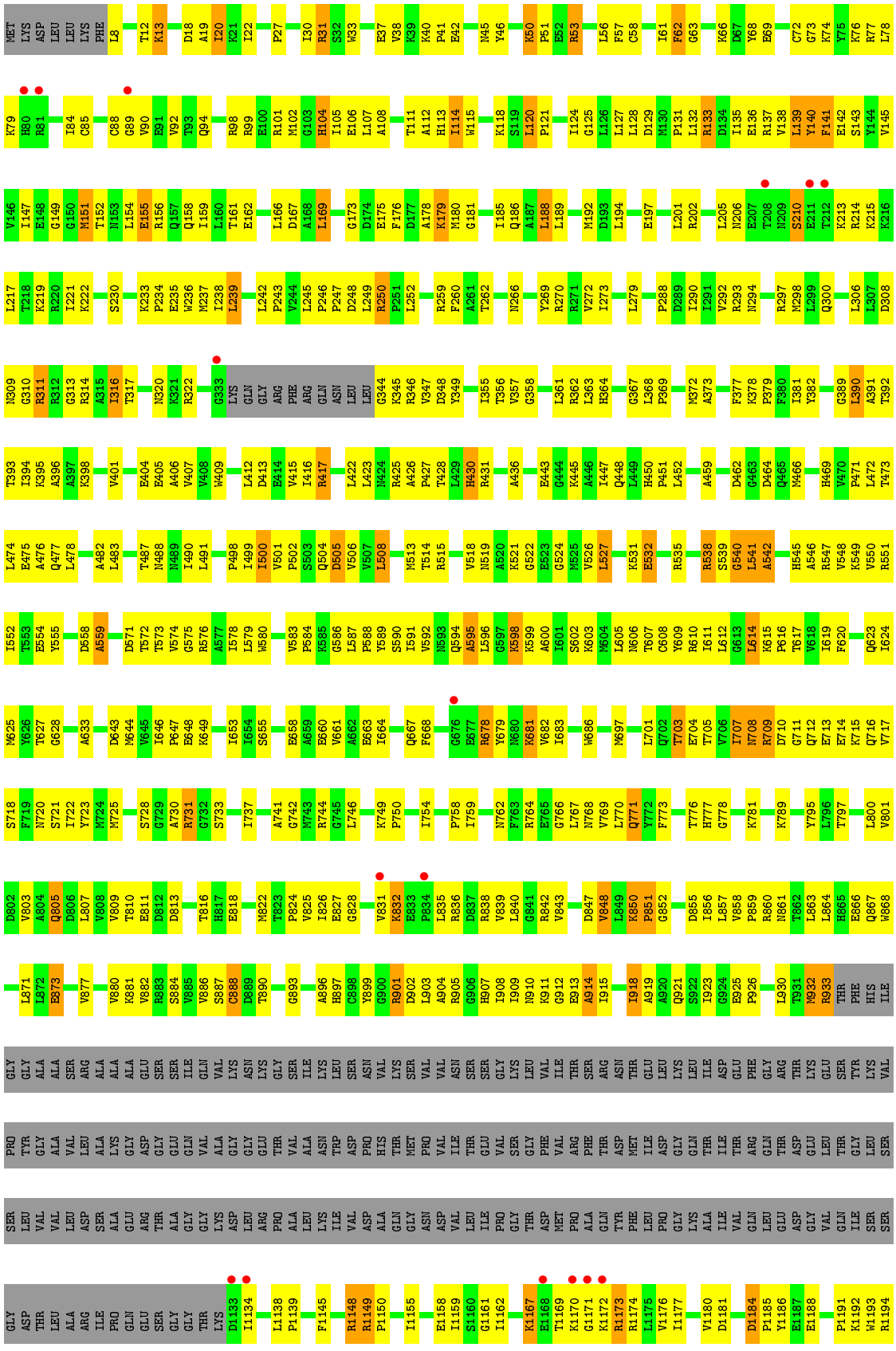


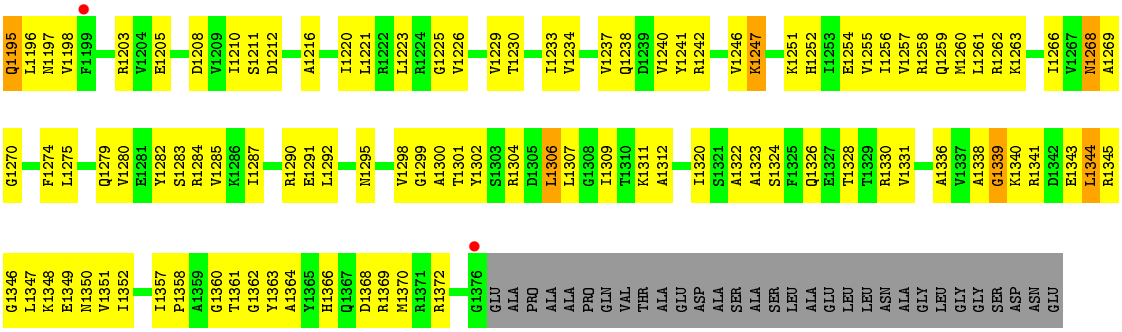
• Molecule 2: Escherichia coli RNA polymerase beta subunit





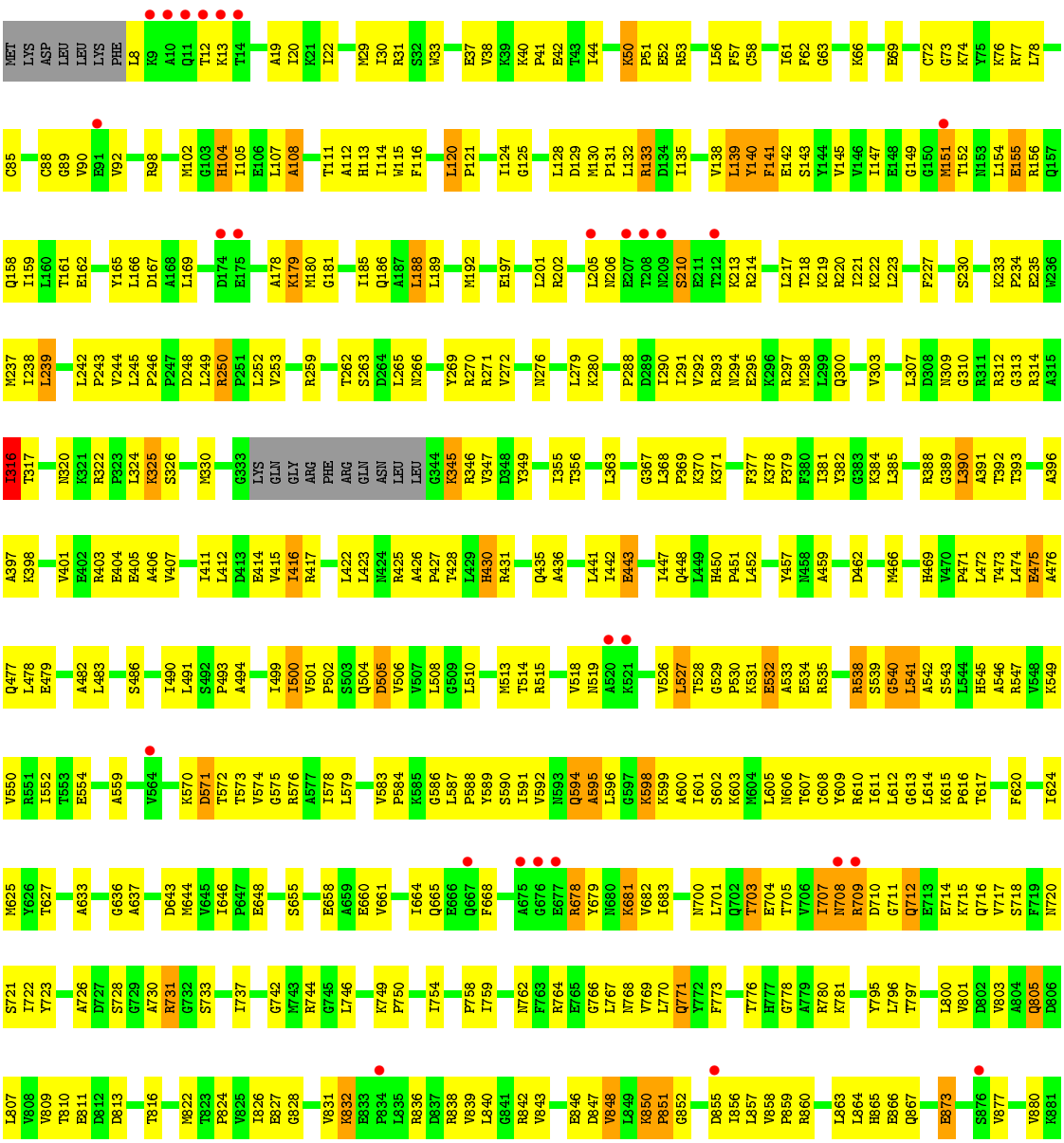
• Molecule 3: Escherichia coli RNA polymerase beta' subunit

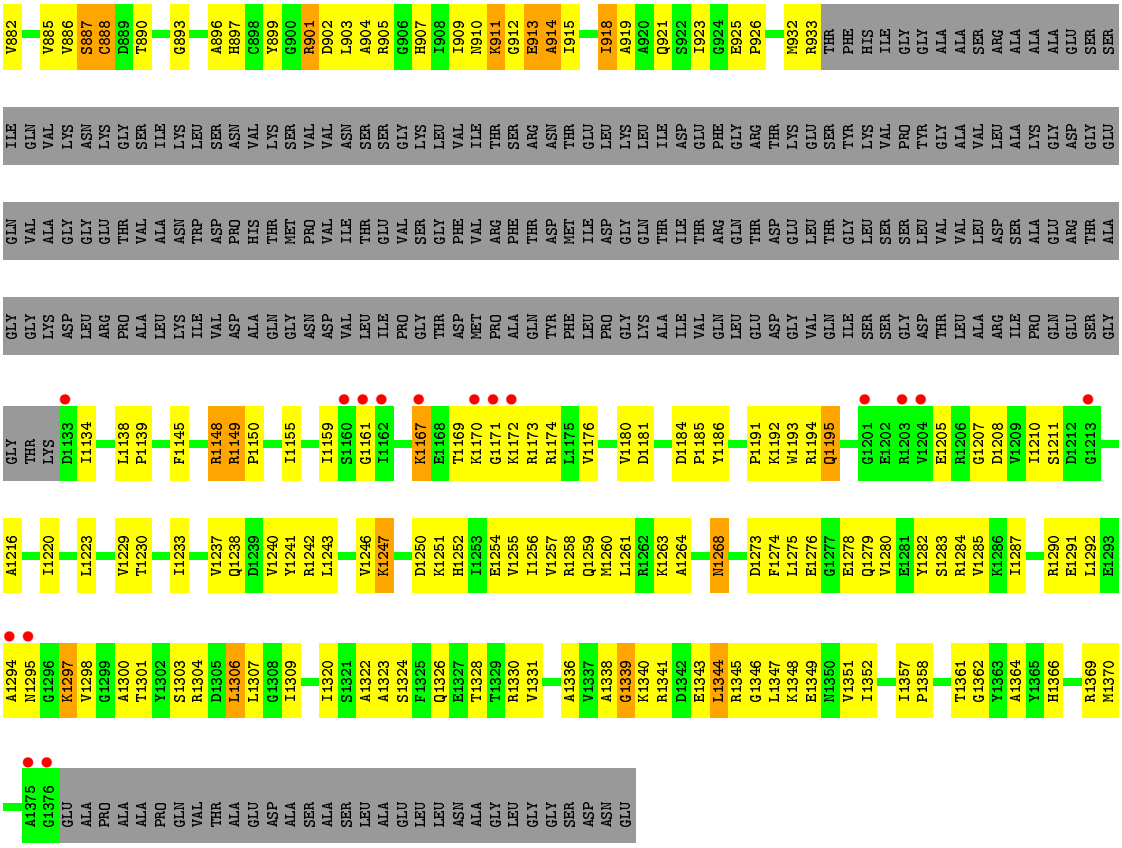




• Molecule 3: Escherichia coli RNA polymerase beta' subunit

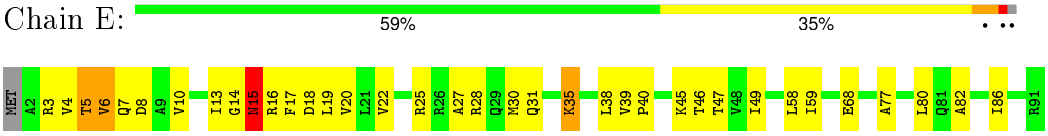
Chain I:





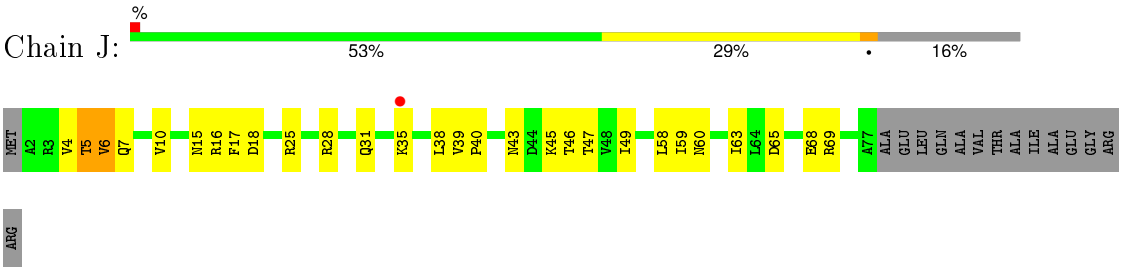
• Molecule 4: Escherichia coli RNA polymerase omega subunit

Chain E:



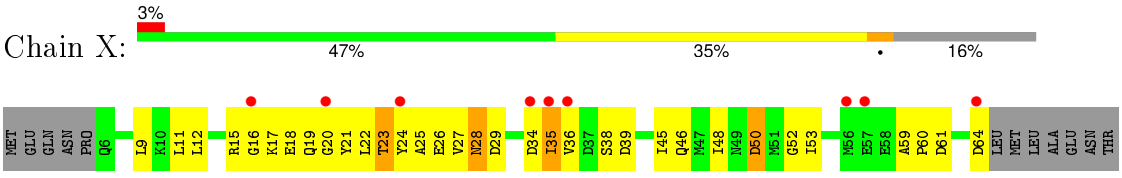
• Molecule 4: Escherichia coli RNA polymerase omega subunit

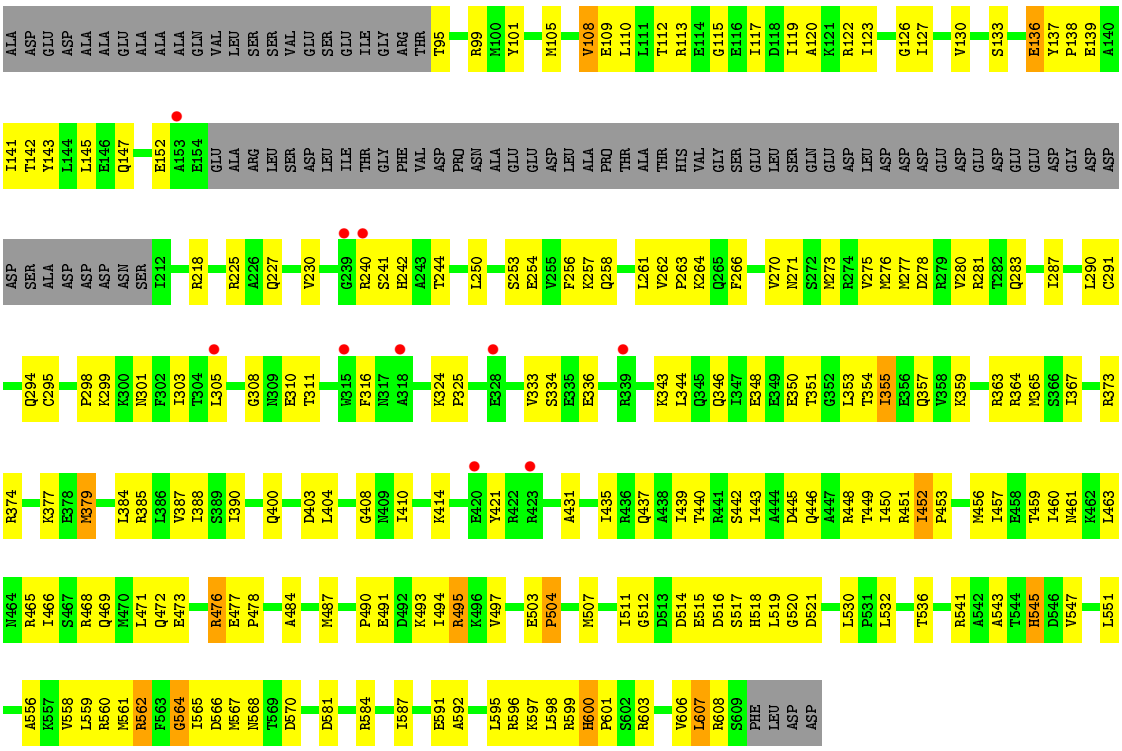
Chain J:



• Molecule 5: Escherichia coli RNA polymerase sigma70 subunit

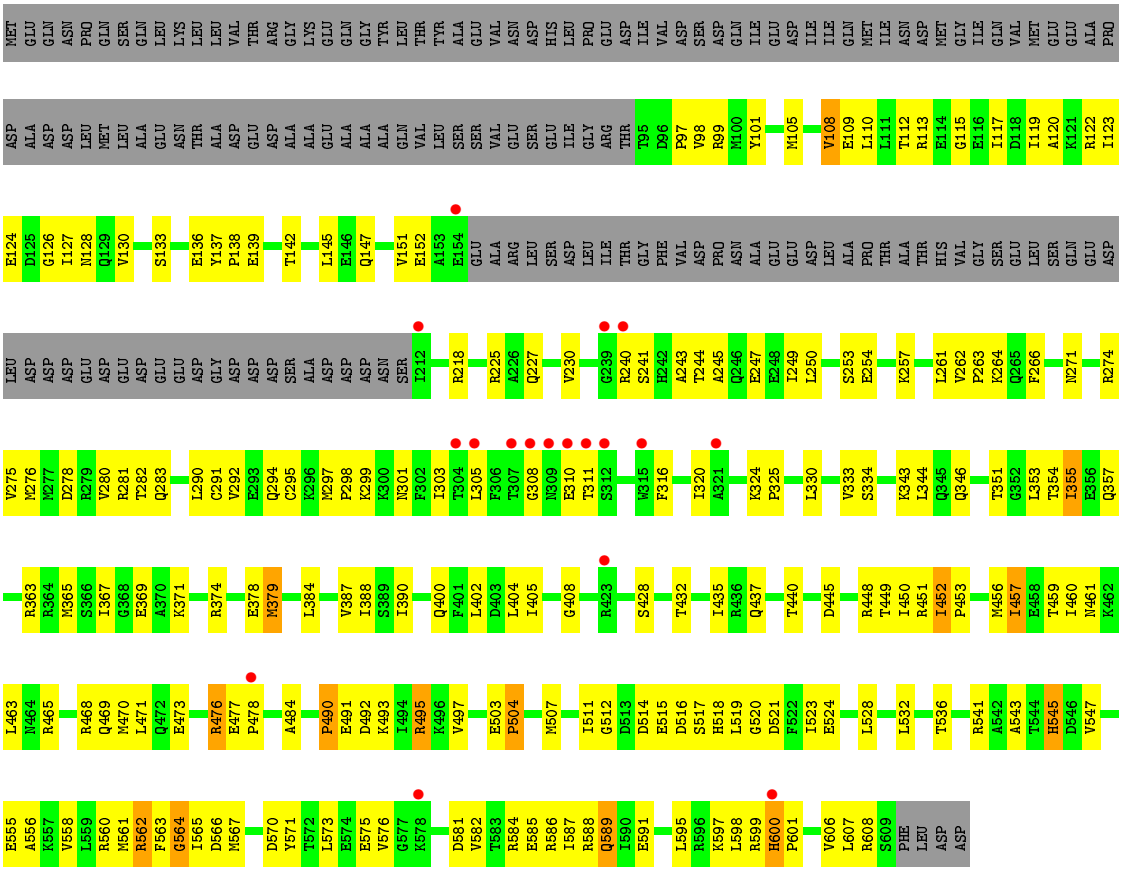
Chain X:





- 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	185.32Å 205.41Å 309.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.94 – 4.20 30.04 – 4.20	Depositor EDS
% Data completeness (in resolution range)	80.4 (29.94-4.20) 70.4 (30.04-4.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.11 (at 4.26Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.244 , 0.322 0.264 , 0.335	Depositor DCC
R_{free} test set	3080 reflections (5.06%)	DCC
Wilson B-factor (Å^2)	159.7	Xtriage
Anisotropy	0.159	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 57.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Outliers	0 of 69537 reflections	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	56129	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.39% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: ZN, 0O2

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	0.36	0/3454
1	B	0.19	0/1725	0.39	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	0.38	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	0.19	0/4253	0.36	0/5719
5	Y	0.20	0/3783	0.36	0/5083
All	All	0.20	0/56889	0.38	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	110	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	91	0
1	F	1775	0	1800	79	0
1	G	1671	0	1706	91	0
2	C	10523	0	10546	600	0
2	H	10523	0	10546	574	0
3	D	9060	0	9257	658	0
3	I	9060	0	9257	591	0
4	E	708	0	719	52	0
4	J	605	0	612	33	0
5	X	4198	0	4250	197	0
5	Y	3732	0	3809	157	0
6	D	2	0	0	0	0
6	I	2	0	0	0	0
7	D	40	10	16	9	0
All	All	56119	10	56822	2973	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 26.

The worst 5 of 2973 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	1.23	1.20
3:D:1173:ARG:HA	3:D:1174:ARG:HB2	1.24	1.18
2:H:488:MET:HB2	2:H:490:GLN:H	1.07	1.11
3:I:850:LYS:HD2	3:I:851:PRO:HD2	1.30	1.09
2:H:1073:LYS:HD3	3:I:462:ASP:HB3	1.28	1.08

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%)	40 (12%)	15 (5%)	3	33
1	B	217/329 (66%)	187 (86%)	22 (10%)	8 (4%)	4	40
1	F	227/329 (69%)	193 (85%)	28 (12%)	6 (3%)	7	47
1	G	213/329 (65%)	186 (87%)	22 (10%)	5 (2%)	8	50
2	C	1333/1342 (99%)	1073 (80%)	208 (16%)	52 (4%)	4	37
2	H	1333/1342 (99%)	1078 (81%)	206 (16%)	49 (4%)	4	40
3	D	1154/1407 (82%)	926 (80%)	182 (16%)	46 (4%)	4	37
3	I	1154/1407 (82%)	925 (80%)	184 (16%)	45 (4%)	4	37
4	E	88/91 (97%)	77 (88%)	6 (7%)	5 (6%)	2	28
4	J	74/91 (81%)	64 (86%)	6 (8%)	4 (5%)	2	30
5	X	511/613 (83%)	450 (88%)	46 (9%)	15 (3%)	6	45
5	Y	454/613 (74%)	409 (90%)	34 (8%)	11 (2%)	7	49
All	All	7079/8222 (86%)	5834 (82%)	984 (14%)	261 (4%)	4	40

5 of 261 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%)	273 (97%)	8 (3%)	51	80
1	B	189/286 (66%)	186 (98%)	3 (2%)	70	88
1	F	197/286 (69%)	194 (98%)	3 (2%)	72	89
1	G	185/286 (65%)	182 (98%)	3 (2%)	70	88

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	C	1150/1157 (99%)	1094 (95%)	56 (5%)	31 69
2	H	1150/1157 (99%)	1097 (95%)	53 (5%)	33 70
3	D	971/1168 (83%)	921 (95%)	50 (5%)	29 68
3	I	971/1168 (83%)	918 (94%)	53 (6%)	27 66
4	E	74/75 (99%)	72 (97%)	2 (3%)	52 80
4	J	65/75 (87%)	65 (100%)	0	100 100
5	X	460/540 (85%)	447 (97%)	13 (3%)	51 80
5	Y	407/540 (75%)	392 (96%)	15 (4%)	41 75
All	All	6100/7024 (87%)	5841 (96%)	259 (4%)	36 72

5 of 259 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	E	8	ASP
2	H	56	VAL
3	I	1149	ARG
5	X	28	ASN
5	X	607	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 91 such sidechains are listed below:

Mol	Chain	Res	Type
5	X	54	GLN
1	G	37	HIS
5	Y	301	ASN
5	X	258	GLN
5	X	437	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	0O2	D	1503	-	31,42,42	2.05	9 (29%)	46,68,68	2.32	10 (21%)

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	0O2	D	1503	-	-	0/29/49/49	0/3/3/3

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	D	1503	0O2	O2'-C2'	-4.34	1.32	1.43
7	D	1503	0O2	C6-C5	-3.06	1.35	1.41
7	D	1503	0O2	O4'-C1'	-2.74	1.37	1.41
7	D	1503	0O2	C2'-C3'	-2.61	1.47	1.53
7	D	1503	0O2	O4'-C4'	-2.53	1.39	1.45

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	D	1503	0O2	C4'-O4'-C1'	-9.46	99.33	109.72
7	D	1503	0O2	PB-O3A-PA	-6.98	113.14	132.73
7	D	1503	0O2	PC-O3C-PD	-5.48	114.28	132.67
7	D	1503	0O2	N3-C2-N1	-3.07	122.77	127.44

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	D	1503	0O2	PB-O3B-PG	-2.43	124.52	132.67

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	D	1503	0O2	9	0

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, 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%)	-0.18	4 (1%) 81 73	0, 55, 172, 230	0
1	B	221/329 (67%)	-0.01	7 (3%) 51 39	0, 86, 193, 260	0
1	F	229/329 (69%)	0.02	13 (5%) 27 20	2, 123, 212, 293	0
1	G	217/329 (65%)	0.06	5 (2%) 64 53	5, 113, 204, 271	0
2	C	1335/1342 (99%)	-0.28	23 (1%) 73 63	0, 38, 168, 304	0
2	H	1335/1342 (99%)	-0.11	40 (2%) 54 42	0, 78, 206, 346	0
3	D	1160/1407 (82%)	-0.20	18 (1%) 74 65	0, 28, 152, 297	0
3	I	1160/1407 (82%)	-0.08	43 (3%) 45 35	0, 54, 183, 316	0
4	E	90/91 (98%)	-0.20	0 100 100	0, 33, 116, 167	0
4	J	76/91 (83%)	0.08	1 (1%) 79 71	12, 83, 181, 230	0
5	X	517/613 (84%)	-0.09	19 (3%) 45 35	0, 98, 238, 341	0
5	Y	458/613 (74%)	-0.07	18 (3%) 43 33	1, 100, 216, 296	0
All	All	7121/8222 (86%)	-0.13	191 (2%) 58 47	0, 63, 198, 346	0

The worst 5 of 191 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	I	10	ALA	9.8
2	H	982	GLY	9.5
2	H	981	ALA	8.6
3	I	521	LYS	8.5
2	H	983	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
6	ZN	I	1502	1/1	0.99	0.17	-0.72	49,49,49,49	0
6	ZN	D	1502	1/1	0.97	0.18	-0.94	8,8,8,8	0
6	ZN	D	1501	1/1	0.99	0.05	-1.65	54,54,54,54	0
6	ZN	I	1501	1/1	0.97	0.04	-1.83	60,60,60,60	0
7	0O2	D	1503	40/40	0.90	0.16	-1.91	20,20,20,20	0

6.5 Other polymers (i)

There are no such residues in this entry.