



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 22, 2024 – 10:49 AM EST

PDB ID : 4YG2
Title : X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme
Authors : Murakami, K.S.
Deposited on : 2015-02-25
Resolution : 3.70 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
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

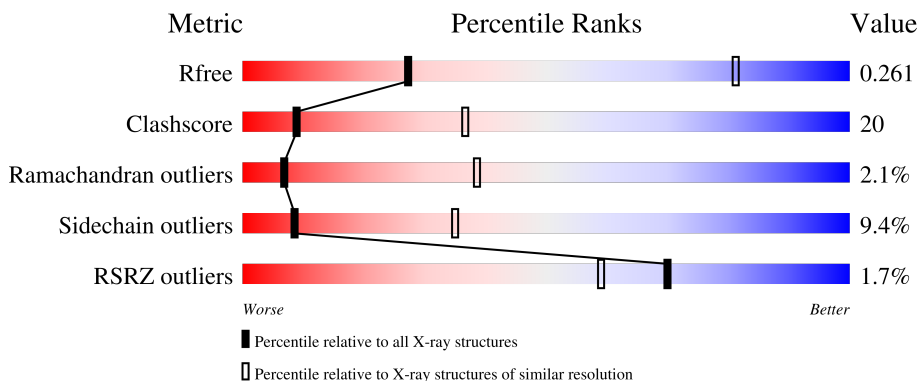
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.70 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	1049 (3.88-3.52)
Clashscore	141614	1027 (3.86-3.54)
Ramachandran outliers	138981	1069 (3.88-3.52)
Sidechain outliers	138945	1065 (3.88-3.52)
RSRZ outliers	127900	1578 (3.90-3.50)

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 $\leq 5\%$. 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	
1	B	329	
1	G	329	
1	H	329	
2	C	1342	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	I	1342	
3	D	1407	
3	J	1407	
4	E	91	
4	K	91	
5	F	613	
5	L	613	

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 55741 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	319	Total	C	N	O	S	0	0	0
			2490	1557	439	486	8			
1	B	217	Total	C	N	O	S	0	0	0
			1677	1047	295	329	6			
1	G	227	Total	C	N	O	S	0	0	0
			1755	1093	311	345	6			
1	H	216	Total	C	N	O	S	0	0	0
			1662	1038	292	326	6			

- 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	Total	C	N	O	S	0	0	0
			10570	6631	1841	2055	43			
2	I	1340	Total	C	N	O	S	0	0	0
			10566	6629	1840	2054	43			

- 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	1166	Total	C	N	O	S	0	0	0
			9060	5697	1621	1696	46			
3	J	1155	Total	C	N	O	S	0	0	0
			9001	5659	1612	1684	46			

- 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	89	Total	C	N	O	S	0	0	0
			691	421	129	140	1			
4	K	79	Total	C	N	O	S	0	0	0
			627	382	118	126	1			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	468	Total	C	N	O	S	0	0	0
			3813	2389	678	723	23			
5	L	469	Total	C	N	O	S	0	0	0
			3821	2393	679	726	23			

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	1	Total	Mg	0	0
			1	1		
6	D	1	Total	Mg	0	0
			1	1		
6	I	1	Total	Mg	0	0
			1	1		
6	J	1	Total	Mg	0	0
			1	1		

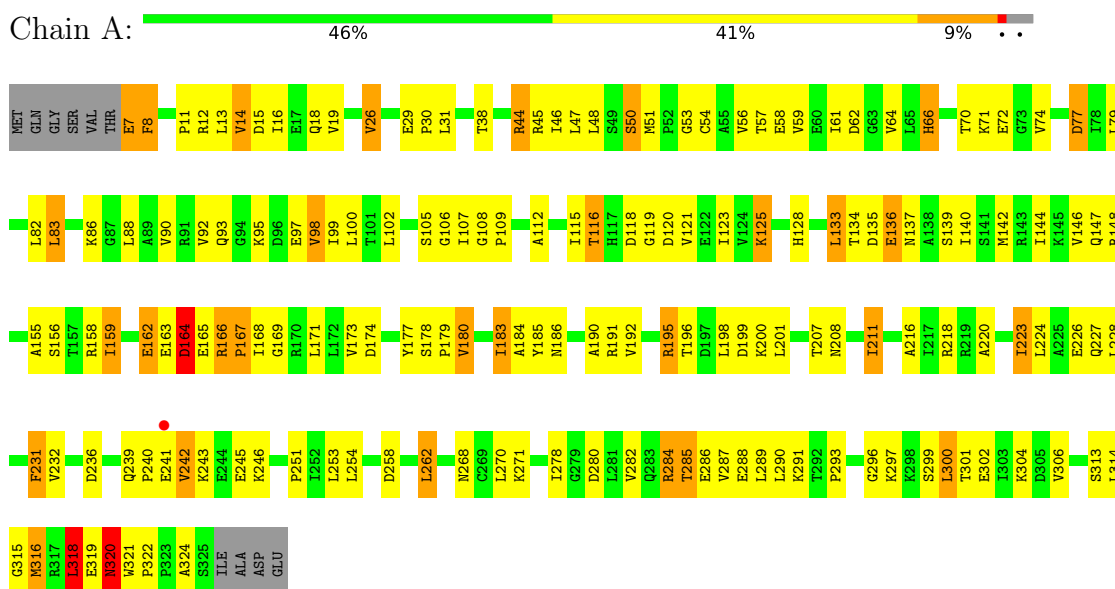
- Molecule 7 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	D	2	Total	Zn	0	0
			2	2		
7	J	2	Total	Zn	0	0
			2	2		

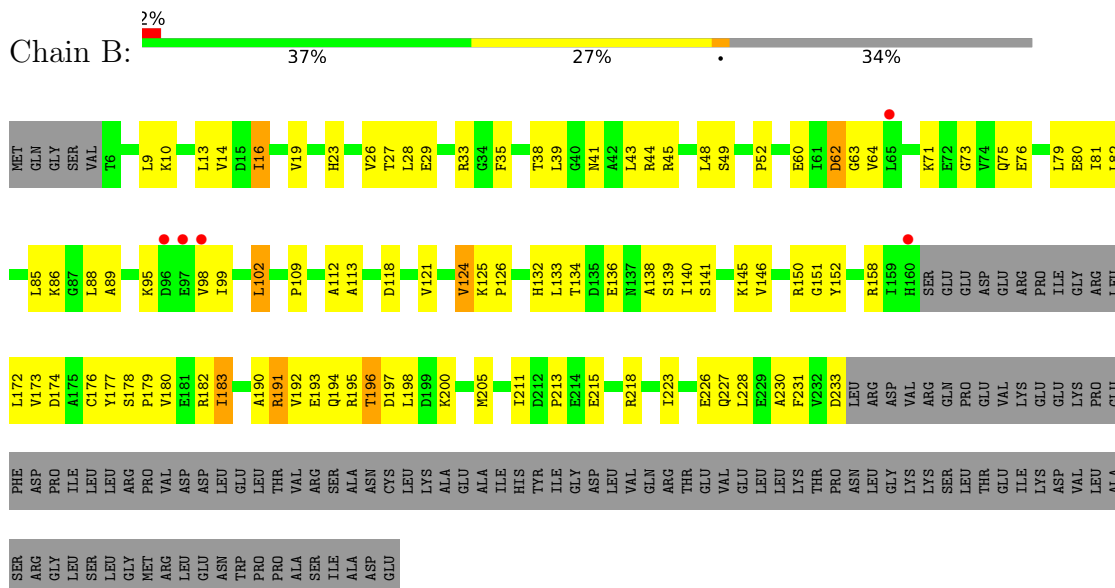
3 Residue-property plots [i](#)

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.

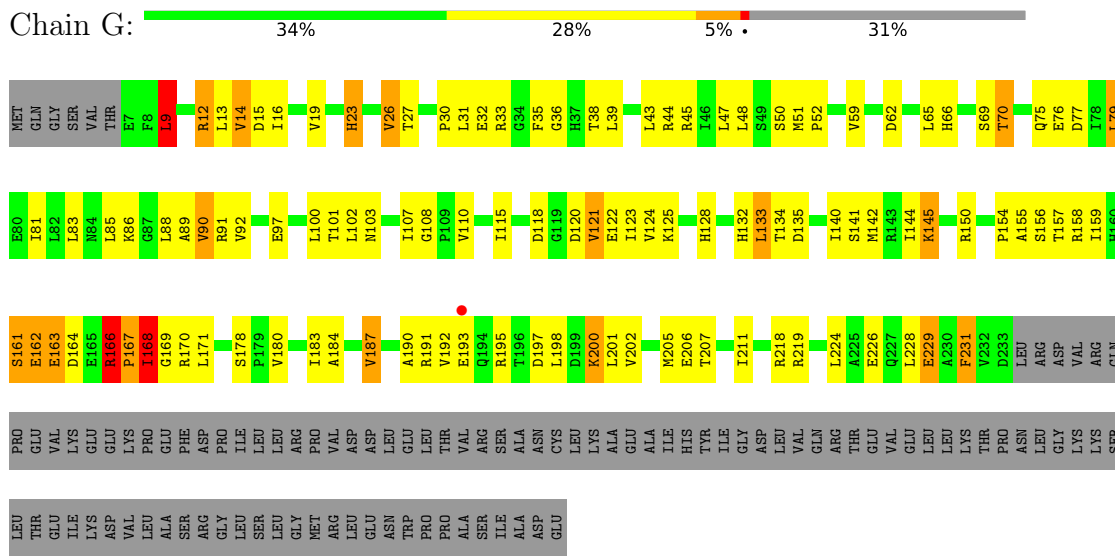
- Molecule 1: DNA-directed RNA polymerase subunit alpha



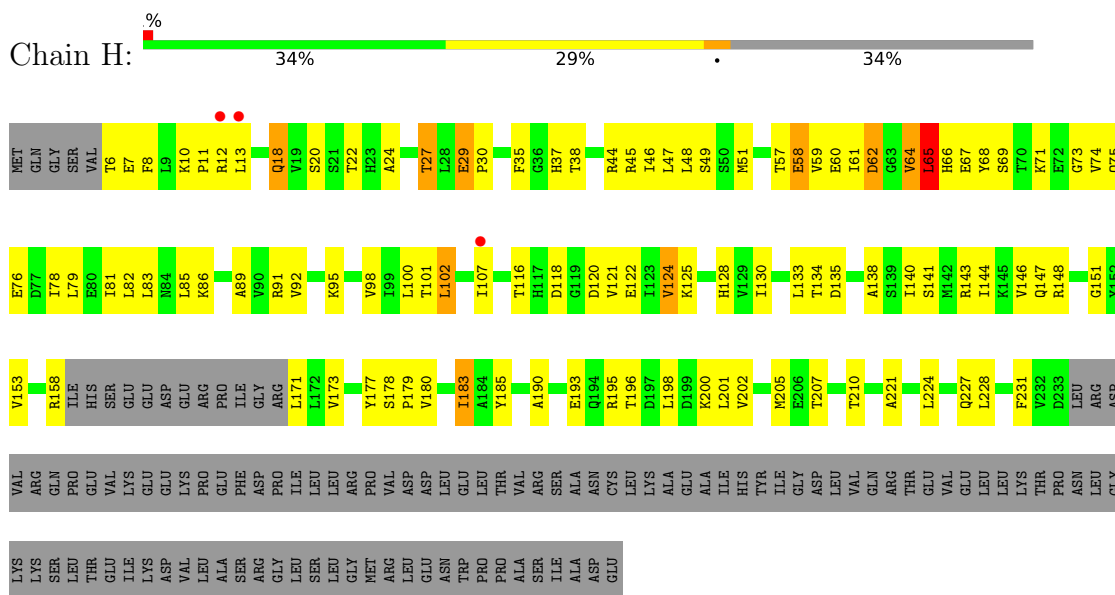
- Molecule 1: DNA-directed RNA polymerase subunit alpha



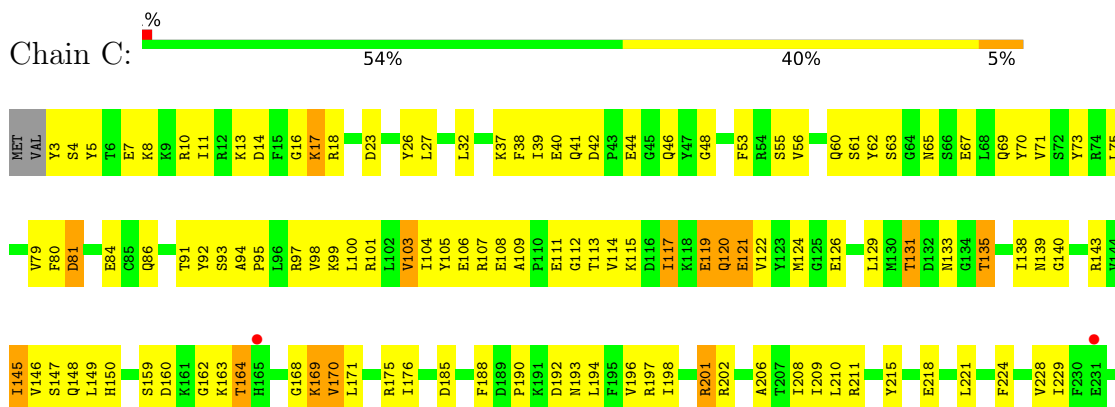
- Molecule 1: DNA-directed RNA polymerase subunit alpha

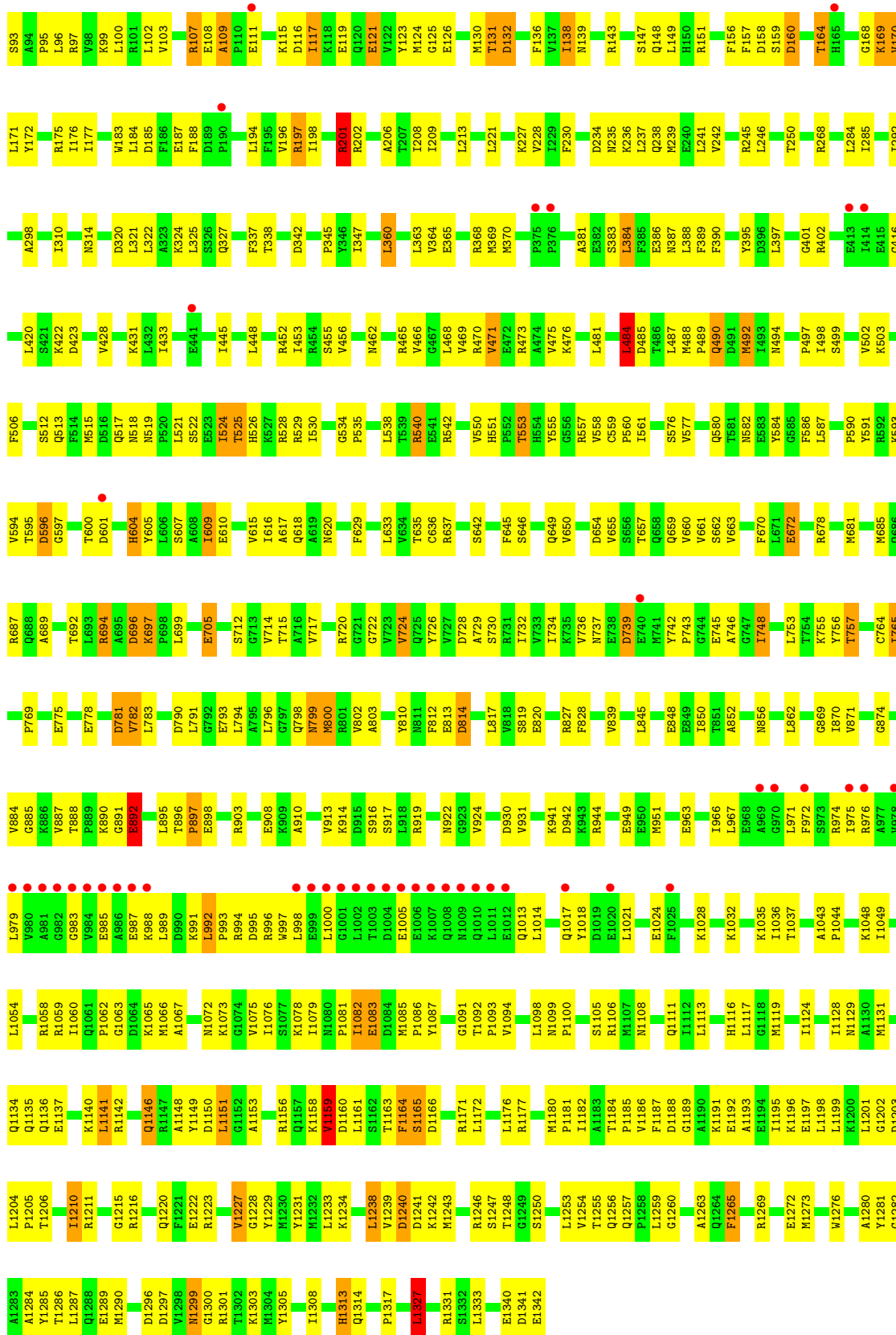


- Molecule 1: DNA-directed RNA polymerase subunit alpha

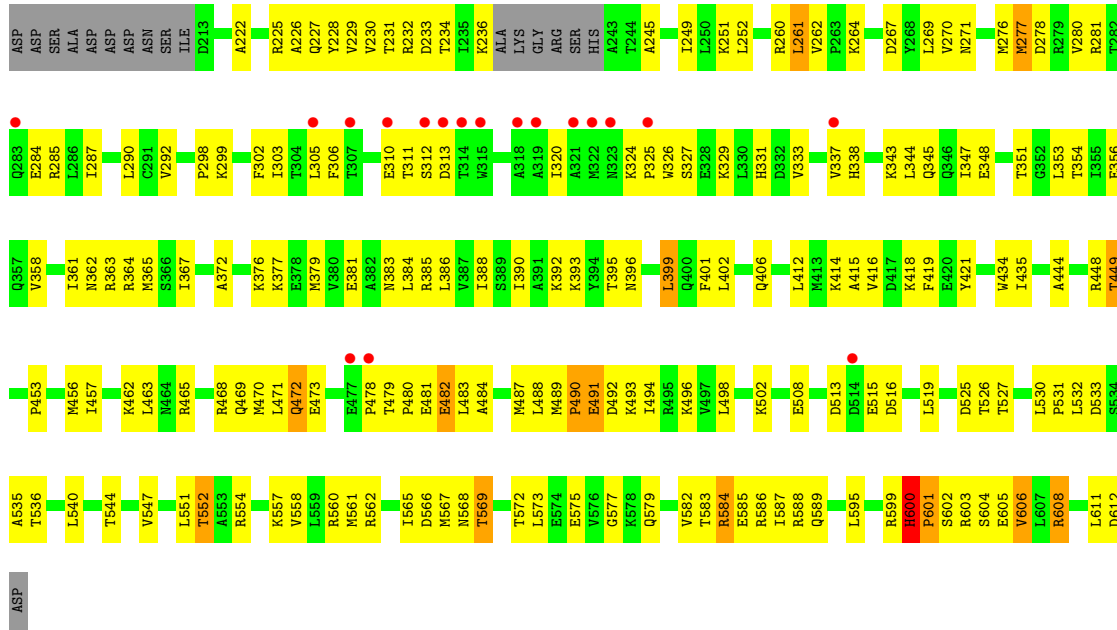


- Molecule 2: DNA-directed RNA polymerase subunit beta

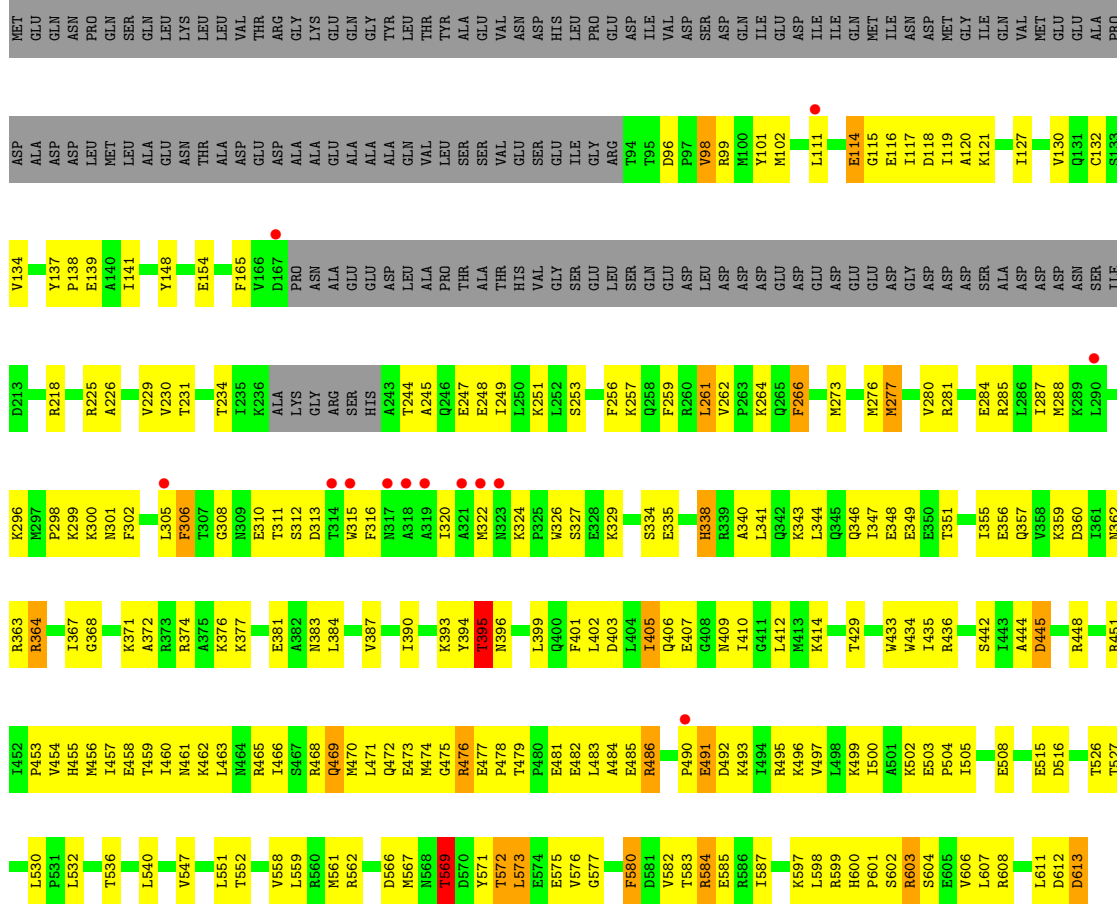




• Molecule 3: DNA-directed RNA polymerase subunit beta'



● Molecule 5: RNA polymerase sigma factor RpoD



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	187.31Å 205.90Å 309.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.98 – 3.70 29.98 – 3.70	Depositor EDS
% Data completeness (in resolution range)	96.8 (29.98-3.70) 92.5 (29.98-3.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.97 (at 3.75Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.191 , 0.260 0.194 , 0.261	Depositor DCC
R_{free} test set	6183 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	116.2	Xtrriage
Anisotropy	0.165	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 61.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	55741	wwPDB-VP
Average B, all atoms (Å ²)	140.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.66% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: 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	0.62	0/2524	0.81	2/3421 (0.1%)
1	B	0.58	0/1697	0.84	0/2300
1	G	0.57	0/1777	0.78	0/2408
1	H	0.51	0/1681	0.83	3/2278 (0.1%)
2	C	0.69	2/10739 (0.0%)	0.86	7/14489 (0.0%)
2	I	0.57	0/10735	0.77	3/14484 (0.0%)
3	D	0.70	0/9200	0.90	12/12423 (0.1%)
3	J	0.64	0/9140	0.83	5/12341 (0.0%)
4	E	0.66	0/693	0.82	0/935
4	K	0.61	0/629	0.77	0/847
5	F	0.56	0/3864	0.81	3/5194 (0.1%)
5	L	0.52	0/3872	0.71	1/5205 (0.0%)
All	All	0.63	2/56551 (0.0%)	0.82	36/76325 (0.0%)

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
2	C	0	2
2	I	0	2
3	D	0	2
3	J	0	3
5	F	0	2
All	All	0	11

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	949	GLU	CB-CG	6.52	1.64	1.52

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	838	CYS	CB-SG	-6.02	1.72	1.82

The worst 5 of 36 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	I	484	LEU	CA-CB-CG	9.03	136.06	115.30
2	C	796	LEU	CB-CG-CD2	-8.02	97.36	111.00
2	C	1161	LEU	CA-CB-CG	-7.43	98.22	115.30
3	D	268	LEU	CA-CB-CG	-7.05	99.09	115.30
2	C	544	GLY	N-CA-C	-6.83	96.01	113.10

There are no chirality outliers.

5 of 11 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	109	ALA	Peptide
2	C	236	LYS	Peptide
3	D	1184	ASP	Peptide
3	D	901	ARG	Peptide
5	F	600	HIS	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	2490	0	2542	129	0
1	B	1677	0	1703	75	0
1	G	1755	0	1773	96	0
1	H	1662	0	1687	79	0
2	C	10570	0	10582	460	0
2	I	10566	0	10576	403	1
3	D	9060	0	9209	451	1
3	J	9001	0	9168	412	0
4	E	691	0	695	24	0
4	K	627	0	634	27	0
5	F	3813	0	3880	150	0
5	L	3821	0	3884	163	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	C	1	0	0	0	0
6	D	1	0	0	0	0
6	I	1	0	0	0	0
6	J	1	0	0	0	0
7	D	2	0	0	0	0
7	J	2	0	0	0	0
All	All	55741	0	56333	2271	1

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 20.

The worst 5 of 2271 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)
1:G:190:ALA:HB2	1:G:200:LYS:HB2	1.42	1.01
1:A:296:GLY:H	1:A:299:SER:HB2	1.26	1.00
5:L:484:ALA:HB1	5:L:491:GLU:HB2	1.41	0.99
3:D:210:SER:HB2	3:D:213:LYS:HB2	1.47	0.97
3:D:1183:SER:HA	3:J:206:ASN:HD21	1.29	0.96

All (1) 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)
3:D:712:GLN:NE2	2:I:862:LEU:O[4_545]	2.14	0.06

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	317/329 (96%)	246 (78%)	53 (17%)	18 (6%)	1 19

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	213/329 (65%)	193 (91%)	16 (8%)	4 (2%)	8	40
1	G	225/329 (68%)	198 (88%)	18 (8%)	9 (4%)	3	26
1	H	212/329 (64%)	196 (92%)	12 (6%)	4 (2%)	8	40
2	C	1338/1342 (100%)	1205 (90%)	112 (8%)	21 (2%)	9	43
2	I	1338/1342 (100%)	1197 (90%)	120 (9%)	21 (2%)	9	43
3	D	1162/1407 (83%)	1030 (89%)	105 (9%)	27 (2%)	6	36
3	J	1151/1407 (82%)	1027 (89%)	98 (8%)	26 (2%)	6	36
4	E	87/91 (96%)	81 (93%)	5 (6%)	1 (1%)	14	50
4	K	77/91 (85%)	71 (92%)	3 (4%)	3 (4%)	3	27
5	F	462/613 (75%)	423 (92%)	32 (7%)	7 (2%)	10	44
5	L	463/613 (76%)	425 (92%)	30 (6%)	8 (2%)	9	42
All	All	7045/8222 (86%)	6292 (89%)	604 (9%)	149 (2%)	7	38

5 of 149 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	14	VAL
1	A	107	ILE
1	A	136	GLU
1	A	162	GLU
1	A	167	PRO

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	278/286 (97%)	227 (82%)	51 (18%)	1	10
1	B	186/286 (65%)	171 (92%)	15 (8%)	11	41
1	G	193/286 (68%)	164 (85%)	29 (15%)	3	18
1	H	183/286 (64%)	170 (93%)	13 (7%)	14	45
2	C	1155/1157 (100%)	1053 (91%)	102 (9%)	10	38

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	I	1154/1157 (100%)	1053 (91%)	101 (9%)	10	38
3	D	962/1168 (82%)	876 (91%)	86 (9%)	9	37
3	J	960/1168 (82%)	873 (91%)	87 (9%)	9	36
4	E	72/75 (96%)	63 (88%)	9 (12%)	4	23
4	K	67/75 (89%)	60 (90%)	7 (10%)	7	30
5	F	417/540 (77%)	385 (92%)	32 (8%)	13	43
5	L	418/540 (77%)	379 (91%)	39 (9%)	9	35
All	All	6045/7024 (86%)	5474 (91%)	571 (9%)	8	35

5 of 571 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	J	376	LEU
3	J	545	HIS
3	J	374	LEU
3	J	1285	VAL
3	D	416	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 64 such sidechains are listed below:

Mol	Chain	Res	Type
3	J	1295	ASN
5	L	258	GLN
5	F	147	GLN
5	F	131	GLN
5	L	301	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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 [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 8 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 [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 [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	319/329 (96%)	-0.36	1 (0%) 94 90	84, 129, 182, 196	0
1	B	217/329 (65%)	-0.18	5 (2%) 60 48	91, 164, 203, 214	0
1	G	227/329 (68%)	-0.28	1 (0%) 92 88	126, 154, 181, 218	0
1	H	216/329 (65%)	-0.10	3 (1%) 75 64	118, 174, 198, 207	0
2	C	1340/1342 (99%)	-0.38	18 (1%) 77 67	58, 114, 218, 253	0
2	I	1340/1342 (99%)	-0.19	44 (3%) 46 35	93, 148, 222, 378	0
3	D	1166/1407 (82%)	-0.39	2 (0%) 95 93	63, 107, 168, 218	0
3	J	1155/1407 (82%)	-0.27	14 (1%) 79 69	85, 128, 189, 228	0
4	E	89/91 (97%)	-0.59	0 100 100	67, 116, 140, 152	0
4	K	79/91 (86%)	-0.27	1 (1%) 77 67	99, 139, 205, 222	0
5	F	468/613 (76%)	-0.17	19 (4%) 37 27	93, 160, 273, 305	0
5	L	469/613 (76%)	-0.15	13 (2%) 53 40	114, 172, 279, 307	0
All	All	7085/8222 (86%)	-0.28	121 (1%) 70 59	58, 135, 209, 378	0

The worst 5 of 121 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	I	982	GLY	11.1
5	F	319	ALA	6.7
2	I	983	GLY	5.3
5	L	318	ALA	5.0
2	I	981	ALA	4.9

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
6	MG	C	1401	1/1	0.95	0.51	94,94,94,94	0
6	MG	D	2001	1/1	0.96	0.35	94,94,94,94	0
6	MG	I	1401	1/1	0.96	0.44	94,94,94,94	0
6	MG	J	2001	1/1	0.96	0.25	94,94,94,94	0
7	ZN	J	2002	1/1	0.97	0.04	105,105,105,105	0
7	ZN	J	2003	1/1	0.98	0.11	94,94,94,94	0
7	ZN	D	2002	1/1	0.99	0.09	113,113,113,113	0
7	ZN	D	2003	1/1	0.99	0.17	94,94,94,94	0

6.5 Other polymers [i](#)

There are no such residues in this entry.