wwPDB X-ray Structure Validation Summary Report (i)

Jun 19, 2024 – 03:26 AM EDT

PDB ID	:	3WOD
Title	:	RNA polymerase-gp39 complex
Authors	:	Tagami, S.; Sekine, S.; Minakhin, L.; Esyunina, D.; Akasaka, R.; Shirouzu, M.; Kulbachinskiy, A.; Severinov, K.; Yokoyama, S.
Deposited on	:	2013-12-26
Resolution	:	3.60 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 (i) 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 (1)) were used in the production of this report:

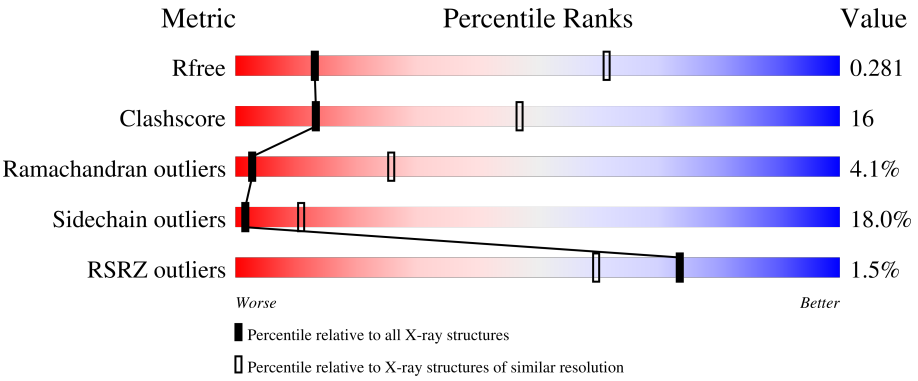
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.60 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.



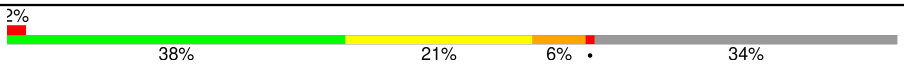
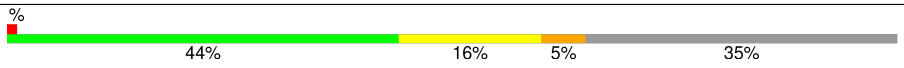
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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 <=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	315	40% 26% 5% 29%
1	B	315	39% 29% 5% 26%
2	C	1119	52% 39% 9%
3	D	1524	2% 49% 39% 10%
4	E	99	3% 56% 31% 7%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	423	
6	G	141	
6	H	141	

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 28954 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	225	Total	C	N	O	S	0	0	0
			1772	1132	308	330	2			
1	B	232	Total	C	N	O	S	0	0	0
			1814	1158	316	338	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1118	Total	C	N	O	S	0	0	0
			8817	5575	1573	1645	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1488	Total	C	N	O	S	0	0	0
			11752	7450	2069	2197	36			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	95	Total	C	N	O	S	0	0	0
			769	488	133	144	4			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	61	GLU	VAL	SEE REMARK 999	UNP Q8RQE7
E	92	ILE	LEU	SEE REMARK 999	UNP Q8RQE7
E	95	GLY	VAL	SEE REMARK 999	UNP Q8RQE7

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	278	Total	C	N	O	S	0	0	0
			2242	1421	399	420	2			

- Molecule 6 is a protein called Putative uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	127	Total	C	N	O	S	0	0	0
			1035	673	175	184	3			
6	H	91	Total	C	N	O	S	0	0	0
			752	495	125	131	1			

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

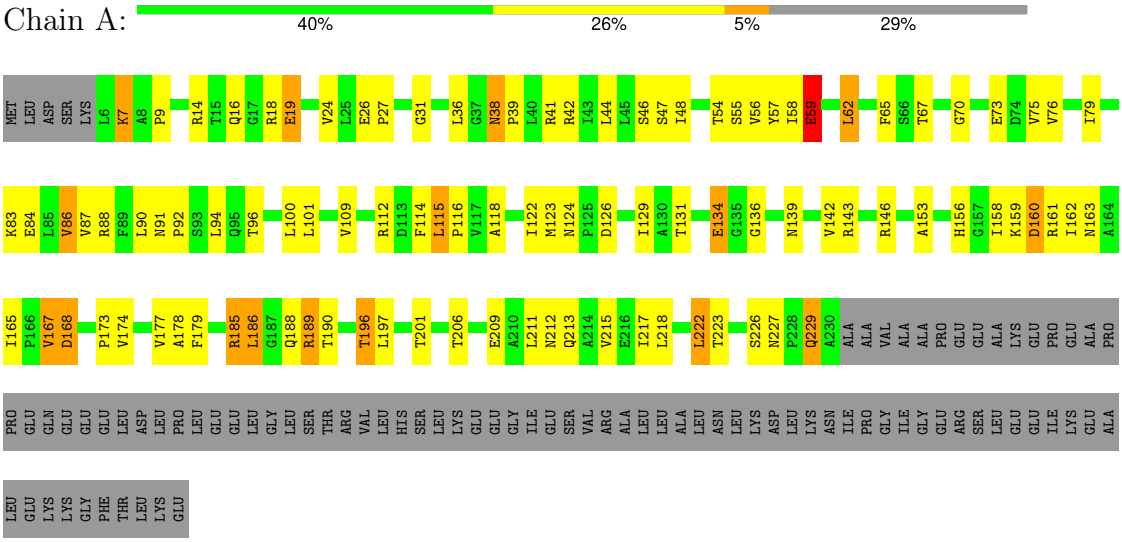
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	D	1	Total	Zn	0	0
			1	1		

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.

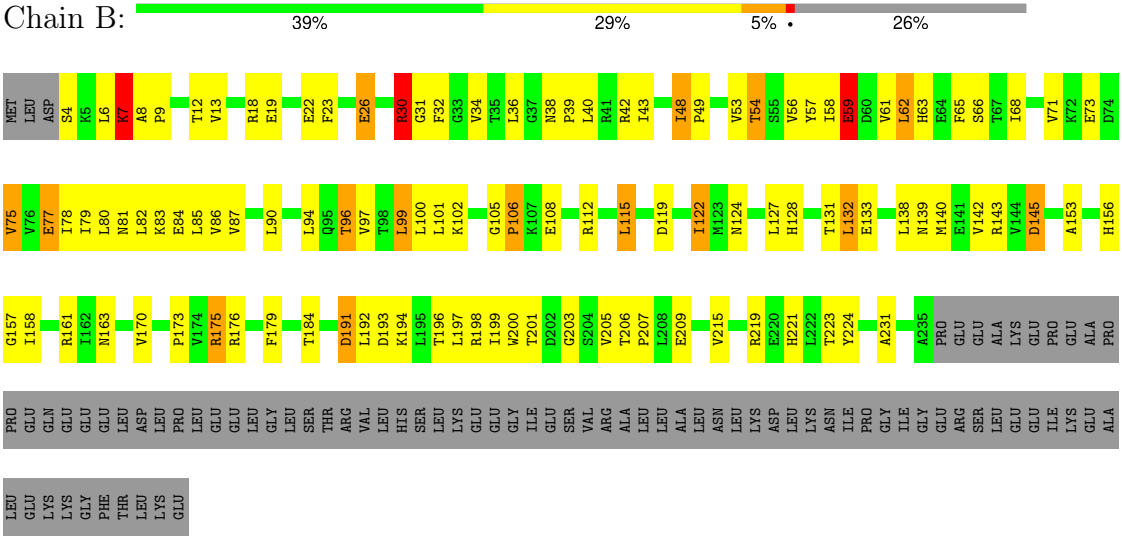
• Molecule 1: DNA-directed RNA polymerase subunit alpha

Chain A:



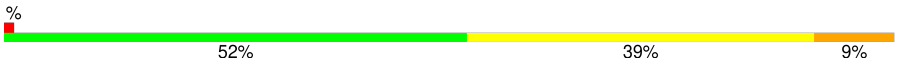
• Molecule 1: DNA-directed RNA polymerase subunit alpha

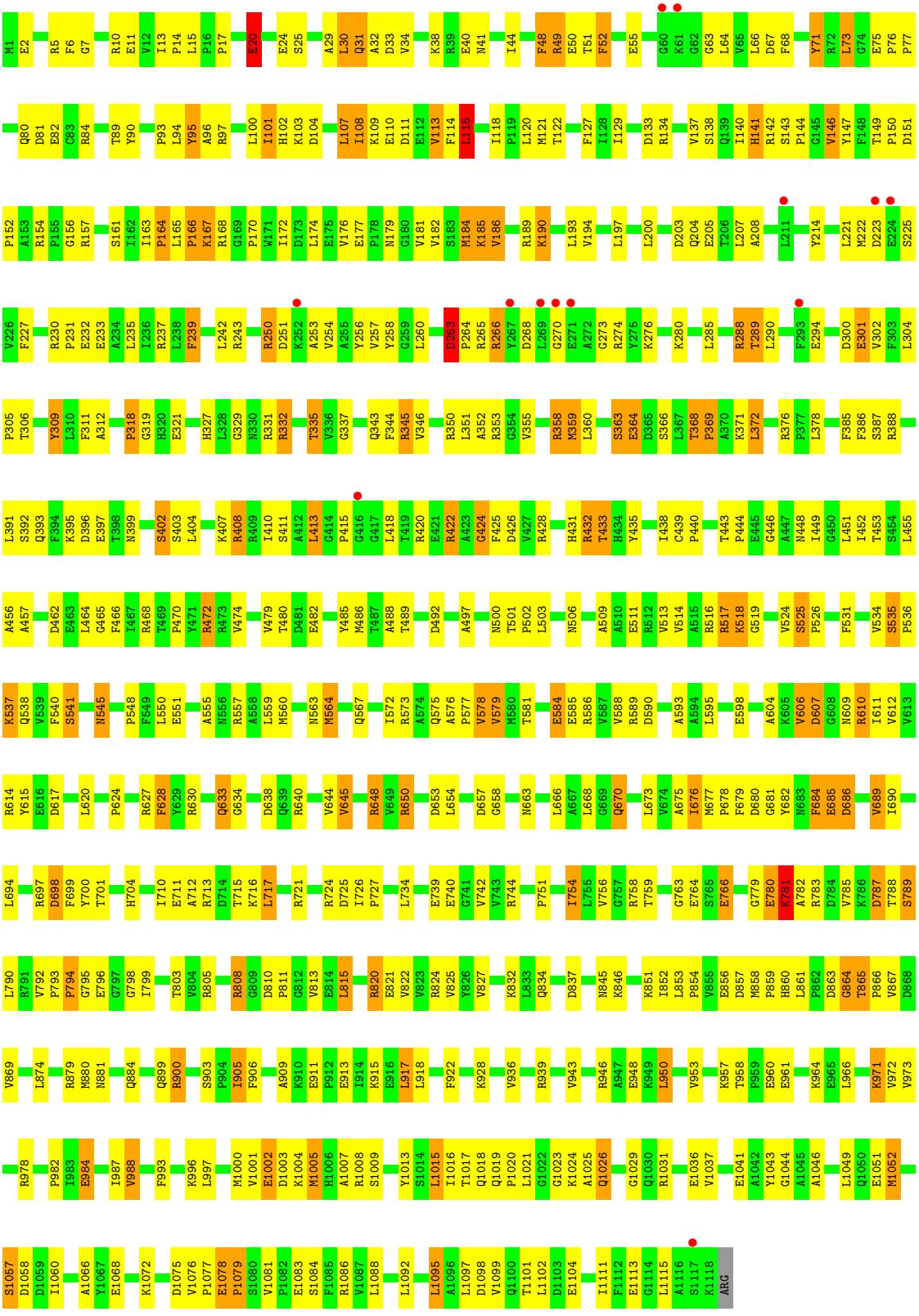
Chain B:



• Molecule 2: DNA-directed RNA polymerase subunit beta

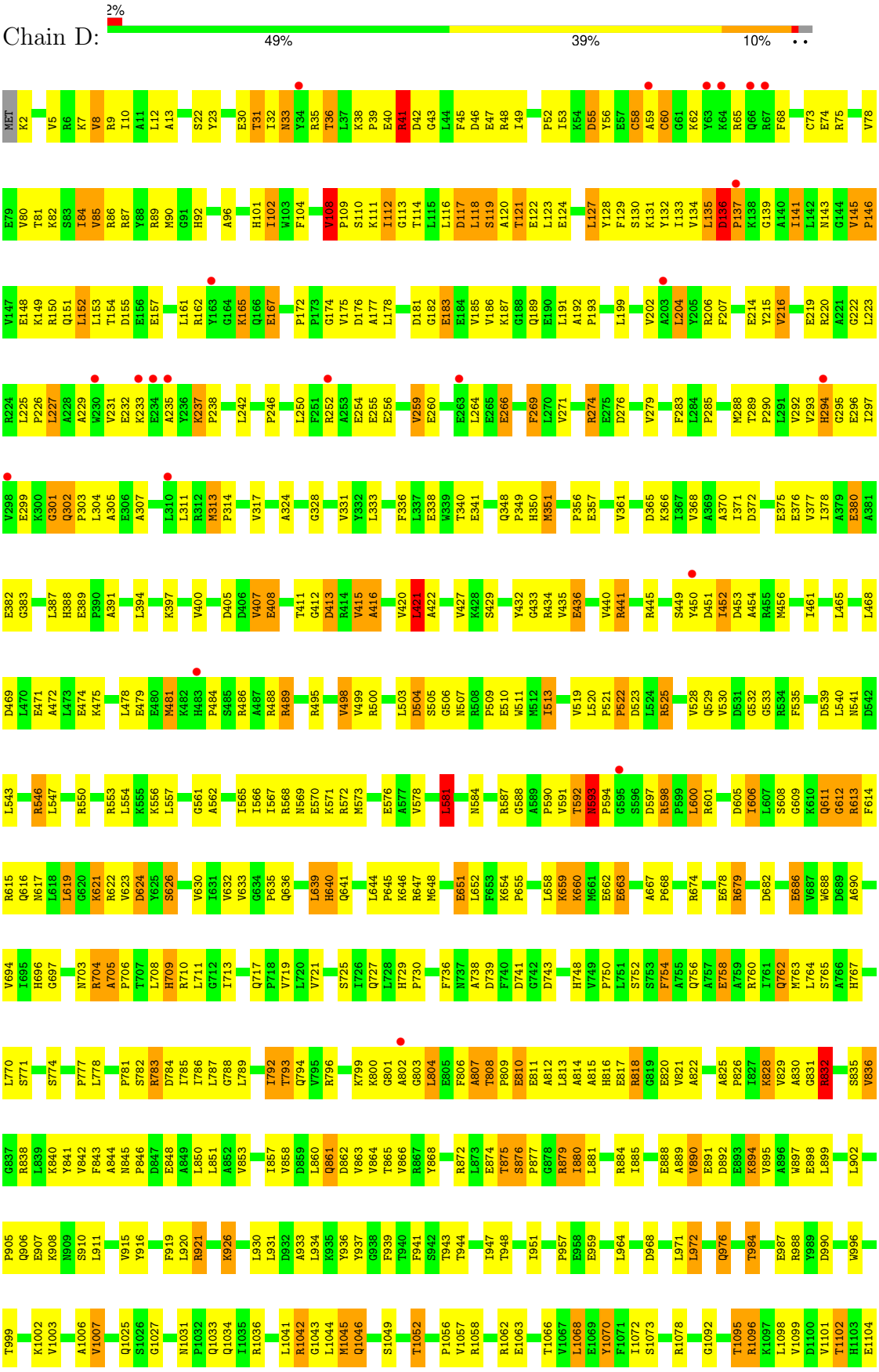
Chain C:

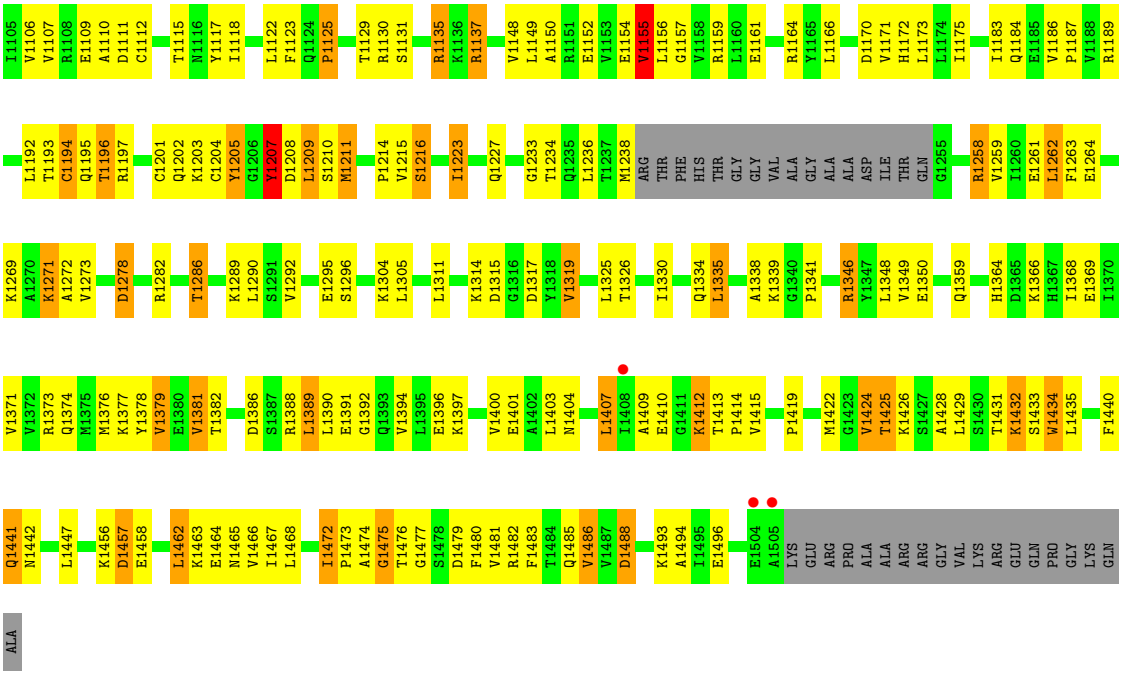




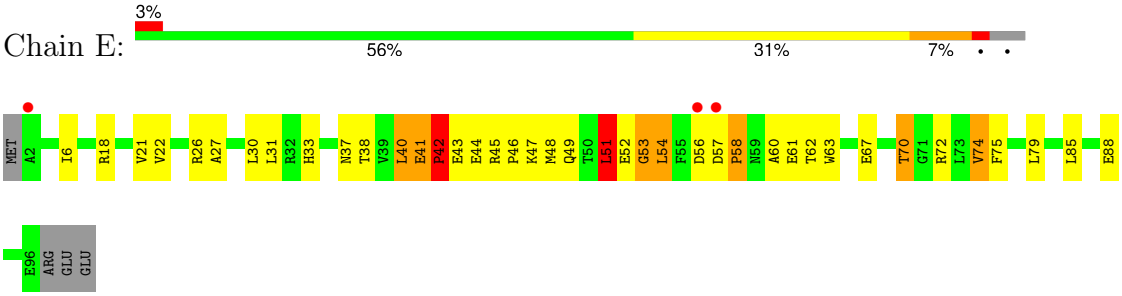
• Molecule 3: DNA-directed RNA polymerase subunit beta'

Chain D:

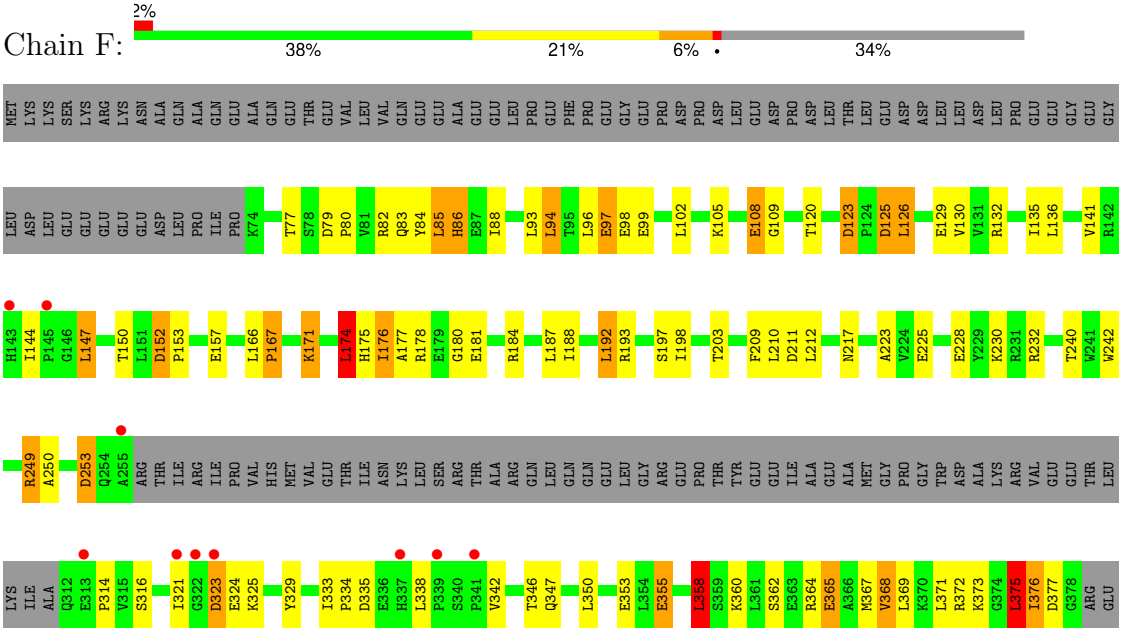


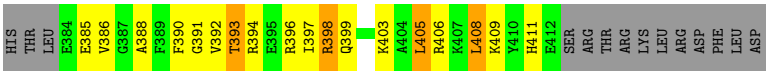


• Molecule 4: DNA-directed RNA polymerase subunit omega



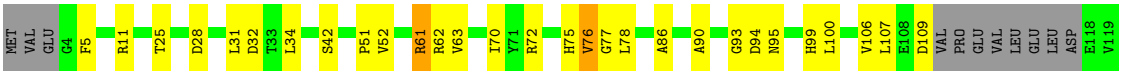
• Molecule 5: RNA polymerase sigma factor

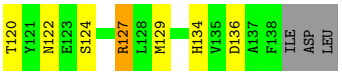




• Molecule 6: Putative uncharacterized protein

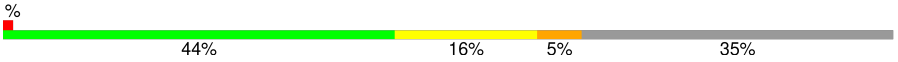
Chain G:

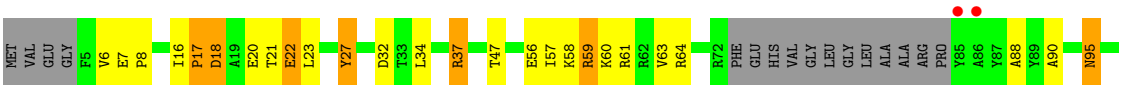




• Molecule 6: Putative uncharacterized protein

Chain H:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	294.44Å 294.44Å 223.25Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.88 – 3.60 19.88 – 3.58	Depositor EDS
% Data completeness (in resolution range)	99.5 (19.88-3.60) 99.3 (19.88-3.58)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.22	Depositor
$< I/\sigma(I) >$ [1]	2.65 (at 3.61Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, R_{free}	0.250 , 0.279 0.256 , 0.281	Depositor DCC
R_{free} test set	6487 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	64.5	Xtriage
Anisotropy	0.443	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.22 , -23.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.33$, $< L^2 > = 0.16$	Xtriage
Estimated twinning fraction	0.166 for -h,-k,l	Xtriage
F_o,F_c correlation	0.84	EDS
Total number of atoms	28954	wwPDB-VP
Average B, all atoms (Å^2)	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.84% 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: 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	0.35	0/1804	0.63	1/2454 (0.0%)
1	B	0.32	0/1846	0.59	0/2511
2	C	0.33	0/8985	0.61	2/12150 (0.0%)
3	D	0.32	0/11958	0.60	4/16166 (0.0%)
4	E	0.31	0/783	0.66	0/1054
5	F	0.33	0/2276	0.59	3/3058 (0.1%)
6	G	0.27	0/1065	0.53	0/1449
6	H	0.27	0/775	0.48	0/1057
All	All	0.32	0/29492	0.60	10/39899 (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	3
3	D	0	3
4	E	0	3
5	F	0	1
All	All	0	10

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	795	GLY	N-CA-C	-6.48	96.90	113.10
1	A	115	LEU	CA-CB-CG	6.18	129.51	115.30
3	D	581	LEU	CA-CB-CG	5.89	128.84	115.30
5	F	174	LEU	CA-CB-CG	5.88	128.81	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	804	LEU	CA-CB-CG	5.79	128.61	115.30

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	260	LEU	Peptide
2	C	685	GLU	Peptide
2	C	794	PRO	Peptide
3	D	1196	THR	Peptide
3	D	561	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	1772	0	1821	63	0
1	B	1814	0	1868	73	0
2	C	8817	0	8920	305	0
3	D	11752	0	11990	447	0
4	E	769	0	775	29	0
5	F	2242	0	2305	62	0
6	G	1035	0	1013	19	0
6	H	752	0	748	20	0
7	D	1	0	0	0	0
All	All	28954	0	29440	943	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 16.

The worst 5 of 943 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:820:GLU:HG3	3:D:836:VAL:HG21	1.51	0.92
2:C:164:PRO:HA	2:C:266:ARG:HH22	1.36	0.90

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:E:54:LEU:HA	4:E:58:PRO:HD2	1.57	0.85
2:C:63:GLY:HA3	2:C:103:LYS:HG2	1.58	0.83
3:D:133:ILE:HG21	3:D:454:ALA:HB1	1.60	0.83

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	223/315 (71%)	198 (89%)	22 (10%)	3 (1%)	12 50
1	B	230/315 (73%)	207 (90%)	14 (6%)	9 (4%)	3 27
2	C	1116/1119 (100%)	947 (85%)	128 (12%)	41 (4%)	3 28
3	D	1484/1524 (97%)	1209 (82%)	200 (14%)	75 (5%)	2 20
4	E	93/99 (94%)	75 (81%)	12 (13%)	6 (6%)	1 16
5	F	272/423 (64%)	225 (83%)	36 (13%)	11 (4%)	3 26
6	G	123/141 (87%)	116 (94%)	5 (4%)	2 (2%)	9 46
6	H	87/141 (62%)	81 (93%)	4 (5%)	2 (2%)	6 38
All	All	3628/4077 (89%)	3058 (84%)	421 (12%)	149 (4%)	3 26

5 of 149 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	59	GLU
1	B	7	LYS
1	B	59	GLU
1	B	75	VAL
2	C	31	GLN

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	197/273 (72%)	164 (83%)	33 (17%)	2 14
1	B	200/273 (73%)	172 (86%)	28 (14%)	3 21
2	C	940/941 (100%)	756 (80%)	184 (20%)	1 8
3	D	1254/1279 (98%)	1018 (81%)	236 (19%)	1 9
4	E	83/87 (95%)	70 (84%)	13 (16%)	2 17
5	F	241/371 (65%)	200 (83%)	41 (17%)	2 14
6	G	107/121 (88%)	92 (86%)	15 (14%)	3 21
6	H	79/121 (65%)	71 (90%)	8 (10%)	7 34
All	All	3101/3466 (90%)	2543 (82%)	558 (18%)	1 11

5 of 558 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	1415	VAL
3	D	1465	ASN
3	D	1412	LYS
5	F	347	GLN
2	C	686	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	1323	GLN
5	F	218	GLN
2	C	45	GLN
2	C	327	HIS
3	D	569	ASN

5.3.3 RNA (i)

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 monosaccharides in this entry.

5.6 Ligand geometry

Of 1 ligands modelled in this entry, 1 is 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

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	225/315 (71%)	-0.18	0 100 100	11, 37, 83, 144	0
1	B	232/315 (73%)	-0.22	0 100 100	14, 47, 104, 158	0
2	C	1118/1119 (99%)	-0.12	13 (1%) 79 66	8, 59, 151, 222	0
3	D	1488/1524 (97%)	-0.12	25 (1%) 70 55	8, 85, 179, 244	0
4	E	95/99 (95%)	-0.00	3 (3%) 47 32	40, 74, 183, 242	0
5	F	278/423 (65%)	-0.10	10 (3%) 42 28	44, 119, 192, 255	0
6	G	127/141 (90%)	-0.20	0 100 100	27, 73, 137, 179	0
6	H	91/141 (64%)	-0.03	2 (2%) 62 45	73, 113, 161, 175	0
All	All	3654/4077 (89%)	-0.13	53 (1%) 73 60	8, 74, 170, 255	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	323	ASP	7.4
4	E	57	ASP	7.4
5	F	322	GLY	7.1
3	D	294	HIS	6.6
2	C	60	GLY	6.2

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

There are no monosaccharides in this entry.

6.4 Ligands

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, 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	B-factors(Å^2)	Q<0.9
7	ZN	D	1601	1/1	0.98	0.20	110,110,110,110	0

6.5 Other polymers

There are no such residues in this entry.