wwPDB X-ray Structure Validation Summary Report (i)

May 28, 2020 – 09:25 pm BST

PDB ID	:	1Y1W
Title	:	Complete RNA Polymerase II elongation complex
Authors	:	Cramer, P.; Kettenberger, H.; Armache, K.-J.
Deposited on	:	2004-11-19
Resolution	:	4.00 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 following versions of software and data (see references (1)) were used in the production of this report:

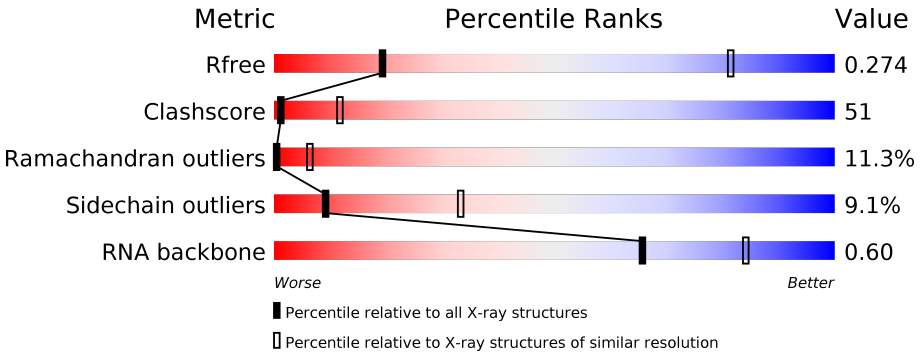
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 4.00 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.



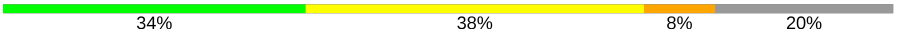
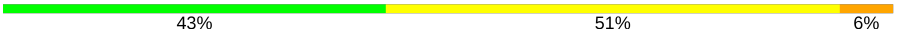
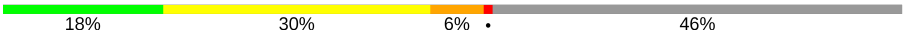
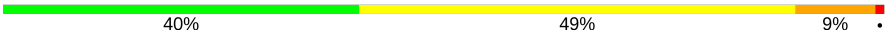
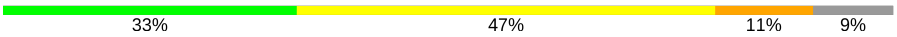
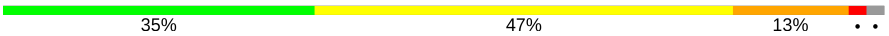
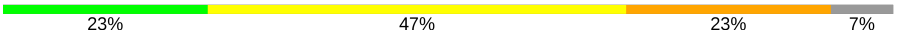
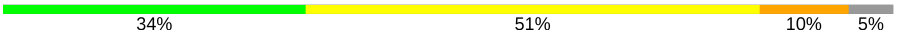
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1087 (4.30-3.70)
Clashscore	141614	1148 (4.30-3.70)
Ramachandran outliers	138981	1108 (4.30-3.70)
Sidechain outliers	138945	1099 (4.30-3.70)
RNA backbone	3102	1048 (5.00-3.00)

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 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%

Mol	Chain	Length	Quality of chain
1	T	19	32% 68%
2	N	7	14% 71% 14%
3	P	10	80% 20%
4	A	1733	25% 45% 10% • 18%
5	B	1224	27% 50% 13% • 9%
6	C	318	27% 45% 10% • 16%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	D	221	
8	E	215	
9	F	155	
10	G	171	
11	H	146	
12	I	122	
13	J	70	
14	K	120	
15	L	70	

2 Entry composition

There are 17 unique types of molecules in this entry. The entry contains 31802 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 DNA chain called 5'-D(P*AP*GP*TP*AP*CP*TP*TP*AP*CP*GP*CP*CP*TP*GP*GP*TP*CP*AP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	T	19	387	185	67	116	19	21	0	0

- Molecule 2 is a DNA chain called 5'-D(*AP*AP*GP*TP*AP*CP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	N	7	141	69	27	39	6	20	0	0

- Molecule 3 is a RNA chain called 5'-R(*AP*AP*GP*AP*CP*CP*AP*GP*GP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	P	10	214	97	44	64	9	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	A	1416	11140	7021	1946	2111	62	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerase II 140 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	B	1112	8836	5594	1548	1639	55	0	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerase II 45 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	C	266	Total	C	N	O	S	0	0	0
			2095	1317	348	417	13			

- Molecule 7 is a protein called DNA-directed RNA polymerase II 32 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	D	177	Total	C	N	O	S	0	0	0
			1356	840	241	273	2			

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	E	214	Total	C	N	O	S	0	0	0
			1752	1111	309	321	11			

- Molecule 9 is a protein called DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	F	84	Total	C	N	O	S	0	0	0
			679	434	115	127	3			

- Molecule 10 is a protein called DNA-directed RNA polymerase II 19 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	G	171	Total	C	N	O	S	0	0	0
			1340	861	222	249	8			

- Molecule 11 is a protein called DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	H	133	Total	C	N	O	S	0	0	0
			1068	673	180	211	4			

- Molecule 12 is a protein called DNA-directed RNA polymerase II subunit 9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	I	119	Total	C	N	O	S	0	0	0
			971	596	179	186	10			

- Molecule 13 is a protein called DNA-directed RNA polymerases I/II/III subunit 10.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	J	65	Total	C	N	O	S	0	0	0
			532	339	93	94	6			

- Molecule 14 is a protein called DNA-directed RNA polymerase II 13.6 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
14	K	114	Total	C	N	O	S	0	0	0
			919	590	156	171	2			

- Molecule 15 is a protein called DNA-directed RNA polymerases I, II, and III 7.7 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
15	L	46	Total	C	N	O	S	0	0	0
			363	224	72	63	4			

- Molecule 16 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
16	J	1	Total	Zn	0	0
			1	1		
16	B	1	Total	Zn	0	0
			1	1		
16	I	2	Total	Zn	0	0
			2	2		
16	C	1	Total	Zn	0	0
			1	1		
16	A	2	Total	Zn	0	0
			2	2		
16	L	1	Total	Zn	0	0
			1	1		

- Molecule 17 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
17	A	1	Total	Mg	0	0
			1	1		

3 Residue-property plots (i)

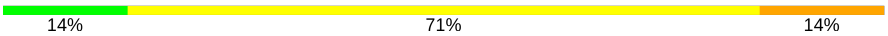
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according 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. 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: 5'-D(P*AP*GP*TP*AP*CP*TP*TP*AP*CP*GP*CP*CP*TP*GP*GP*TP*CP*AP*T)-3'

Chain T:

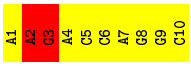
• Molecule 2: 5'-D(*AP*AP*GP*TP*AP*CP*T)-3'

Chain N:



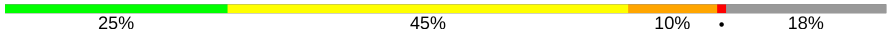
• Molecule 3: 5'-R(*AP*AP*GP*AP*CP*CP*AP*GP*GP*C)-3'

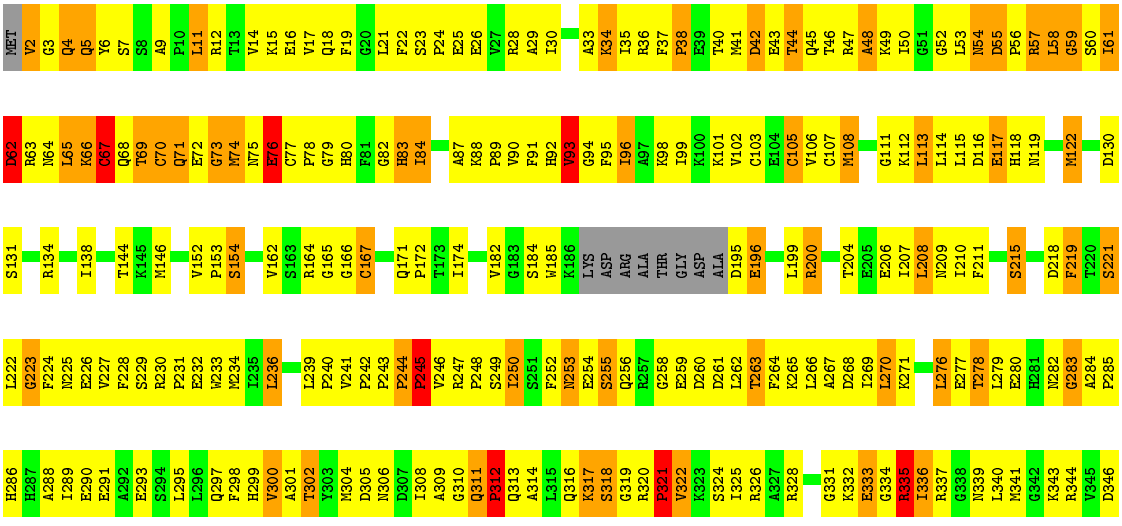
Chain P:

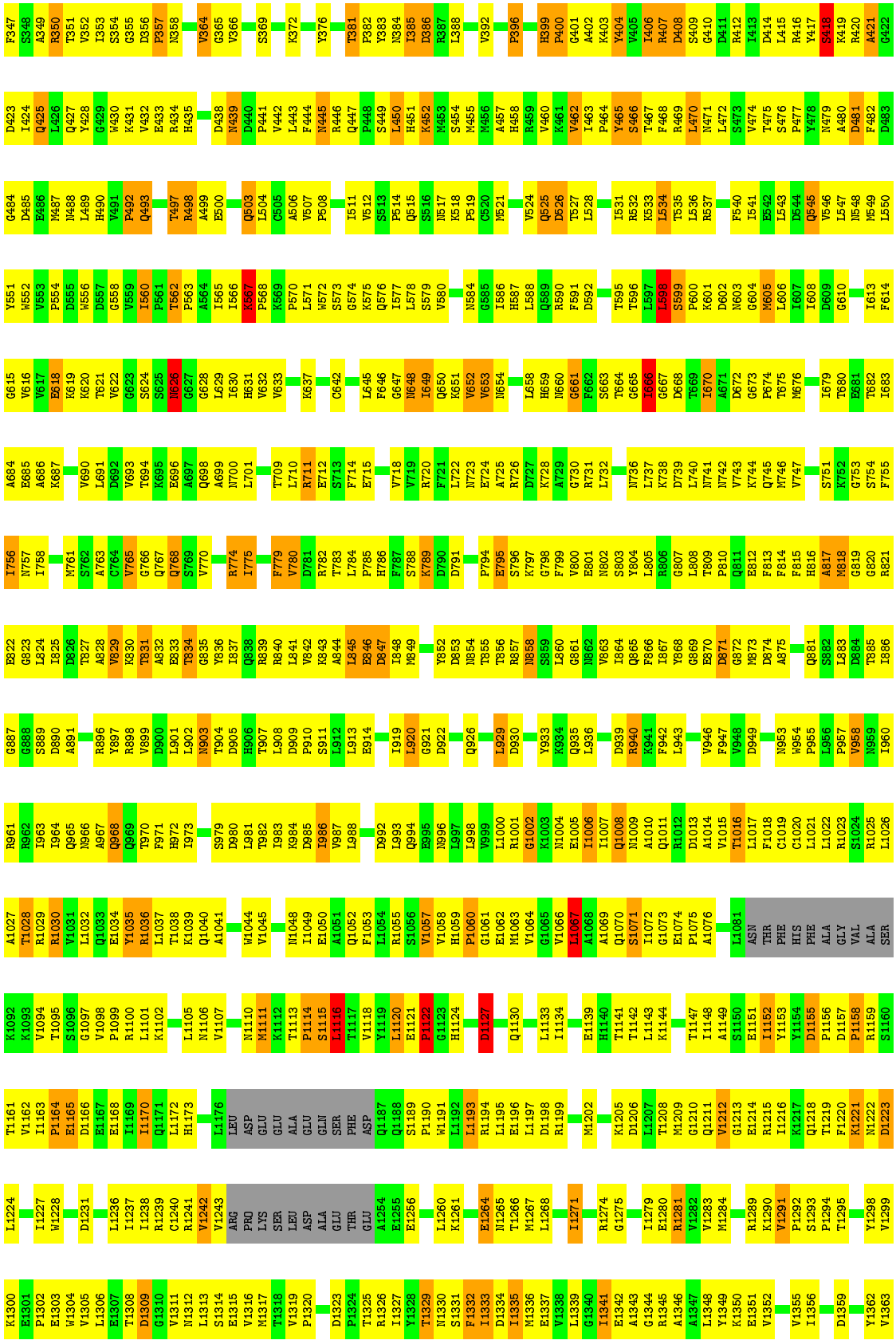


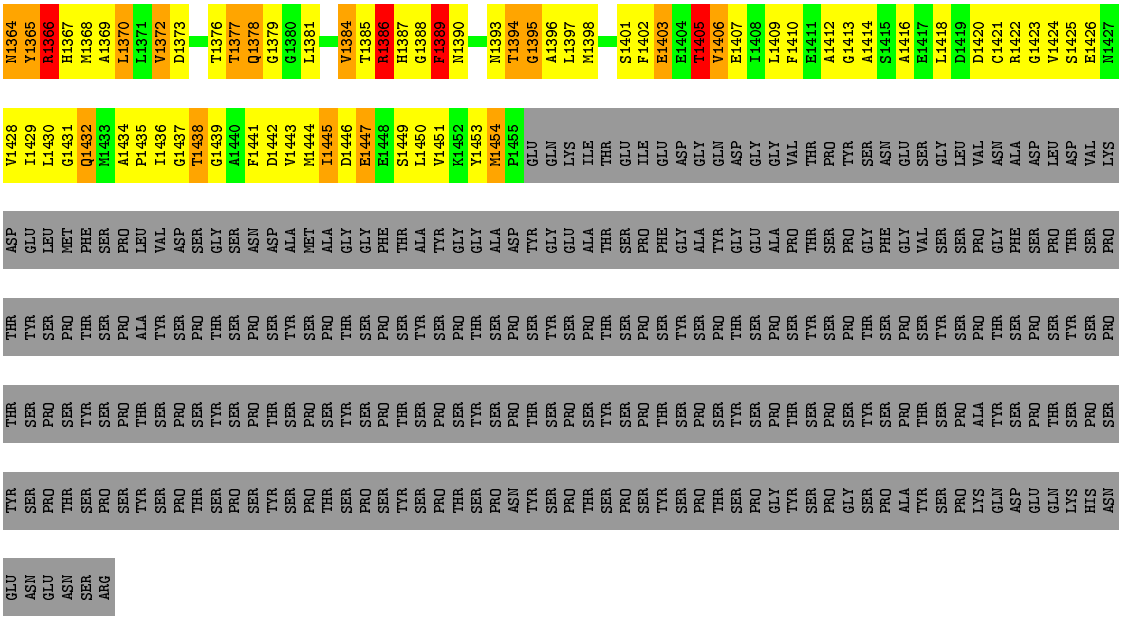
• Molecule 4: DNA-directed RNA polymerase II largest subunit

Chain A:



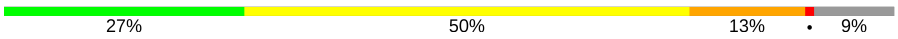


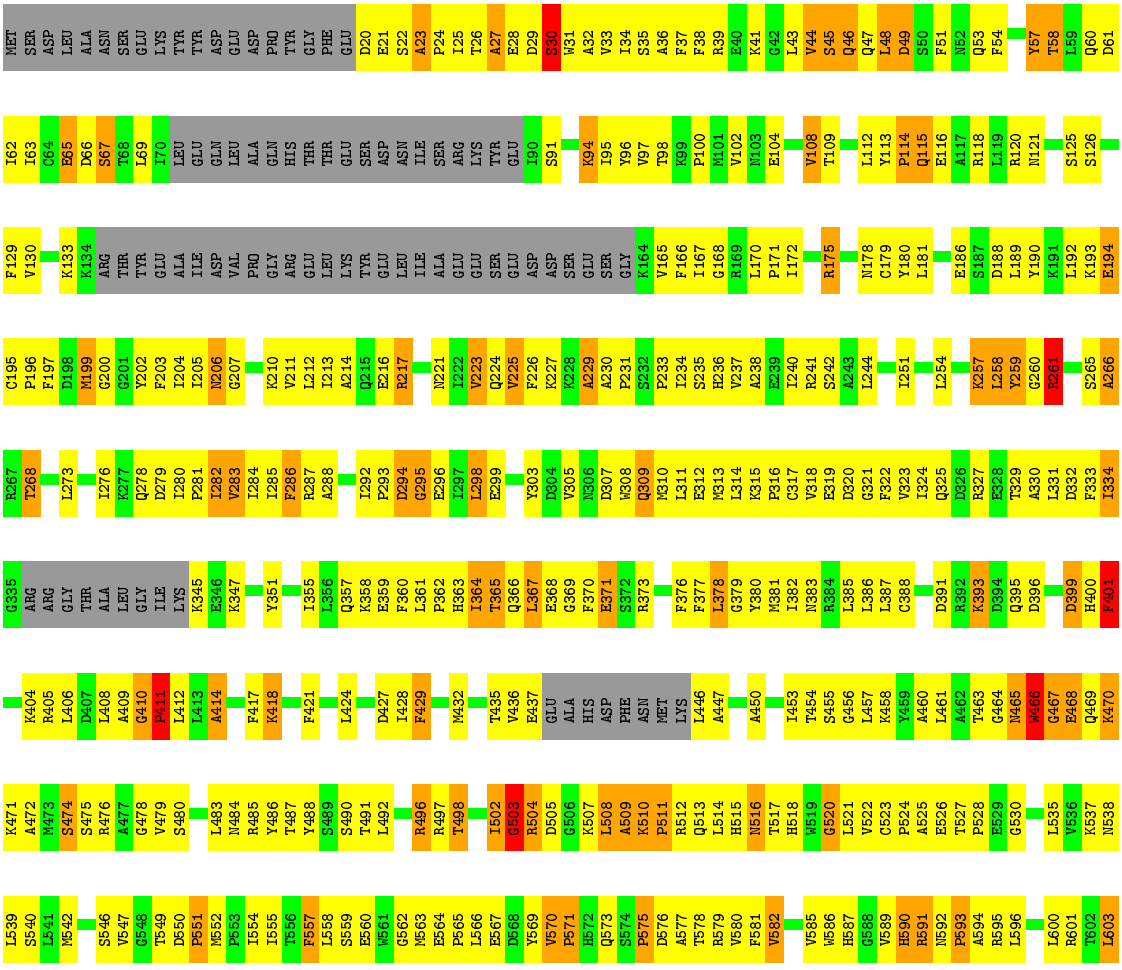


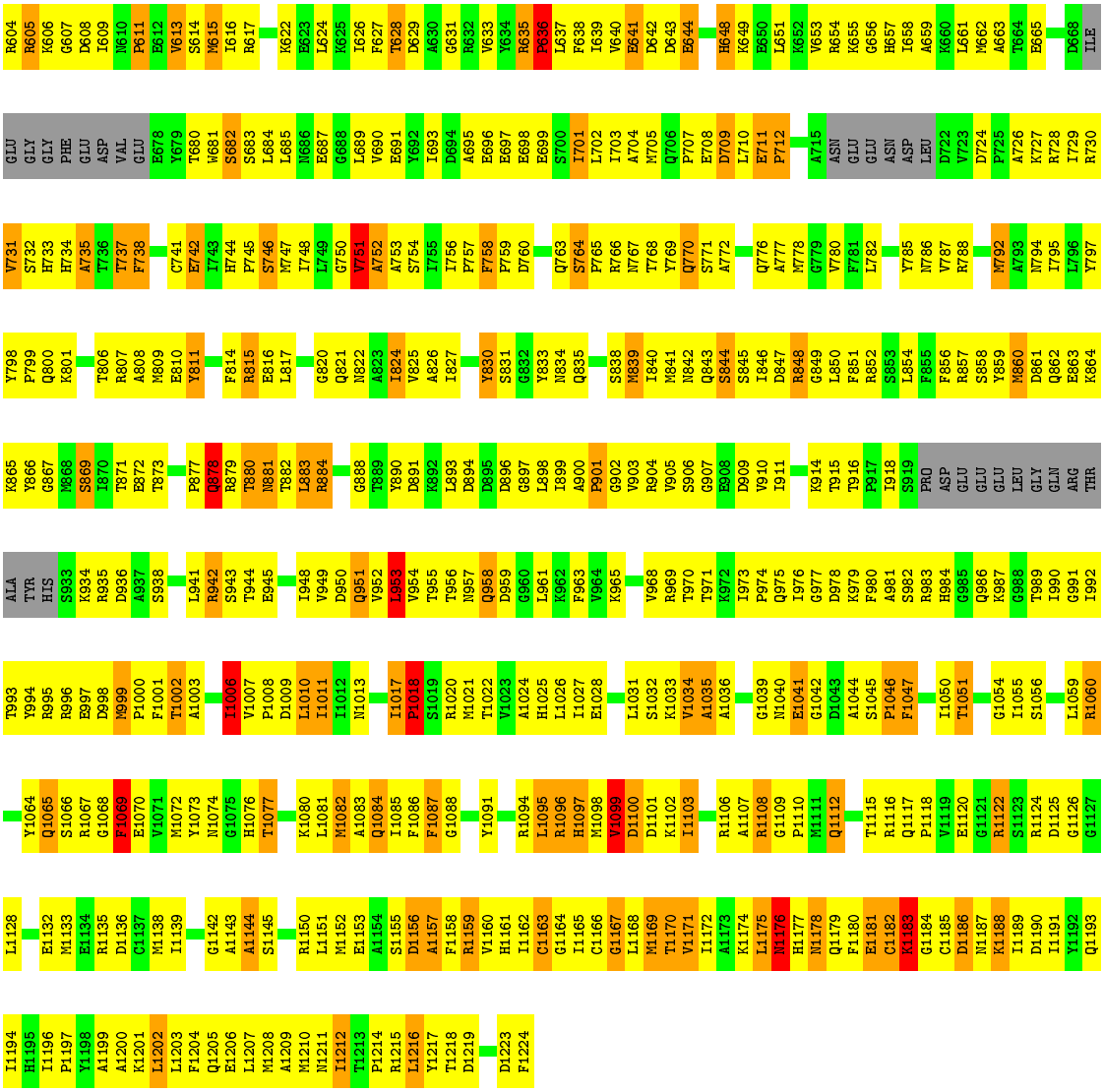


• Molecule 5: DNA-directed RNA polymerase II 140 kDa polypeptide

Chain B:

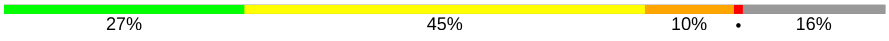


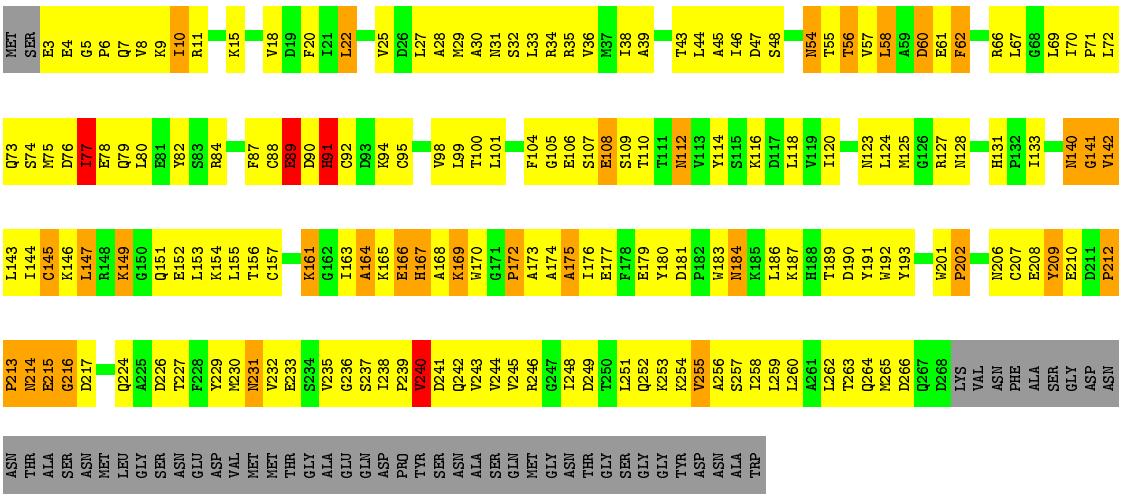




• Molecule 6: DNA-directed RNA polymerase II 45 kDa polypeptide

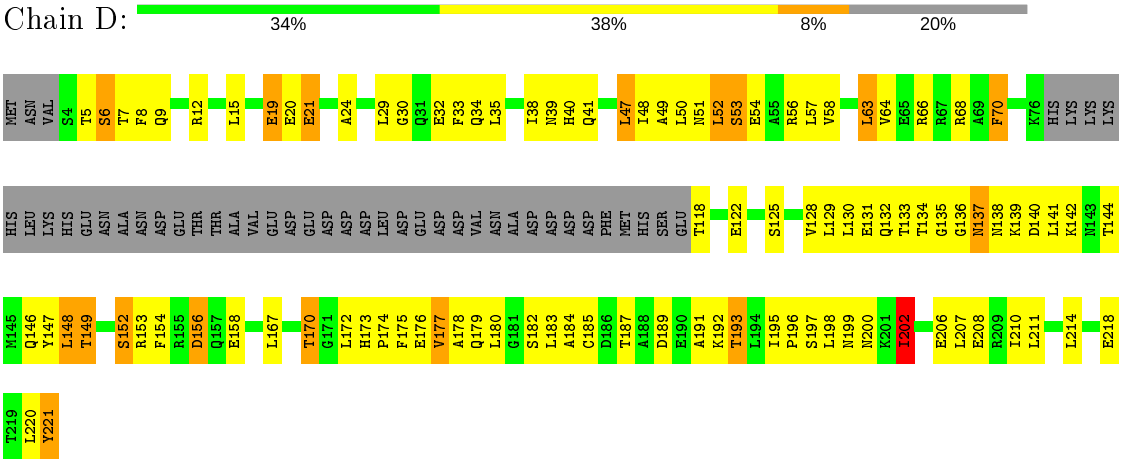
Chain C:





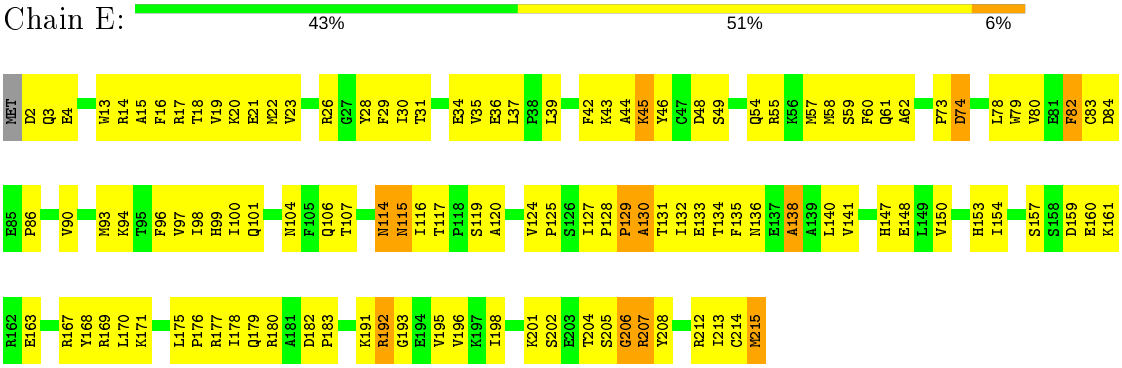
• Molecule 7: DNA-directed RNA polymerase II 32 kDa polypeptide

Chain D:



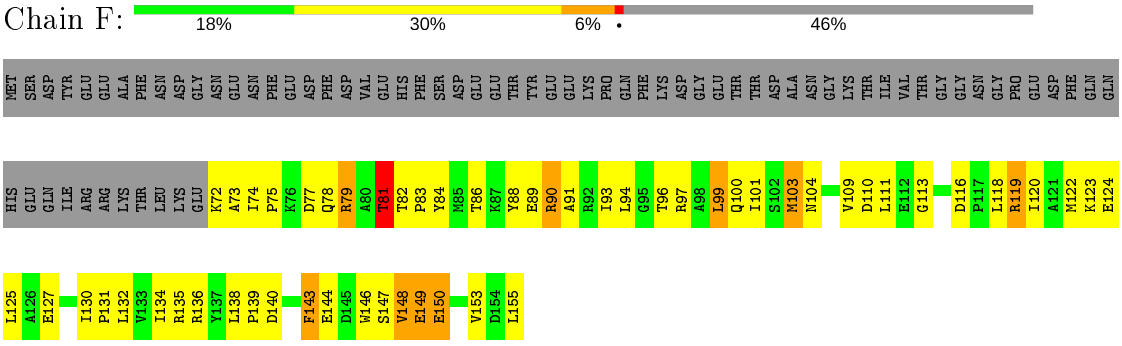
• Molecule 8: DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide

Chain E:



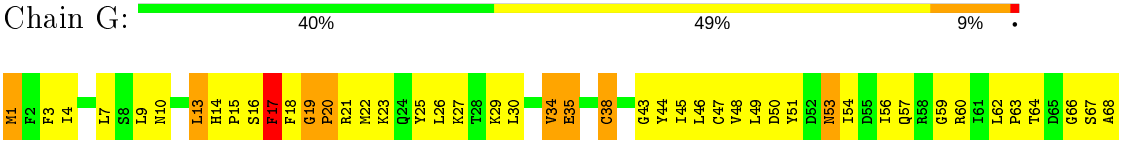
• Molecule 9: DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide

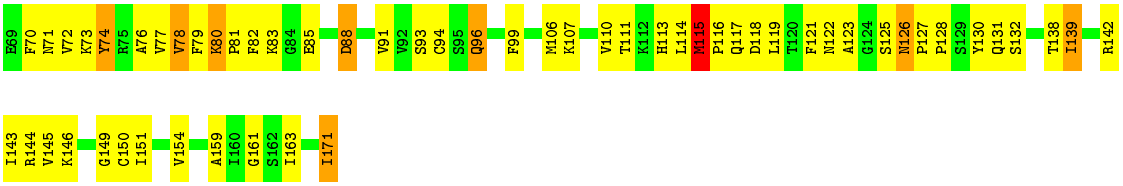
Chain F:



• Molecule 10: DNA-directed RNA polymerase II 19 kDa polypeptide

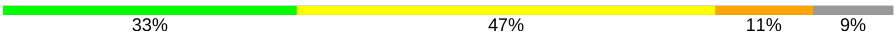
Chain G:

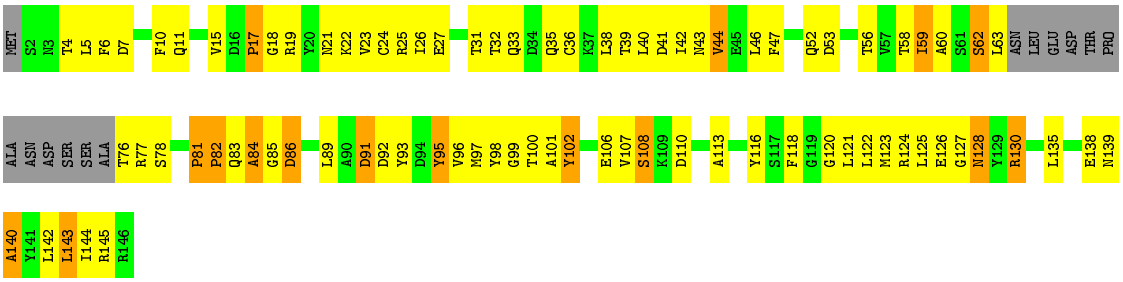




• Molecule 11: DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide

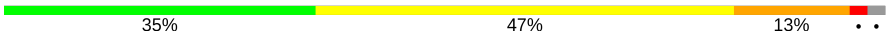
Chain H:

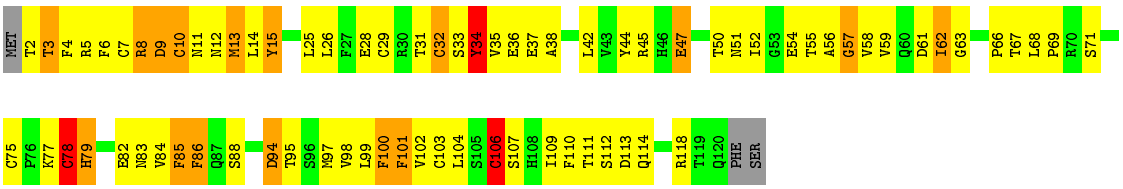




• Molecule 12: DNA-directed RNA polymerase II subunit 9

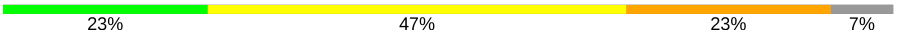
Chain I:

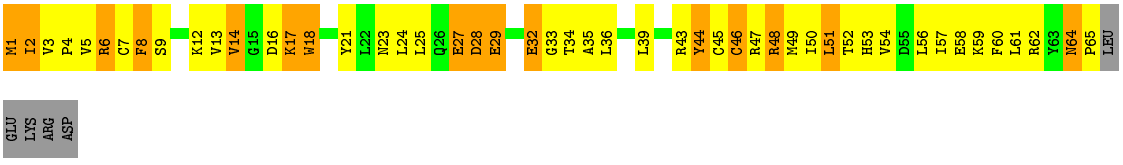




• Molecule 13: DNA-directed RNA polymerases I/II/III subunit 10

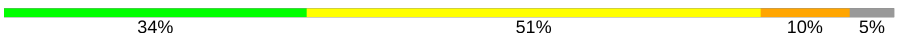
Chain J:

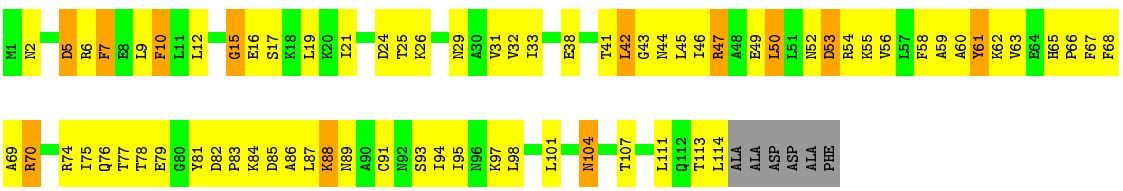




• Molecule 14: DNA-directed RNA polymerase II 13.6 kDa polypeptide

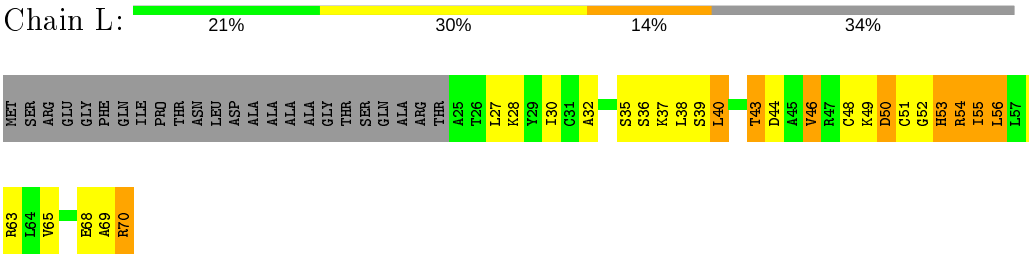
Chain K:





• Molecule 15: DNA-directed RNA polymerases I, II, and III 7.7 kDa polypeptide

Chain L:


21%
30%
14%
34%
MET SER ARG GLU GLY PHE GLN ILE PRO THR ASN LEU ASP ALA ALA ALA ALA GLY THR SER GLN ALA ARG THR A25 T26 L27 K28 Y29 I30 C31 A32 S35 S36 K37 L38 S39 L40 T43 D44 A45 V46 R47 C48 K49 D50 C51 G52 H53 R54 I55 L56 L57 K58 A59 R60 T61 K62
R63 L64 V65 E68 A69 R70

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	221.37Å 392.50Å 283.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 4.00 48.83 – 3.78	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-4.00) 94.4 (48.83-3.78)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.00 (at 3.77Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.253 , 0.276 0.260 , 0.274	Depositor DCC
R_{free} test set	2439 reflections (2.00%)	wwPDB-VP
Wilson B-factor (Å^2)	77.4	Xtriage
Anisotropy	0.284	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.18 , -11.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.28$, $< L^2 > = 0.12$	Xtriage
Estimated twinning fraction	0.207 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.206 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
F_o,F_c correlation	0.83	EDS
Total number of atoms	31802	wwPDB-VP
Average B, all atoms (Å^2)	89.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.51% 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, 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	T	1.15	1/432 (0.2%)	1.03	1/664 (0.2%)
2	N	1.74	1/158 (0.6%)	0.91	1/242 (0.4%)
3	P	1.17	2/240 (0.8%)	1.06	3/373 (0.8%)
4	A	0.50	0/11339	0.75	5/15334 (0.0%)
5	B	0.51	1/9008 (0.0%)	0.74	5/12146 (0.0%)
6	C	0.56	0/2133	0.76	0/2891
7	D	0.46	0/1365	0.71	0/1837
8	E	0.45	0/1788	0.64	0/2406
9	F	0.56	0/691	0.80	0/933
10	G	0.55	0/1368	0.76	0/1844
11	H	0.40	0/1086	0.65	0/1470
12	I	0.49	1/989 (0.1%)	0.72	0/1331
13	J	0.52	0/541	0.80	0/727
14	K	0.52	0/937	0.70	0/1265
15	L	0.47	0/365	0.74	0/485
All	All	0.54	6/32440 (0.0%)	0.75	15/43948 (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
5	B	0	1
6	C	0	1
All	All	0	2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	T	10	DA	O3'-P	-9.11	1.50	1.61

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	N	6	DC	O3'-P	7.11	1.69	1.61
12	I	78	CYS	CB-SG	-6.30	1.71	1.82
3	P	3	G	P-OP1	-6.03	1.38	1.49
5	B	503	GLY	CA-C	6.02	1.61	1.51

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	T	10	DA	OP1-P-O3'	7.38	121.44	105.20
3	P	3	G	O5'-P-OP1	-7.04	99.36	105.70
3	P	2	A	C2'-C3'-O3'	6.95	124.82	113.70
2	N	6	DC	P-O3'-C3'	6.32	127.28	119.70
5	B	1185	CYS	N-CA-C	-6.23	94.17	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	B	503	GLY	Mainchain
6	C	82	TYR	Sidechain

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	T	387	0	216	24	0
2	N	141	0	81	8	0
3	P	214	0	111	13	0
4	A	11140	0	11217	1300	0
5	B	8836	0	8871	1003	0
6	C	2095	0	2051	255	0
7	D	1356	0	1319	117	0
8	E	1752	0	1776	148	0
9	F	679	0	701	86	0
10	G	1340	0	1357	157	0
11	H	1068	0	1040	110	0
12	I	971	0	930	105	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
13	J	532	0	542	94	0
14	K	919	0	929	93	0
15	L	363	0	387	45	0
16	A	2	0	0	0	0
16	B	1	0	0	0	0
16	C	1	0	0	0	0
16	I	2	0	0	0	0
16	J	1	0	0	0	0
16	L	1	0	0	0	0
17	A	1	0	0	0	0
All	All	31802	0	31528	3238	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 51.

The worst 5 of 3238 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)
7:D:40:HIS:HB3	10:G:73:LYS:NZ	1.61	1.14
14:K:47:ARG:HB3	14:K:47:ARG:HH11	1.00	1.14
4:A:53:LEU:HD23	4:A:54:ASN:H	1.08	1.12
4:A:76:GLU:O	4:A:76:GLU:HG3	1.53	1.08
4:A:53:LEU:HD23	4:A:54:ASN:N	1.70	1.07

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
4	A	1406/1733 (81%)	936 (67%)	311 (22%)	159 (11%)	0 6
5	B	1096/1224 (90%)	740 (68%)	215 (20%)	141 (13%)	0 4

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
6	C	264/318 (83%)	166 (63%)	64 (24%)	34 (13%)	0 4
7	D	173/221 (78%)	118 (68%)	38 (22%)	17 (10%)	0 10
8	E	212/215 (99%)	154 (73%)	44 (21%)	14 (7%)	1 17
9	F	82/155 (53%)	63 (77%)	16 (20%)	3 (4%)	3 28
10	G	169/171 (99%)	133 (79%)	24 (14%)	12 (7%)	1 16
11	H	129/146 (88%)	85 (66%)	28 (22%)	16 (12%)	0 5
12	I	117/122 (96%)	79 (68%)	27 (23%)	11 (9%)	0 11
13	J	63/70 (90%)	34 (54%)	15 (24%)	14 (22%)	0 1
14	K	112/120 (93%)	87 (78%)	17 (15%)	8 (7%)	1 16
15	L	44/70 (63%)	17 (39%)	18 (41%)	9 (20%)	0 2
All	All	3867/4565 (85%)	2612 (68%)	817 (21%)	438 (11%)	0 6

5 of 438 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	A	5	GLN
4	A	48	ALA
4	A	54	ASN
4	A	55	ASP
4	A	57	ARG

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
4	A	1239/1520 (82%)	1125 (91%)	114 (9%)	9 32
5	B	964/1061 (91%)	873 (91%)	91 (9%)	8 31
6	C	234/274 (85%)	213 (91%)	21 (9%)	9 34
7	D	140/200 (70%)	124 (89%)	16 (11%)	5 25
8	E	196/197 (100%)	188 (96%)	8 (4%)	30 57
9	F	74/137 (54%)	65 (88%)	9 (12%)	5 23

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
10	G	152/152 (100%)	139 (91%)	13 (9%)	10 37
11	H	117/128 (91%)	109 (93%)	8 (7%)	16 44
12	I	113/116 (97%)	98 (87%)	15 (13%)	4 21
13	J	60/65 (92%)	55 (92%)	5 (8%)	11 38
14	K	99/102 (97%)	90 (91%)	9 (9%)	9 33
15	L	40/57 (70%)	37 (92%)	3 (8%)	13 41
All	All	3428/4009 (86%)	3116 (91%)	312 (9%)	9 33

5 of 312 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
5	B	463	THR
5	B	901	PRO
12	I	78	CYS
5	B	485	ARG
5	B	636	PRO

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

Mol	Chain	Res	Type
5	B	484	ASN
5	B	975	GLN
11	H	131	ASN
5	B	516	ASN
5	B	734	HIS

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	P	9/10 (90%)	1 (11%)	1 (11%)

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	P	3	G

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
3	P	2	A

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 9 ligands modelled in this entry, 9 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

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.