



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 08:26 am GMT

PDB ID : 4LK0
Title : Crystal Structure Analysis of the E.coli holoenzyme/T7 Gp2 complex
Authors : Bae, B.; Darst, S.A.
Deposited on : 2013-07-05
Resolution : 3.91 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
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

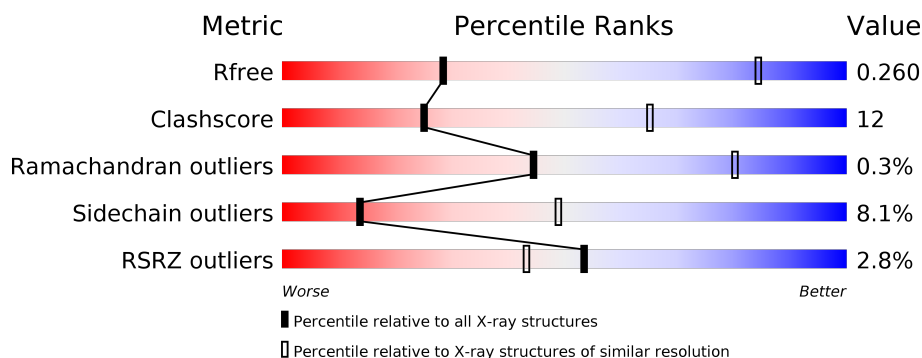
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.91 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	1012 (4.24-3.60)
Clashscore	112137	1108 (4.24-3.60)
Ramachandran outliers	110173	1067 (4.24-3.60)
Sidechain outliers	110143	1058 (4.24-3.60)
RSRZ outliers	101464	1025 (4.24-3.60)

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 $\leq 5\%$. 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	239	<div> <div>2%</div> <div> <div></div> <div>60%</div> <div>32%</div> <div>6%</div> </div> </div>
1	B	239	<div> <div>6%</div> <div> <div></div> <div>64%</div> <div>24%</div> <div>8%</div> </div> </div>
1	G	239	<div> <div>5%</div> <div> <div></div> <div>61%</div> <div>33%</div> <div>5%</div> </div> </div>
1	H	239	<div> <div>6%</div> <div> <div></div> <div>60%</div> <div>27%</div> <div>9%</div> </div> </div>
2	C	1342	<div> <div>3%</div> <div> <div></div> <div>68%</div> <div>29%</div> <div></div> </div> </div>
2	I	1342	<div> <div>4%</div> <div> <div></div> <div>69%</div> <div>28%</div> <div></div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	D	1407	<p>2% 64% 28% • •</p>
3	J	1407	<p>1% 63% 28% • 6%</p>
4	E	91	<p>76% 20% • •</p>
4	K	91	<p>70% 14% • 13%</p>
5	F	522	<p>2% 59% 27% • 10%</p>
5	L	522	<p>2% 57% 30% • 10%</p>
6	M	64	<p>2% 53% 25% • 20%</p>
6	N	64	<p>5% 48% 30% • 20%</p>

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 58505 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
1	A	224	Total	C	N	O	S	0	0	0
			1730	1076	308	340	6			
1	B	220	Total	C	N	O	S	0	0	0
			1687	1053	298	330	6			
1	G	228	Total	C	N	O	S	0	0	0
			1750	1088	312	344	6			
1	H	217	Total	C	N	O	S	0	0	0
			1667	1041	293	327	6			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	235	GLU	-	EXPRESSION TAG	UNP C9QXI7
A	236	VAL	-	EXPRESSION TAG	UNP C9QXI7
A	237	LEU	-	EXPRESSION TAG	UNP C9QXI7
A	238	PHE	-	EXPRESSION TAG	UNP C9QXI7
A	239	GLN	-	EXPRESSION TAG	UNP C9QXI7
B	235	GLU	-	EXPRESSION TAG	UNP C9QXI7
B	236	VAL	-	EXPRESSION TAG	UNP C9QXI7
B	237	LEU	-	EXPRESSION TAG	UNP C9QXI7
B	238	PHE	-	EXPRESSION TAG	UNP C9QXI7
B	239	GLN	-	EXPRESSION TAG	UNP C9QXI7
G	235	GLU	-	EXPRESSION TAG	UNP C9QXI7
G	236	VAL	-	EXPRESSION TAG	UNP C9QXI7
G	237	LEU	-	EXPRESSION TAG	UNP C9QXI7
G	238	PHE	-	EXPRESSION TAG	UNP C9QXI7
G	239	GLN	-	EXPRESSION TAG	UNP C9QXI7
H	235	GLU	-	EXPRESSION TAG	UNP C9QXI7
H	236	VAL	-	EXPRESSION TAG	UNP C9QXI7
H	237	LEU	-	EXPRESSION TAG	UNP C9QXI7
H	238	PHE	-	EXPRESSION TAG	UNP C9QXI7
H	239	GLN	-	EXPRESSION TAG	UNP C9QXI7

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
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
3	D	1345	Total	C	N	O	S	0	0	0
			10447	6560	1864	1974	49			
3	J	1325	Total	C	N	O	S	0	0	0
			10295	6470	1831	1945	49			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
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	470	Total	C	N	O	S	0	0	0
			3822	2394	680	725	23			
5	L	469	Total	C	N	O	S	0	0	0
			3821	2393	679	726	23			

- Molecule 6 is a protein called Bacterial RNA polymerase inhibitor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	M	51	Total	C	N	O	S	0	0	0
			413	268	65	79	1			
6	N	51	Total	C	N	O	S	0	0	0
			413	268	65	79	1			

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

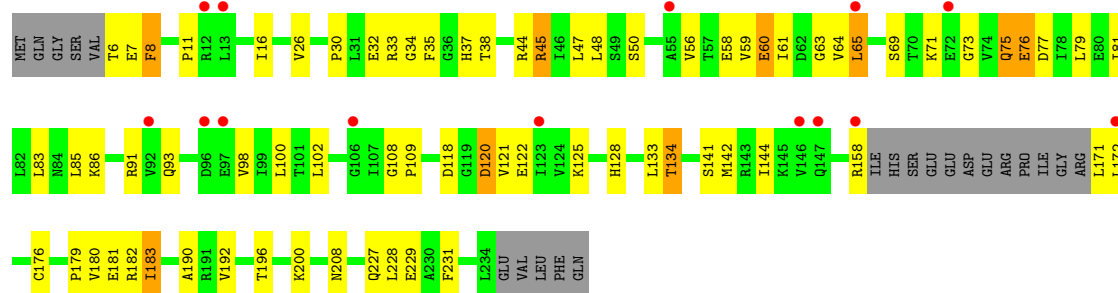
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	J	1	Total	Mg	0	0
			1	1		
7	D	1	Total	Mg	0	0
			1	1		

- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

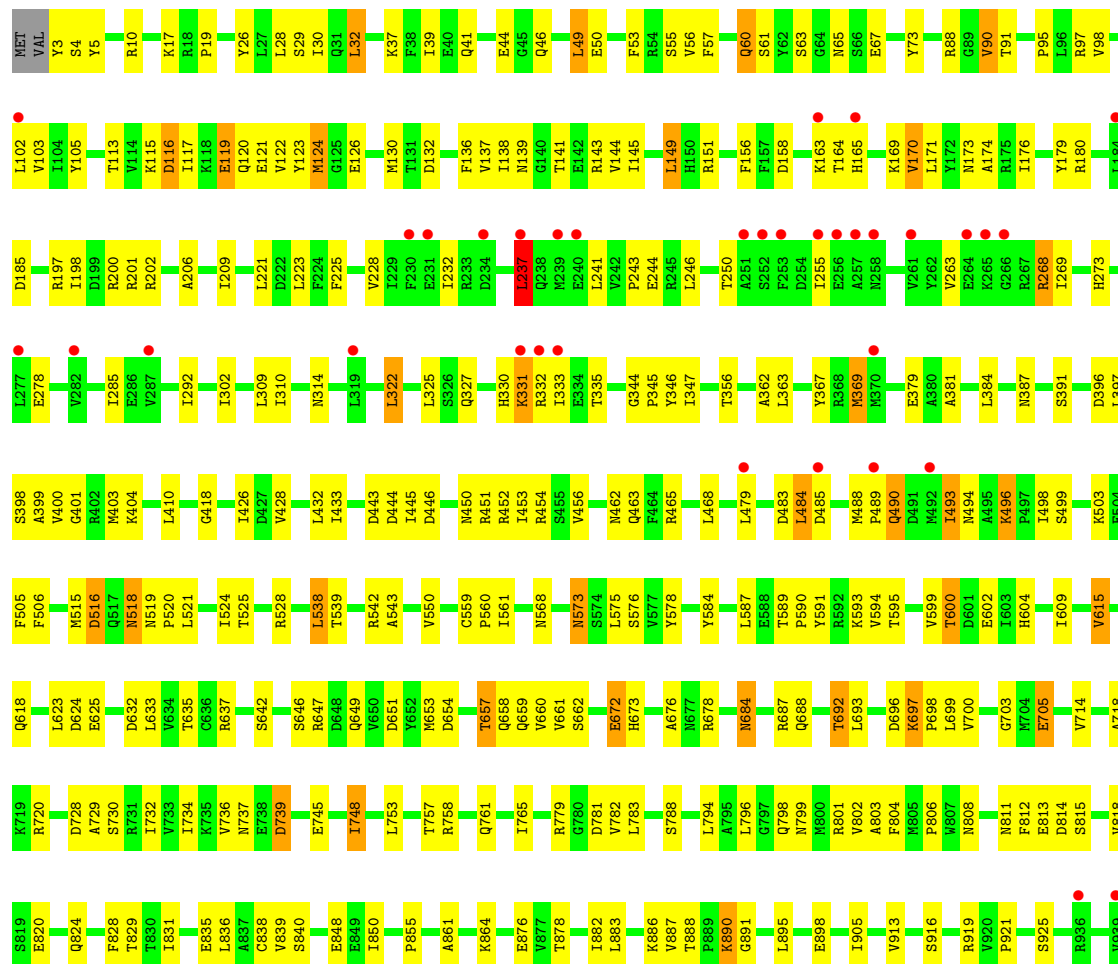
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	J	2	Total	Zn	0	0
			2	2		
8	D	2	Total	Zn	0	0
			2	2		



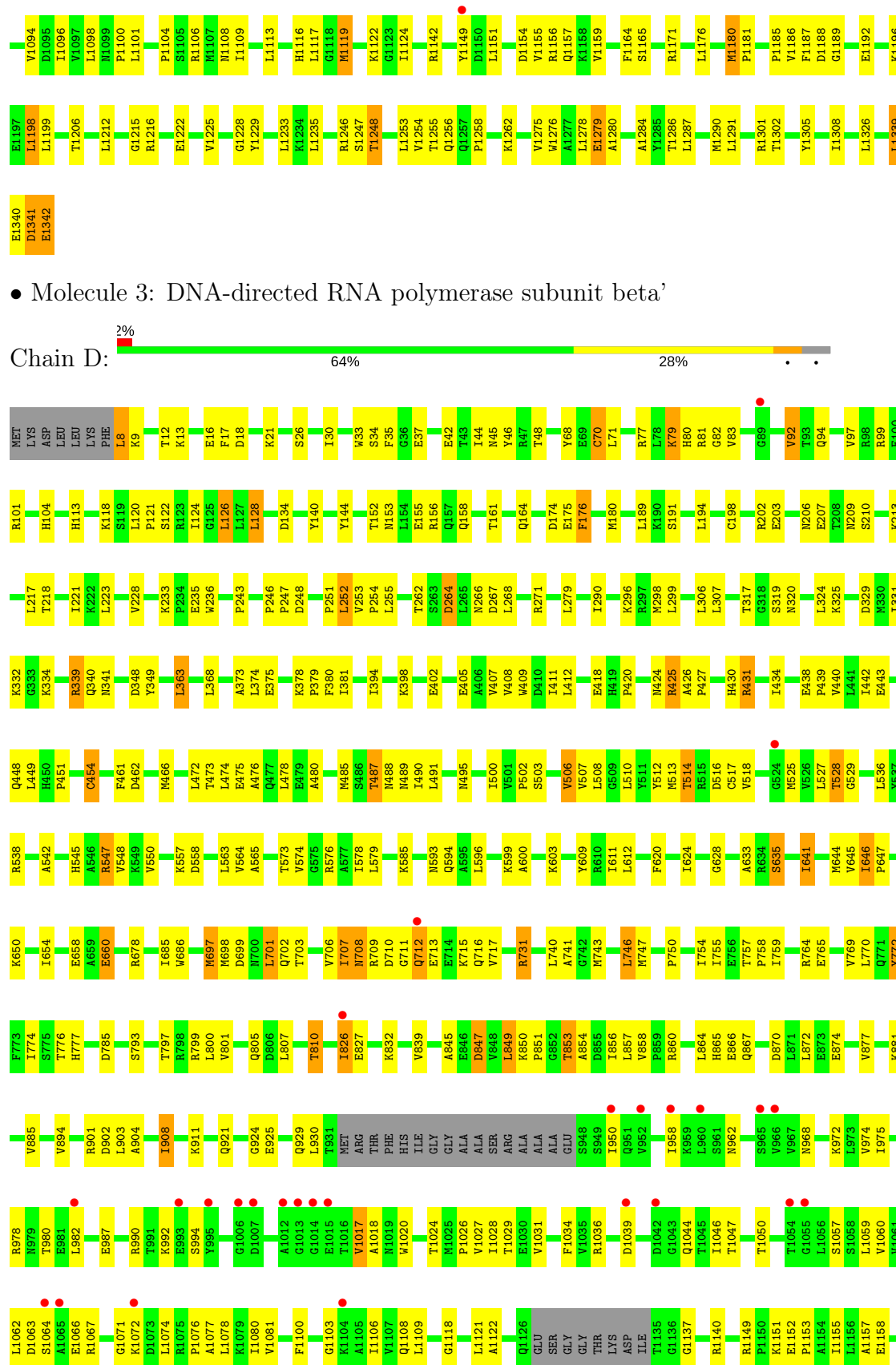
• Molecule 1: DNA-directed RNA polymerase subunit alpha



• Molecule 2: DNA-directed RNA polymerase subunit beta

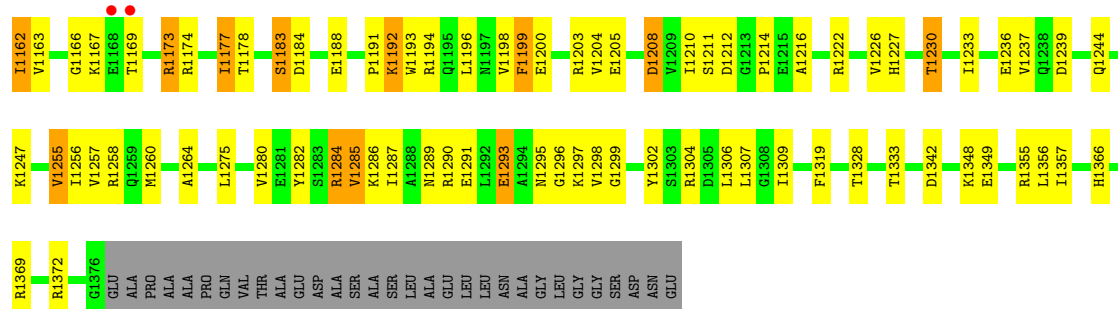




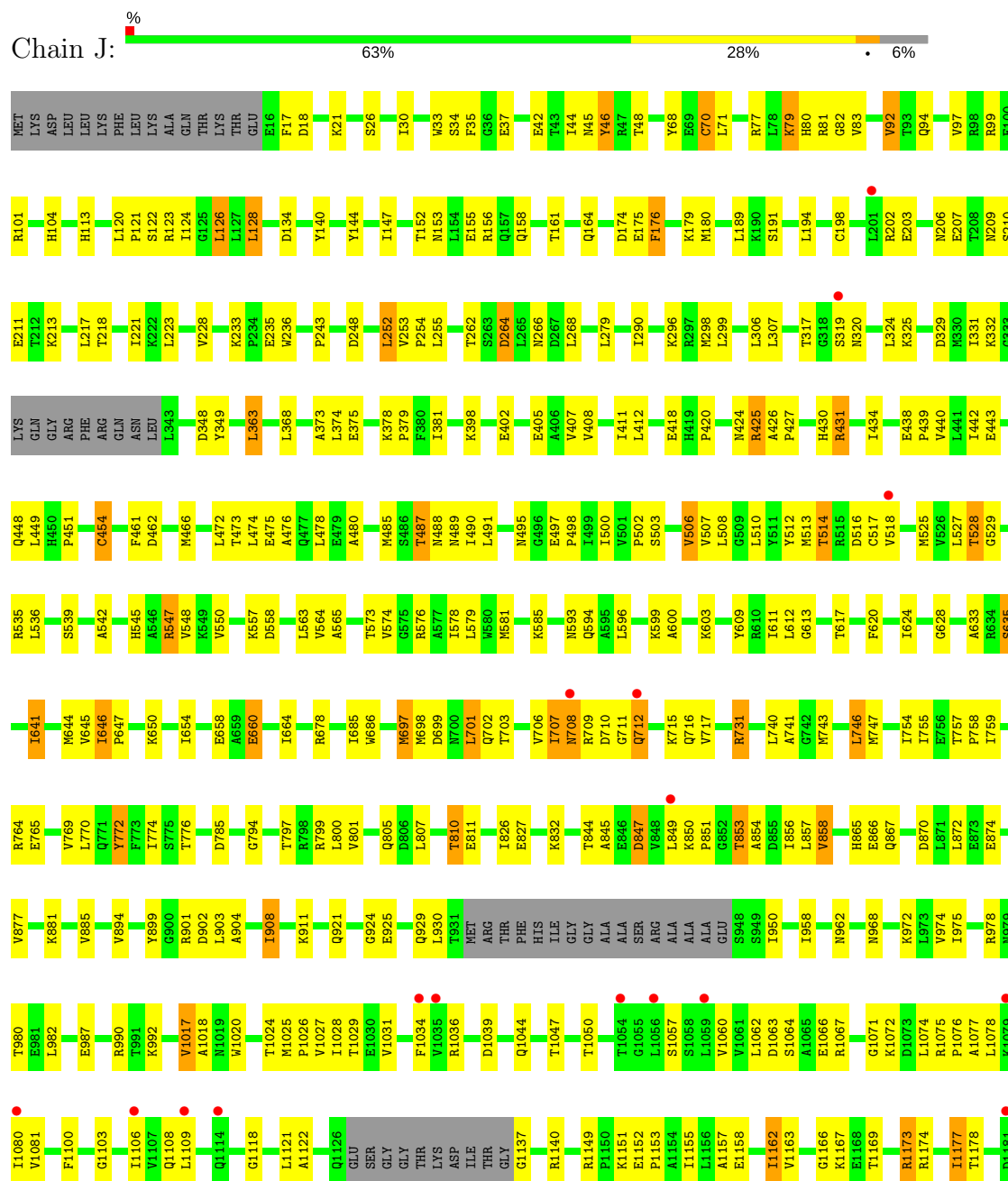


• Molecule 3: DNA-directed RNA polymerase subunit beta'





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





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	186.36Å 206.28Å 308.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.94 – 3.91 39.94 – 3.91	Depositor EDS
% Data completeness (in resolution range)	92.8 (39.94-3.91) 92.9 (39.94-3.91)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.65 (at 3.87Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.219 , 0.260 0.216 , 0.260	Depositor DCC
R_{free} test set	5023 reflections (5.03%)	DCC
Wilson B-factor (Å ²)	120.8	Xtriage
Anisotropy	0.647	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 47.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	58505	wwPDB-VP
Average B, all atoms (Å ²)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.69% 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

5.1 Standard geometry

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	0.25	0/1751	0.51	0/2373
1	B	0.27	0/1707	0.51	0/2314
1	G	0.25	0/1771	0.52	0/2401
1	H	0.27	0/1686	0.51	0/2285
2	C	0.26	0/10739	0.49	2/14489 (0.0%)
2	I	0.26	0/10735	0.49	2/14484 (0.0%)
3	D	0.25	0/10603	0.48	0/14316
3	J	0.25	0/10450	0.47	0/14112
4	E	0.24	0/693	0.47	0/935
4	K	0.23	0/629	0.47	0/847
5	F	0.26	0/3873	0.49	1/5206 (0.0%)
5	L	0.27	0/3872	0.49	1/5205 (0.0%)
6	M	0.25	0/426	0.41	0/583
6	N	0.24	0/426	0.41	0/583
All	All	0.25	0/59361	0.48	6/80133 (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
3	D	0	1
3	J	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	I	237	LEU	CA-CB-CG	6.80	130.94	115.30
2	C	237	LEU	CA-CB-CG	6.68	130.67	115.30
2	I	516	ASP	CB-CG-OD2	5.27	123.04	118.30
2	C	516	ASP	CB-CG-OD2	5.18	122.96	118.30
5	F	149	ASP	CB-CG-OD2	5.17	122.95	118.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	1296	GLY	Peptide
3	J	1296	GLY	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	1730	0	1756	53	0
1	B	1687	0	1700	43	0
1	G	1750	0	1764	57	0
1	H	1667	0	1689	49	0
2	C	10570	0	10582	269	0
2	I	10566	0	10576	248	0
3	D	10447	0	10672	304	0
3	J	10295	0	10511	292	0
4	E	691	0	695	15	0
4	K	627	0	634	12	0
5	F	3822	0	3885	101	0
5	L	3821	0	3884	105	0
6	M	413	0	389	10	0
6	N	413	0	389	13	0
7	D	1	0	0	0	0
7	J	1	0	0	0	0
8	D	2	0	0	0	0
8	J	2	0	0	0	0
All	All	58505	0	59126	1420	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 12.

The worst 5 of 1420 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:206:ASN:ND2	3:J:1183:SER:OG	2.04	0.90
3:D:1183:SER:OG	3:J:206:ASN:ND2	2.08	0.87
2:I:10:ARG:HD3	2:I:1181:PRO:HG2	1.60	0.83
2:C:452:ARG:NH1	2:C:584:TYR:O	2.11	0.83
2:C:10:ARG:HD3	2:C:1181:PRO:HG2	1.60	0.81

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	222/239 (93%)	197 (89%)	21 (10%)	4 (2%)	10	51
1	B	216/239 (90%)	189 (88%)	27 (12%)	0	100	100
1	G	226/239 (95%)	202 (89%)	20 (9%)	4 (2%)	10	51
1	H	213/239 (89%)	187 (88%)	26 (12%)	0	100	100
2	C	1338/1342 (100%)	1213 (91%)	121 (9%)	4 (0%)	44	80
2	I	1338/1342 (100%)	1216 (91%)	117 (9%)	5 (0%)	38	76
3	D	1339/1407 (95%)	1217 (91%)	120 (9%)	2 (0%)	55	88
3	J	1317/1407 (94%)	1201 (91%)	115 (9%)	1 (0%)	55	88
4	E	87/91 (96%)	79 (91%)	8 (9%)	0	100	100
4	K	77/91 (85%)	70 (91%)	7 (9%)	0	100	100
5	F	464/522 (89%)	415 (89%)	48 (10%)	1 (0%)	51	84
5	L	463/522 (89%)	415 (90%)	47 (10%)	1 (0%)	51	84
6	M	49/64 (77%)	42 (86%)	7 (14%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
6	N	49/64 (77%)	42 (86%)	7 (14%)	0	100	100
All	All	7398/7808 (95%)	6685 (90%)	691 (9%)	22 (0%)	44	80

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	170	VAL
2	I	170	VAL
3	D	340	GLN
1	G	167	PRO
2	I	484	LEU

5.3.2 Protein sidechains ⓘ

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	191/206 (93%)	180 (94%)	11 (6%)	23	60
1	B	184/206 (89%)	167 (91%)	17 (9%)	11	43
1	G	191/206 (93%)	180 (94%)	11 (6%)	23	60
1	H	183/206 (89%)	166 (91%)	17 (9%)	10	42
2	C	1155/1157 (100%)	1052 (91%)	103 (9%)	11	44
2	I	1154/1157 (100%)	1056 (92%)	98 (8%)	12	47
3	D	1125/1168 (96%)	1036 (92%)	89 (8%)	14	50
3	J	1110/1168 (95%)	1023 (92%)	87 (8%)	15	50
4	E	72/75 (96%)	69 (96%)	3 (4%)	34	68
4	K	67/75 (89%)	65 (97%)	2 (3%)	46	75
5	F	417/462 (90%)	383 (92%)	34 (8%)	13	48
5	L	418/462 (90%)	386 (92%)	32 (8%)	15	51
6	M	44/56 (79%)	38 (86%)	6 (14%)	4	28
6	N	44/56 (79%)	38 (86%)	6 (14%)	4	28
All	All	6355/6660 (95%)	5839 (92%)	516 (8%)	14	49

5 of 516 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	F	246	GLN
2	I	46	GLN
5	L	130	VAL
5	F	459	THR
1	G	120	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
5	F	446	GLN
5	L	446	GLN
2	I	686	GLN
2	C	1257	GLN
3	J	206	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 ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 6 ligands modelled in this entry, 6 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

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, 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	224/239 (93%)	0.05	5 (2%) 62 53	48, 85, 138, 169	0
1	B	220/239 (92%)	0.37	15 (6%) 18 14	46, 110, 148, 166	0
1	G	228/239 (95%)	0.07	13 (5%) 24 19	48, 94, 141, 167	0
1	H	217/239 (90%)	0.33	14 (6%) 20 14	48, 111, 150, 167	0
2	C	1340/1342 (99%)	-0.08	37 (2%) 53 43	12, 74, 129, 163	0
2	I	1340/1342 (99%)	0.04	58 (4%) 36 28	12, 78, 143, 189	0
3	D	1345/1407 (95%)	-0.08	29 (2%) 62 53	13, 61, 142, 194	0
3	J	1325/1407 (94%)	-0.14	19 (1%) 75 66	10, 63, 139, 195	0
4	E	89/91 (97%)	-0.22	0 100 100	30, 77, 122, 138	0
4	K	79/91 (86%)	-0.29	0 100 100	32, 76, 121, 164	0
5	F	470/522 (90%)	-0.16	8 (1%) 70 61	25, 98, 147, 179	0
5	L	469/522 (89%)	-0.16	10 (2%) 64 54	28, 96, 150, 181	0
6	M	51/64 (79%)	0.53	1 (1%) 65 56	66, 115, 146, 158	0
6	N	51/64 (79%)	0.31	3 (5%) 23 17	64, 112, 143, 149	0
All	All	7448/7808 (95%)	-0.04	212 (2%) 53 43	10, 79, 142, 195	0

The worst 5 of 212 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	I	982	GLY	9.4
2	I	981	ALA	7.9
3	J	1059	LEU	6.5
1	B	146	VAL	5.6
2	I	995	ASP	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 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, 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	LLDF	B-factors(Å ²)	Q<0.9
8	ZN	D	1503	1/1	0.99	0.21	-0.40	36,36,36,36	0
8	ZN	J	1503	1/1	1.00	0.15	-0.40	23,23,23,23	0
8	ZN	D	1502	1/1	0.90	0.18	-0.82	193,193,193,193	0
8	ZN	J	1502	1/1	0.98	0.06	-1.51	71,71,71,71	0
7	MG	J	1501	1/1	0.81	0.36	-	40,40,40,40	0
7	MG	D	1501	1/1	0.70	0.44	-	41,41,41,41	0

6.5 Other polymers [i](#)

There are no such residues in this entry.