



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 9, 2023 – 10:10 am GMT

PDB ID : 2JA8  
Title : CPD lesion containing RNA Polymerase II elongation complex D  
Authors : Brueckner, F.; Hennecke, U.; Carell, T.; Cramer, P.  
Deposited on : 2006-11-23  
Resolution : 3.80 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](mailto: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  
Mogul : 1.8.4, CSD as541be (2020)  
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

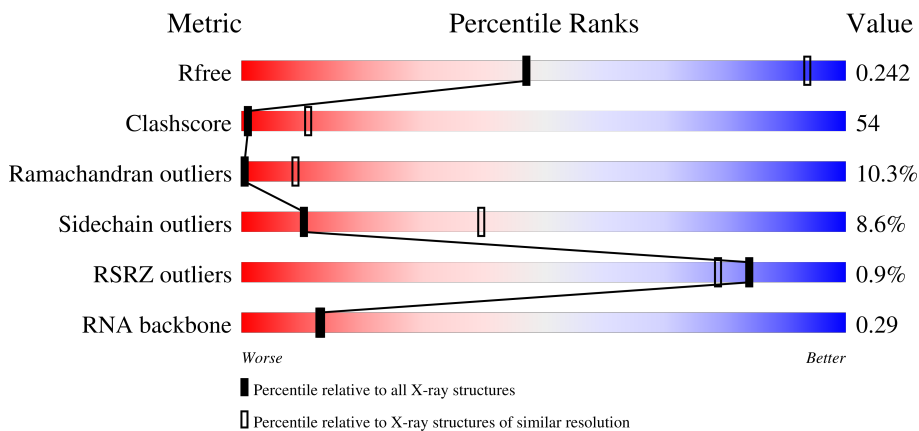
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.80 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.



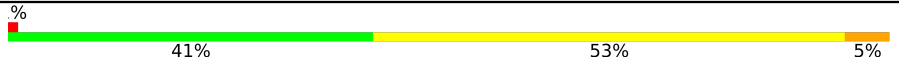


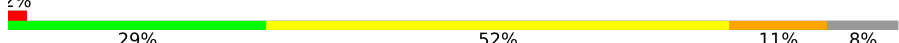
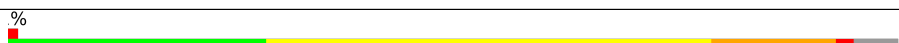
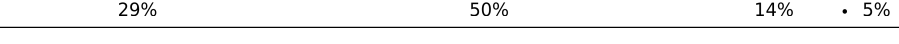
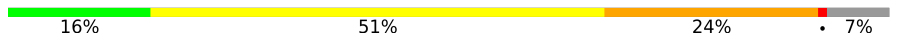
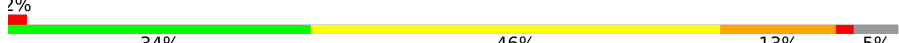



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)
RNA backbone	3102	1036 (4.60-3.00)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	1733	 25% 46% 10% 18%
2	B	1224	 28% 52% 11% 9%
3	C	318	 22% 51% 10% 16%
4	D	221	 27% 40% 12% 20%

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	E	215	
6	F	155	
7	G	171	
8	H	146	
9	I	122	
10	J	70	
11	K	120	
12	L	70	
13	N	14	
14	P	11	
15	T	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
15	TT	T	19	-	-	X	-

## 2 Entry composition

There are 17 unique types of molecules in this entry. The entry contains 32000 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 II LARGEST SUB-UNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1421	11186	7048	1958	2118	62	0	0	0

- Molecule 2 is a protein called DNA-DIRECTED RNA POLYMERASE II 140 KDA POLYPEPTIDE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1115	8866	5614	1553	1644	55	0	0	0

- Molecule 3 is a protein called DNA-DIRECTED RNA POLYMERASE II 45KDA POLYPEPTIDE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	267	2101	1320	349	419	13	0	0	0

- Molecule 4 is a protein called DNA-DIRECTED RNA POLYMERASE II 32KDA POLYPEPTIDE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	177	1427	882	256	287	2	0	0	0

- Molecule 5 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III 27 KDA POLYPEPTIDE.

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

- Molecule 6 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III 23

## KDA POLYPEPTIDE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	F	87	705	451	119	132	3	0	0	0

- Molecule 7 is a protein called DNA-DIRECTED RNA POLYMERASE II 19KDA POLYPEPTIDE.

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

- Molecule 8 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III 14.5 KDA POLYPEPTIDE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	H	135	1084	683	183	214	4	0	0	0

- Molecule 9 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT 9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
9	I	116	944	581	172	181	10	0	0	0

- Molecule 10 is a protein called DNA-DIRECTED RNA POLYMERASES I/II/III SUBUNIT 10.

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

- Molecule 11 is a protein called DNA-DIRECTED RNA POLYMERASE II 13.6 KDA POLYPEPTIDE.

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

- Molecule 12 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III 7.7 KDA POLYPEPTIDE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
12	L	46	364	224	72	64	4	0	0	0

- Molecule 13 is a DNA chain called 5'-D(\*TP\*AP\*AP\*GP\*TP\*AP\*CP\*TP\*TP\*GP\*AP\*GP\*CP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
13	N	8	161	79	29	46	7	0	0	0

- Molecule 14 is a RNA chain called 5'-R(\*UP\*UP\*CP\*GP\*AP\*CP\*CP\*AP\*GP\*AP\*UP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
14	P	10	209	95	38	67	9	0	0	0

- Molecule 15 is a DNA chain called 5'-D(\*AP\*GP\*CP\*TP\*CP\*AP\*AP\*GP\*TP\*AP\*CP\*TP\*TP\*TP\*TP\*TTP\*CP\*BRUP\*GP\*GP\*TP\*CP\*AP\*TP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	Br	C	N	O				P
15	T	19	401	1	197	61	124	18	0	0	0

- Molecule 16 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

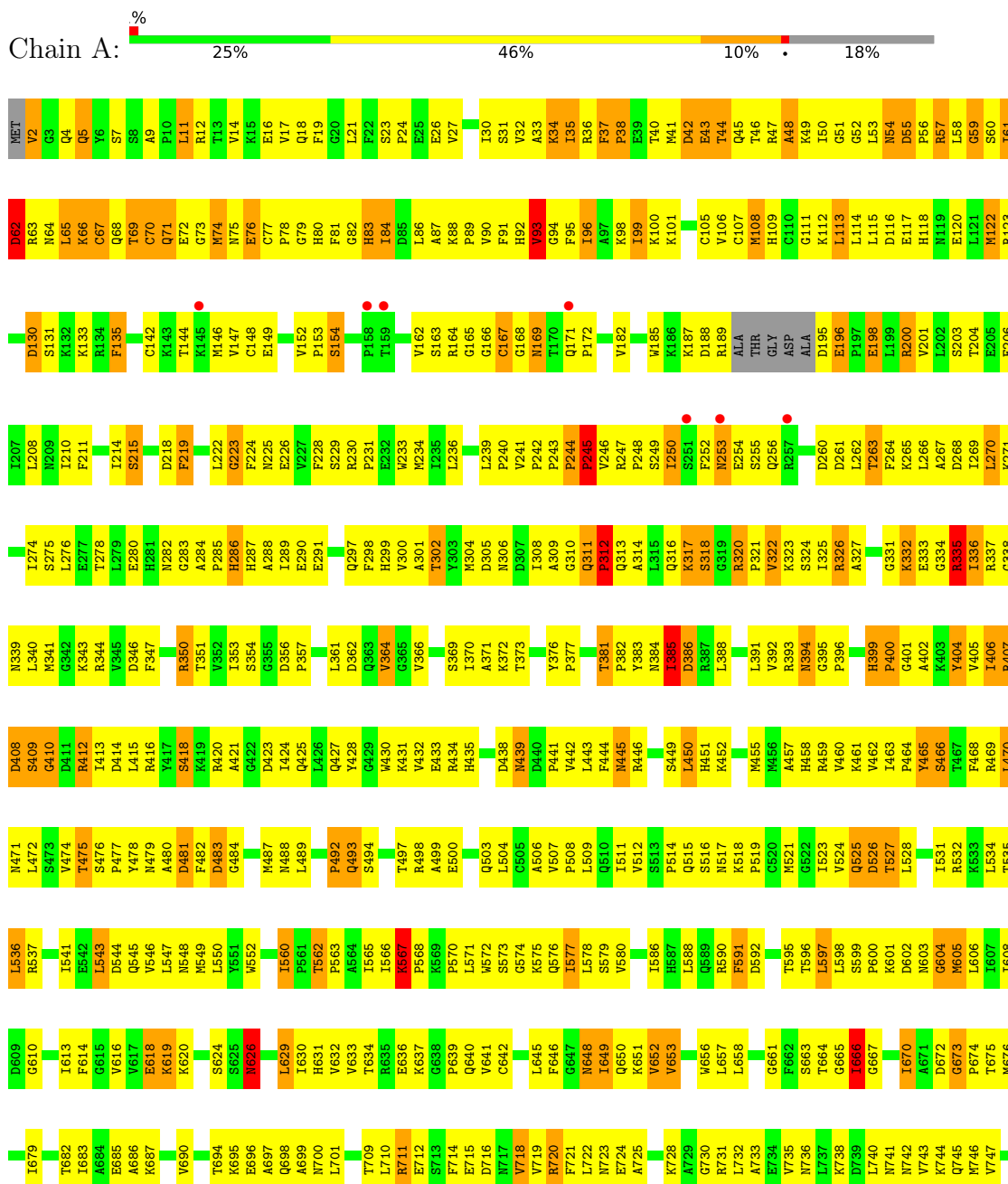
- Molecule 17 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
17	A	8	Total	Zn	0	0
			8	8		

### 3 Residue-property plots

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.

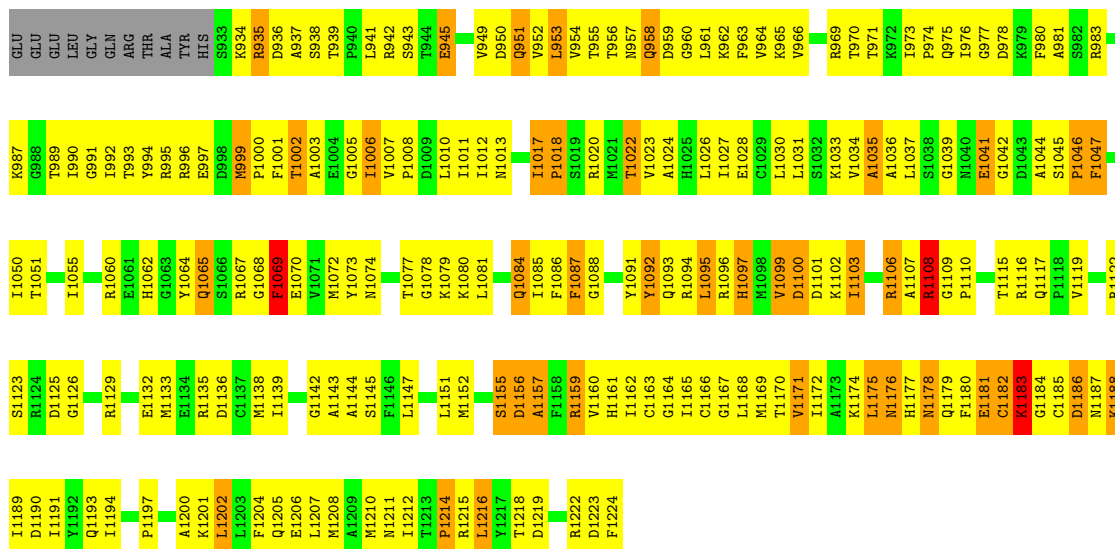
- Molecule 1: DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT



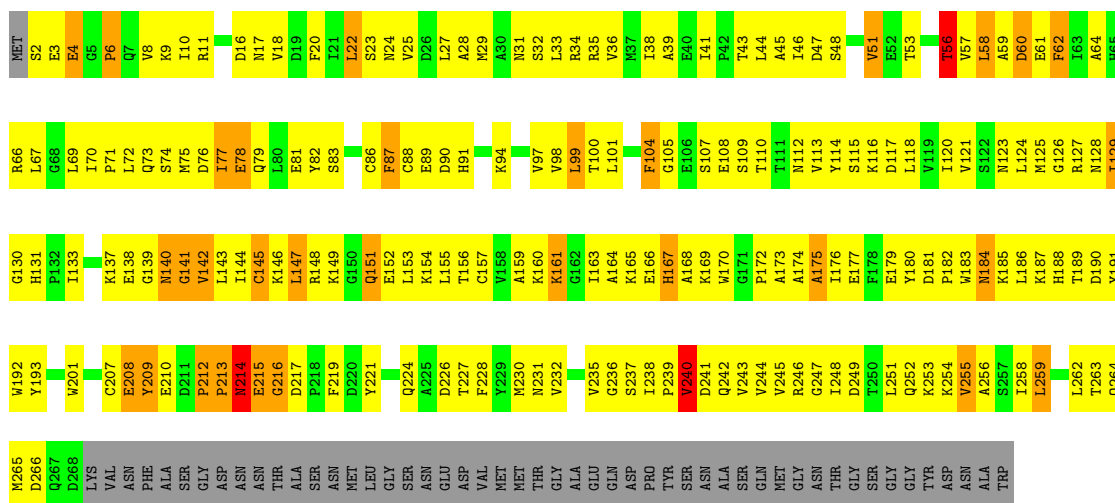




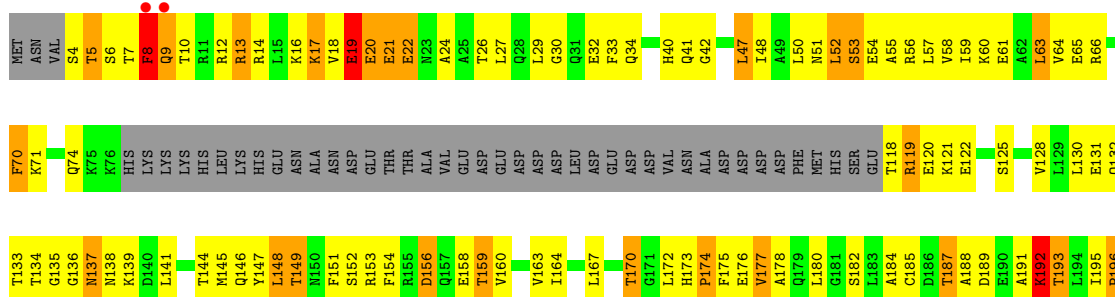




● Molecule 3: DNA-DIRECTED RNA POLYMERASE II 45KDA POLYPEPTIDE

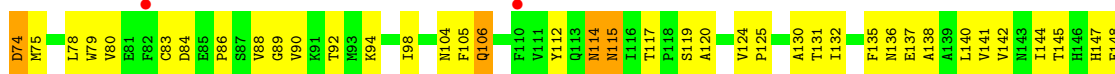
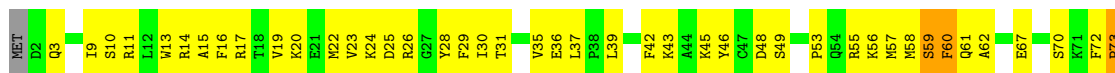


● Molecule 4: DNA-DIRECTED RNA POLYMERASE II 32KDA POLYPEPTIDE

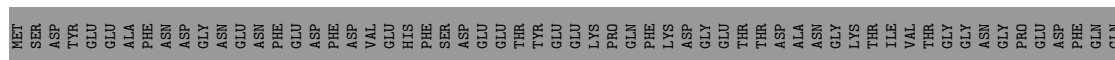
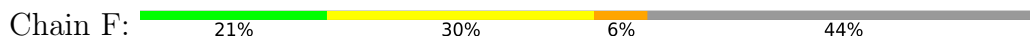




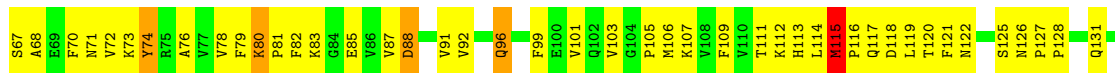
- Molecule 5: DNA-DIRECTED RNA POLYMERASES I, II, AND III 27 KDA POLYPEPTIDE



- Molecule 6: DNA-DIRECTED RNA POLYMERASES I, II, AND III 23 KDA POLYPEPTIDE

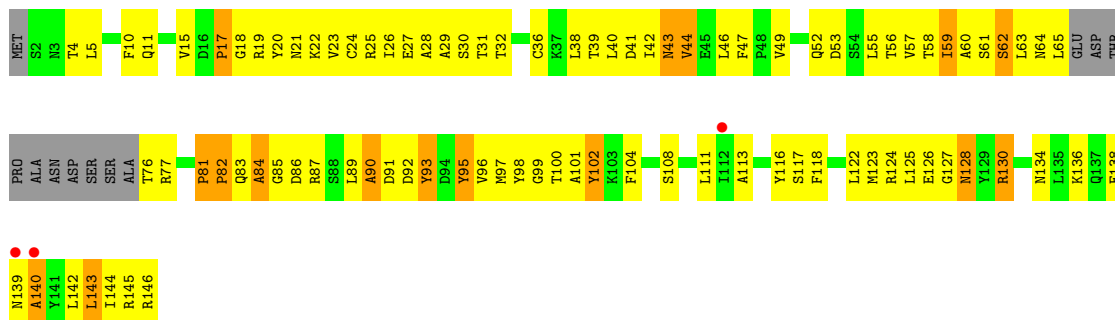


- Molecule 7: DNA-DIRECTED RNA POLYMERASE II 19KDA POLYPEPTIDE

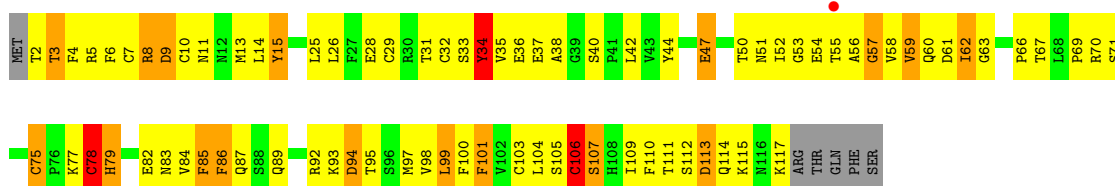


- Molecule 8: DNA-DIRECTED RNA POLYMERASES I, II, AND III 14.5 KDA POLYPEPTIDE

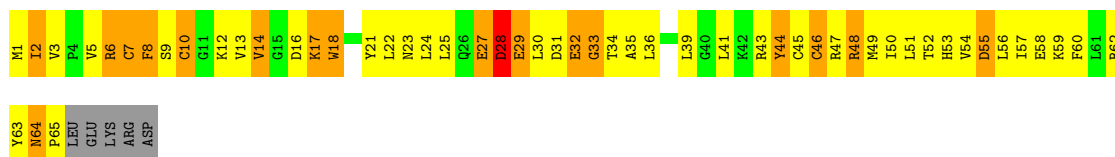
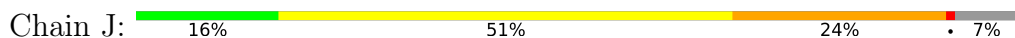




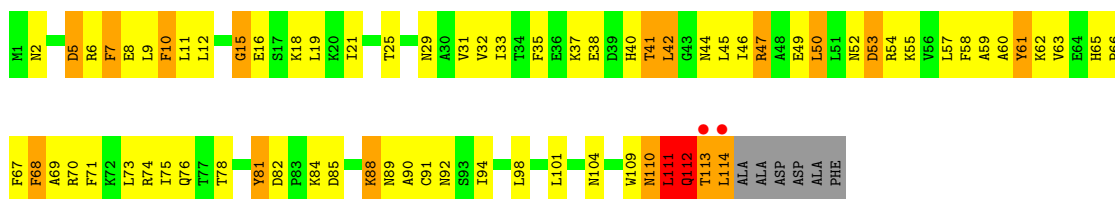
- Molecule 9: DNA-DIRECTED RNA POLYMERASE II SUBUNIT 9



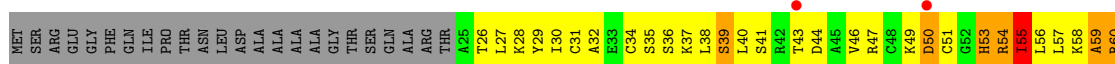
- Molecule 10: DNA-DIRECTED RNA POLYMERASES I/II/III SUBUNIT 10

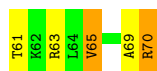


- Molecule 11: DNA-DIRECTED RNA POLYMERASE II 13.6 KDA POLYPEPTIDE

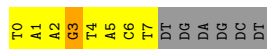
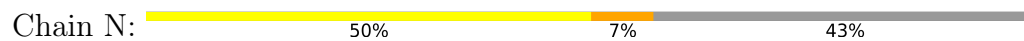


- Molecule 12: DNA-DIRECTED RNA POLYMERASES I, II, AND III 7.7 KDA POLYPEPTIDE

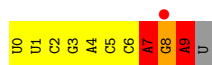




- Molecule 13: 5'-D(\*TP\*AP\*AP\*GP\*TP\*AP\*CP\*TP\*TP\*GP \*AP\*GP\*CP\*T)-3'



- Molecule 14: 5'-R(\*UP\*UP\*CP\*GP\*AP\*CP\*CP\*AP\*GP\*AP\*UP)-3'



- Molecule 15: 5'-D(\*AP\*GP\*CP\*TP\*CP\*AP\*AP\*GP\*TP\*AP \*CP\*TP\*TP\*TP\*TP\*TTP\*CP \*BRUP\*GP\*GP\*TP\*CP\*AP\*TP\*T)-3'



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	222.64Å 392.85Å 282.89Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.80 48.85 – 3.80	Depositor EDS
% Data completeness (in resolution range)	99.2 (50.00-3.80) 99.9 (48.85-3.80)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.30 (at 3.77Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.269 , 0.288 0.210 , 0.242	Depositor DCC
$R_{free}$ test set	4685 reflections (1.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	110.0	Xtrriage
Anisotropy	0.478	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 70.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	0.015 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.023 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	32000	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	95.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.78% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: MG, TT, BRU, 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	0.46	0/11385	0.73	1/15393 (0.0%)
2	B	0.45	0/9037	0.71	2/12181 (0.0%)
3	C	0.48	0/2138	0.71	0/2896
4	D	0.42	0/1437	0.67	0/1925
5	E	0.43	1/1788 (0.1%)	0.62	0/2406
6	F	0.53	0/716	0.77	0/964
7	G	0.49	0/1368	0.73	0/1844
8	H	0.38	0/1102	0.65	0/1492
9	I	0.39	0/962	0.68	0/1295
10	J	0.48	0/541	0.80	0/727
11	K	0.67	3/937 (0.3%)	0.86	4/1265 (0.3%)
12	L	0.41	0/366	0.68	0/485
13	N	1.21	1/180 (0.6%)	1.08	0/276
14	P	0.85	0/233	1.41	2/361 (0.6%)
15	T	1.12	2/380 (0.5%)	1.39	3/580 (0.5%)
All	All	0.49	7/32570 (0.0%)	0.74	12/44090 (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	B	0	1
15	T	0	1
All	All	0	2

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	K	113	THR	N-CA	5.90	1.58	1.46

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	K	113	THR	CA-C	5.73	1.67	1.52
15	T	12	DT	N1-C2	5.38	1.42	1.38
5	E	214	CYS	CB-SG	-5.28	1.73	1.81
11	K	114	LEU	N-CA	5.10	1.56	1.46

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
11	K	113	THR	N-CA-C	9.56	136.81	111.00
15	T	24	DG	O4'-C1'-N9	7.22	113.06	108.00
15	T	21	DC	O5'-P-OP1	7.09	119.21	110.70
14	P	9	A	C2'-C3'-O3'	7.01	124.92	109.50
14	P	7	A	N9-C1'-C2'	-6.77	104.56	112.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	486	TYR	Sidechain
15	T	13	DA	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	A	11186	0	11266	1310	0
2	B	8866	0	8898	1015	0
3	C	2101	0	2055	265	0
4	D	1427	0	1451	139	0
5	E	1752	0	1776	134	0
6	F	705	0	730	89	0
7	G	1340	0	1357	173	0
8	H	1084	0	1057	127	0
9	I	944	0	900	110	0
10	J	532	0	542	99	0
11	K	919	0	929	107	0
12	L	364	0	386	47	0

Continued on next page...



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
13	N	161	0	93	17	0
14	P	209	0	109	28	0
15	T	401	0	231	64	0
16	A	1	0	0	0	0
17	A	8	0	0	0	0
All	All	32000	0	31780	3425	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 54.

The worst 5 of 3425 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:A:53:LEU:HD23	1:A:54:ASN:N	1.61	1.13
1:A:855:THR:HG21	1:A:857:ARG:HE	1.10	1.11
7:G:14:HIS:CD2	7:G:16:SER:HB2	1.87	1.10
15:T:19:TT:H5R1	15:T:19:TT:C1'	1.83	1.09
1:A:1445:ILE:H	1:A:1445:ILE:HD12	1.16	1.09

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	1410/1733 (81%)	962 (68%)	299 (21%)	149 (11%)	0 8
2	B	1096/1224 (90%)	767 (70%)	219 (20%)	110 (10%)	0 9
3	C	264/318 (83%)	171 (65%)	62 (24%)	31 (12%)	0 6
4	D	173/221 (78%)	125 (72%)	29 (17%)	19 (11%)	0 8
5	E	212/215 (99%)	157 (74%)	44 (21%)	11 (5%)	2 23

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
6	F	84/155 (54%)	67 (80%)	15 (18%)	2 (2%)	6	37
7	G	169/171 (99%)	125 (74%)	34 (20%)	10 (6%)	1	21
8	H	131/146 (90%)	84 (64%)	30 (23%)	17 (13%)	0	5
9	I	114/122 (93%)	80 (70%)	21 (18%)	13 (11%)	0	7
10	J	63/70 (90%)	35 (56%)	13 (21%)	15 (24%)	0	1
11	K	112/120 (93%)	86 (77%)	16 (14%)	10 (9%)	1	12
12	L	44/70 (63%)	17 (39%)	17 (39%)	10 (23%)	0	1
All	All	3872/4565 (85%)	2676 (69%)	799 (21%)	397 (10%)	0	9

5 of 397 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	44	THR
1	A	48	ALA
1	A	57	ARG
1	A	62	ASP
1	A	65	LEU

### 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	1244/1520 (82%)	1138 (92%)	106 (8%)	10	40
2	B	967/1061 (91%)	890 (92%)	77 (8%)	12	42
3	C	235/274 (86%)	215 (92%)	20 (8%)	10	40
4	D	159/200 (80%)	136 (86%)	23 (14%)	3	19
5	E	196/197 (100%)	191 (97%)	5 (3%)	46	69
6	F	77/137 (56%)	68 (88%)	9 (12%)	5	27
7	G	152/152 (100%)	140 (92%)	12 (8%)	12	42
8	H	119/128 (93%)	112 (94%)	7 (6%)	19	51
9	I	110/116 (95%)	97 (88%)	13 (12%)	5	26

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
10	J	60/65 (92%)	53 (88%)	7 (12%)	5	27
11	K	99/102 (97%)	86 (87%)	13 (13%)	4	23
12	L	40/57 (70%)	36 (90%)	4 (10%)	7	32
All	All	3458/4009 (86%)	3162 (91%)	296 (9%)	10	40

5 of 296 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	F	79	ARG
11	K	68	PHE
6	F	148	VAL
9	I	9	ASP
1	A	1385	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 83 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	1193	GLN
5	E	147	HIS
3	C	112	ASN
4	D	40	HIS
7	G	126	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
14	P	9/11 (81%)	3 (33%)	1 (11%)

All (3) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
14	P	7	A
14	P	8	G
14	P	9	A

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
14	P	7	A

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

2 non-standard protein/DNA/RNA residues are modelled in this entry.

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	$\# Z  > 2$
15	BRU	T	22	15,14	18,21,22	0.44	0	26,30,33	0.99	1 (3%)
15	TT	T	19	15	40,43,44	4.90	12 (30%)	59,69,72	2.36	17 (28%)

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
15	BRU	T	22	15,14	-	0/7/21/22	0/2/2/2
15	TT	T	19	15	-	11/18/105/106	0/5/6/6

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
15	T	19	TT	C5-C6	-21.02	1.31	1.55
15	T	19	TT	C5T-C6T	-20.02	1.32	1.55
15	T	19	TT	C6-N1	-5.36	1.38	1.46
15	T	19	TT	C1'-N1	4.04	1.50	1.45
15	T	19	TT	C6T-N1T	-3.72	1.40	1.46

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
15	T	19	TT	C5-C6-N1	8.44	127.43	115.61
15	T	19	TT	C5T-C6T-N1T	6.14	124.22	115.61

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
15	T	19	TT	C5T-C5-C6	6.02	95.86	88.38
15	T	19	TT	C5-C6-C6T	-5.71	79.91	89.28
15	T	19	TT	C5-C5T-C6T	-5.45	81.59	88.38

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
15	T	19	TT	C4'-C5'-O5'-P
15	T	19	TT	O3'-C7-O5R-C5R
15	T	19	TT	C2R-C1R-N1T-C6T
15	T	19	TT	C2'-C1'-N1-C6
15	T	19	TT	C2'-C1'-N1-C2

There are no ring outliers.

2 monomers are involved in 31 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
15	T	22	BRU	5	0
15	T	19	TT	26	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

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 [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	B	2
3	C	1
6	F	1
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	2:SER	C	3:GLU	N	4.15
1	B	18:PHE	C	19:GLU	N	3.85
1	F	69:LEU	C	70:LYS	N	3.50
1	A	1175:SER	C	1176:LEU	N	3.39
1	B	337:ARG	C	338:GLY	N	2.64

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	1421/1733 (81%)	-0.37	9 (0%) 89 85	17, 78, 152, 199	0
2	B	1115/1224 (91%)	-0.28	12 (1%) 80 74	20, 91, 167, 200	0
3	C	267/318 (83%)	-0.41	0 100 100	39, 77, 139, 165	0
4	D	177/221 (80%)	-0.17	2 (1%) 80 74	61, 111, 160, 175	0
5	E	214/215 (99%)	-0.28	2 (0%) 84 79	50, 133, 184, 188	0
6	F	87/155 (56%)	-0.52	0 100 100	27, 56, 100, 138	0
7	G	171/171 (100%)	-0.28	0 100 100	57, 80, 126, 137	0
8	H	135/146 (92%)	0.24	3 (2%) 62 54	95, 138, 173, 183	0
9	I	116/122 (95%)	-0.12	1 (0%) 84 79	75, 134, 166, 186	0
10	J	65/70 (92%)	-0.61	0 100 100	42, 73, 117, 125	0
11	K	114/120 (95%)	-0.33	2 (1%) 68 61	38, 83, 112, 167	0
12	L	46/70 (65%)	0.16	2 (4%) 35 30	78, 156, 175, 181	0
13	N	8/14 (57%)	0.51	0 100 100	129, 140, 153, 155	0
14	P	10/11 (90%)	0.12	1 (10%) 7 6	115, 126, 152, 158	0
15	T	17/25 (68%)	0.37	0 100 100	125, 138, 151, 153	0
All	All	3963/4615 (85%)	-0.29	34 (0%) 84 79	17, 89, 165, 200	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
11	K	114	LEU	7.5
2	B	471	LYS	6.3
11	K	113	THR	5.0
2	B	882	THR	4.0
2	B	883	LEU	3.4

## 6.2 Non-standard residues in protein, DNA, RNA chains [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
15	BRU	T	22	20/21	0.78	0.24	112,118,124,128	0
15	TT	T	19	38/39	0.84	0.27	130,144,163,166	0

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
17	ZN	A	2461	1/1	0.95	0.12	200,200,200,200	0
17	ZN	A	2459	1/1	0.98	0.03	125,125,125,125	0
17	ZN	A	2464	1/1	0.98	0.07	87,87,87,87	0
17	ZN	A	2460	1/1	0.99	0.14	88,88,88,88	0
17	ZN	A	2458	1/1	0.99	0.16	57,57,57,57	0
17	ZN	A	2462	1/1	0.99	0.07	39,39,39,39	0
17	ZN	A	2463	1/1	0.99	0.14	47,47,47,47	0
16	MG	A	2457	1/1	0.99	0.16	35,35,35,35	0
17	ZN	A	2465	1/1	0.99	0.06	39,39,39,39	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.