



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 27, 2023 – 12:49 PM EDT

PDB ID : 3HKZ  
Title : The X-ray crystal structure of RNA polymerase from Archaea  
Authors : Murakami, K.S.  
Deposited on : 2009-05-26  
Resolution : 3.40 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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.35  
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) : 2.35

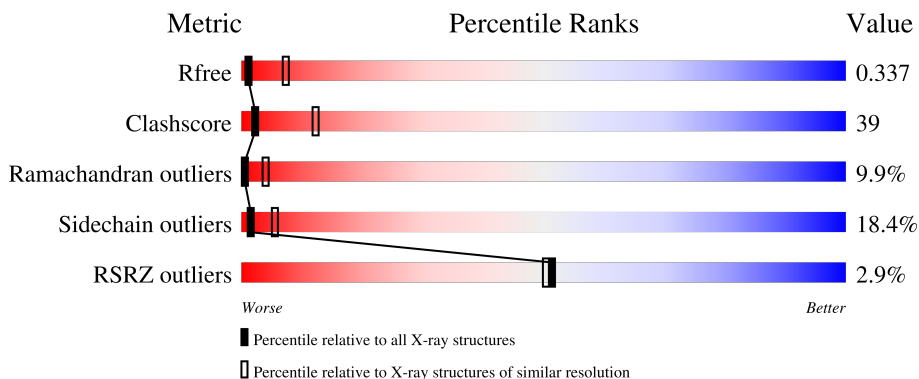
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.40 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	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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	880	 33% 46% 14% • 5%
1	I	880	 32% 48% 14% • 5%
2	C	395	 32% 43% 16% • 6%
2	M	395	 27% 44% 19% • 6%
3	B	1124	 32% 48% 15% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	J	1124	
4	D	265	
4	O	265	
5	E	180	
5	Q	180	
6	F	113	
6	R	113	
7	G	132	
7	S	132	
8	H	84	
8	T	84	
9	K	95	
9	U	95	
10	L	92	
10	V	92	
11	N	66	
11	W	66	
12	P	48	
12	X	48	
13	Y	104	
13	Z	104	

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
14	ZN	B	2001	-	-	X	-
14	ZN	J	2001	-	-	X	-

Continued on next page...

*Continued from previous page...*

<b>Mol</b>	<b>Type</b>	<b>Chain</b>	<b>Res</b>	<b>Chirality</b>	<b>Geometry</b>	<b>Clashes</b>	<b>Electron density</b>
16	F3S	O	1001	-	-	X	-

## 2 Entry composition [i](#)

There are 16 unique types of molecules in this entry. The entry contains 53072 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 A'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	836	6673	4248	1178	1221	26	0	0	0
1	I	836	6673	4248	1178	1221	26	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	370	2868	1818	490	551	9	0	0	0
2	M	370	2868	1818	490	551	9	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MET	-	expression tag	UNP P58192
C	2	GLU	-	expression tag	UNP P58192
C	3	GLY	-	expression tag	UNP P58192
M	1	MET	-	expression tag	UNP P58192
M	2	GLU	-	expression tag	UNP P58192
M	3	GLY	-	expression tag	UNP P58192

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	B	1090	8645	5483	1529	1602	31	0	0	0
3	J	1090	8645	5483	1529	1602	31	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	264	Total	C	N	O	S	0	0	0
			2114	1355	342	403	14			
4	O	264	Total	C	N	O	S	0	0	0
			2114	1355	342	403	14			

- Molecule 5 is a protein called DNA-directed RNA polymerase, subunit E' (RpoE1).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	176	Total	C	N	O	S	0	0	0
			1402	903	236	259	4			
5	Q	176	Total	C	N	O	S	0	0	0
			1402	903	236	259	4			

- Molecule 6 is a protein called DNA-directed RNA polymerase, subunit F (RpoF).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	89	Total	C	N	O	S	0	0	0
			694	433	115	142	4			
6	R	89	Total	C	N	O	S	0	0	0
			694	433	115	142	4			

- Molecule 7 is a protein called DNA-directed RNA polymerase, subunit G (RpoG).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	113	Total	C	N	O	S	0	0	0
			884	556	149	174	5			
7	S	113	Total	C	N	O	S	0	0	0
			884	556	149	174	5			

- Molecule 8 is a protein called DNA-directed RNA polymerase subunit H.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
8	H	74	Total	C	N	O	0	0	0
			611	397	109	105			
8	T	74	Total	C	N	O	0	0	0
			611	397	109	105			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	K	82	Total	C	N	O	S	0	0	0
			658	420	121	116	1			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	U	82	Total	C	N	O	S	0	0	0
			658	420	121	116	1			

- Molecule 10 is a protein called DNA-directed RNA polymerase subunit L.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	L	92	Total	C	N	O	S	0	0	0
			723	459	121	141	2			
10	V	92	Total	C	N	O	S	0	0	0
			723	459	121	141	2			

- Molecule 11 is a protein called DNA-directed RNA polymerase subunit N.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	N	64	Total	C	N	O	S	0	0	0
			514	326	94	88	6			
11	W	64	Total	C	N	O	S	0	0	0
			514	326	94	88	6			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	P	43	Total	C	N	O	S	0	0	0
			346	230	58	53	5			
12	X	43	Total	C	N	O	S	0	0	0
			346	230	58	53	5			

- Molecule 13 is a protein called DNA-directed RNA polymerase subunit 13.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	Y	45	Total	C	N	O	S	0	0	0
			391	245	68	77	1			
13	Z	45	Total	C	N	O	S	0	0	0
			391	245	68	77	1			

- Molecule 14 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
14	A	2	Total	Zn	0	0
			2	2		
14	B	1	Total	Zn	0	0
			1	1		

*Continued on next page...*

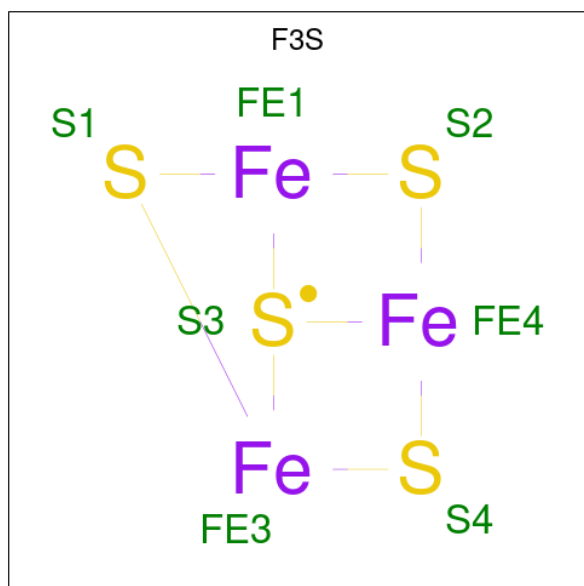
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
14	N	1	Total	Zn	0	0
			1	1		
14	P	1	Total	Zn	0	0
			1	1		
14	I	2	Total	Zn	0	0
			2	2		
14	J	1	Total	Zn	0	0
			1	1		
14	W	1	Total	Zn	0	0
			1	1		
14	X	1	Total	Zn	0	0
			1	1		

- Molecule 15 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
15	A	1	Total	Mg	0	0
			1	1		
15	I	1	Total	Mg	0	0
			1	1		

- Molecule 16 is FE3-S4 CLUSTER (three-letter code: F3S) (formula: Fe<sub>3</sub>S<sub>4</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
16	D	1	Total	Fe	S	0	0
			7	3	4		

Continued on next page...



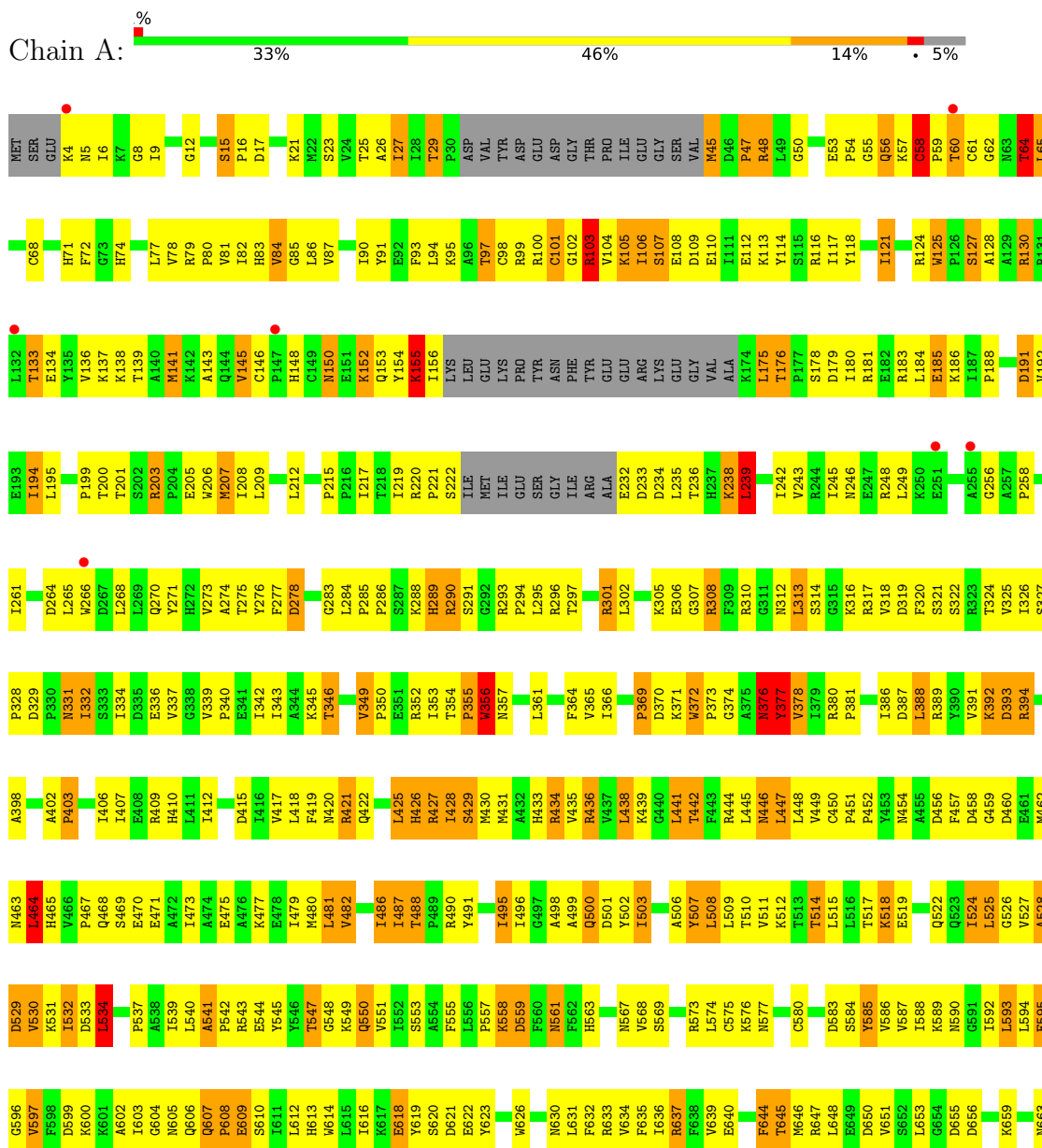
*Continued from previous page...*

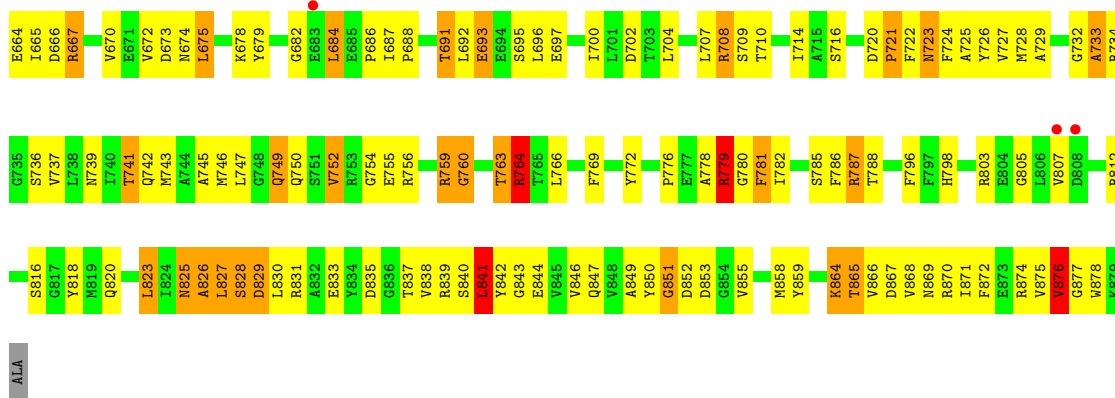
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	Fe	S		
16	O	1	7	3	4	0	0

### 3 Residue-property plots

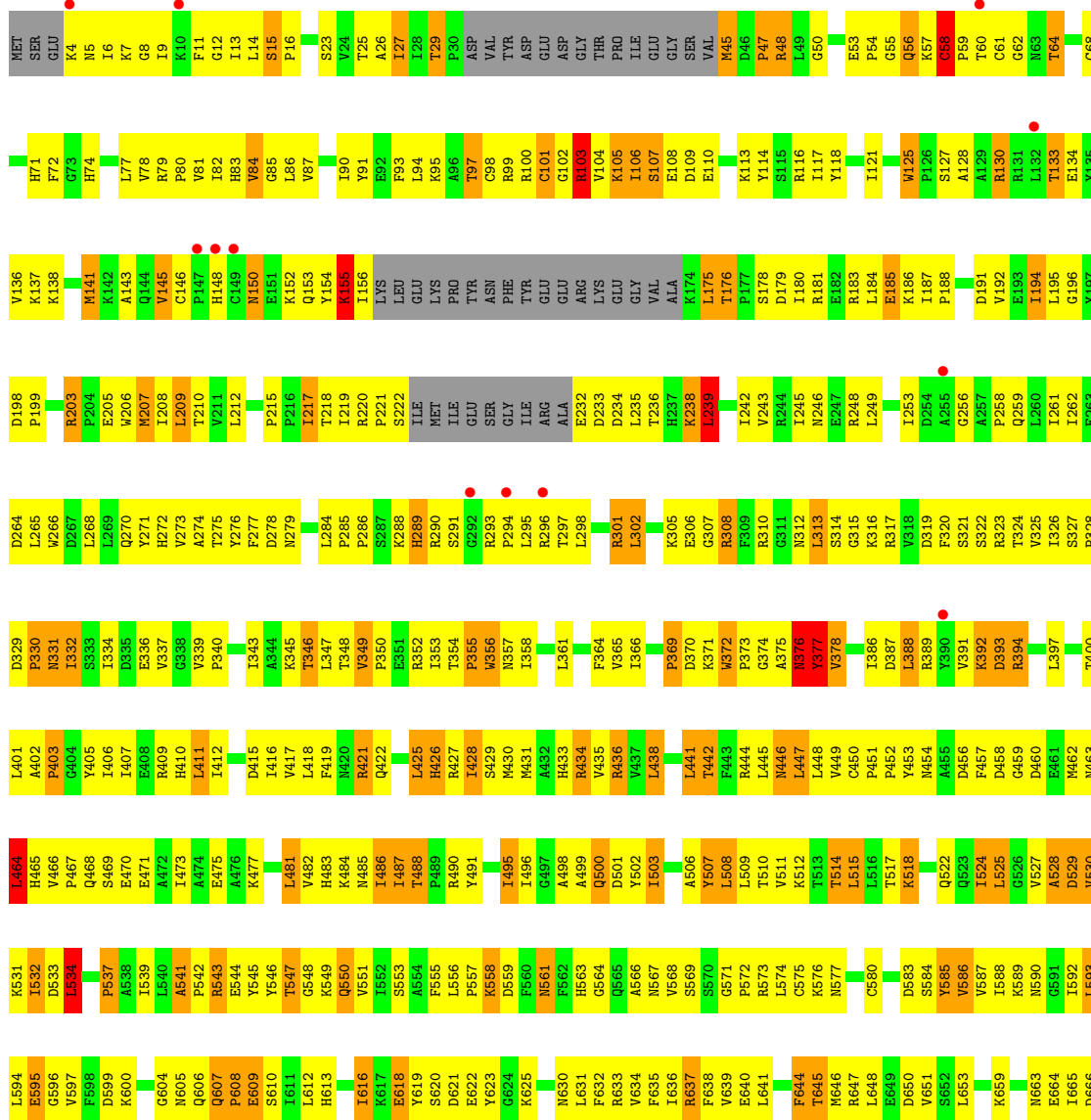
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.

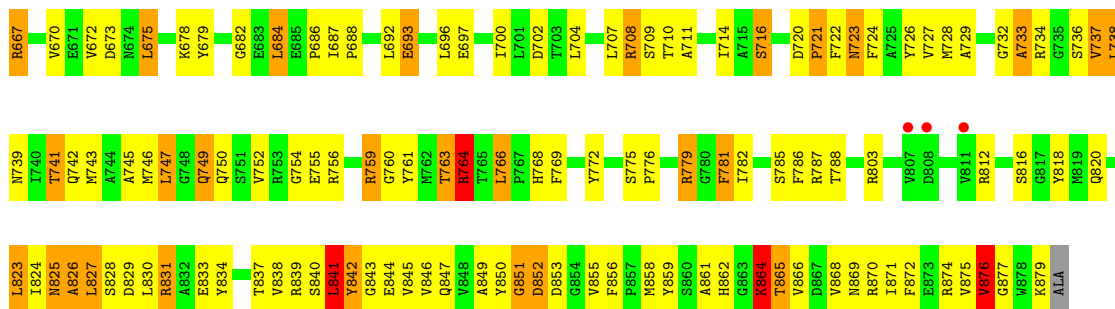
- Molecule 1: DNA-directed RNA polymerase subunit A'



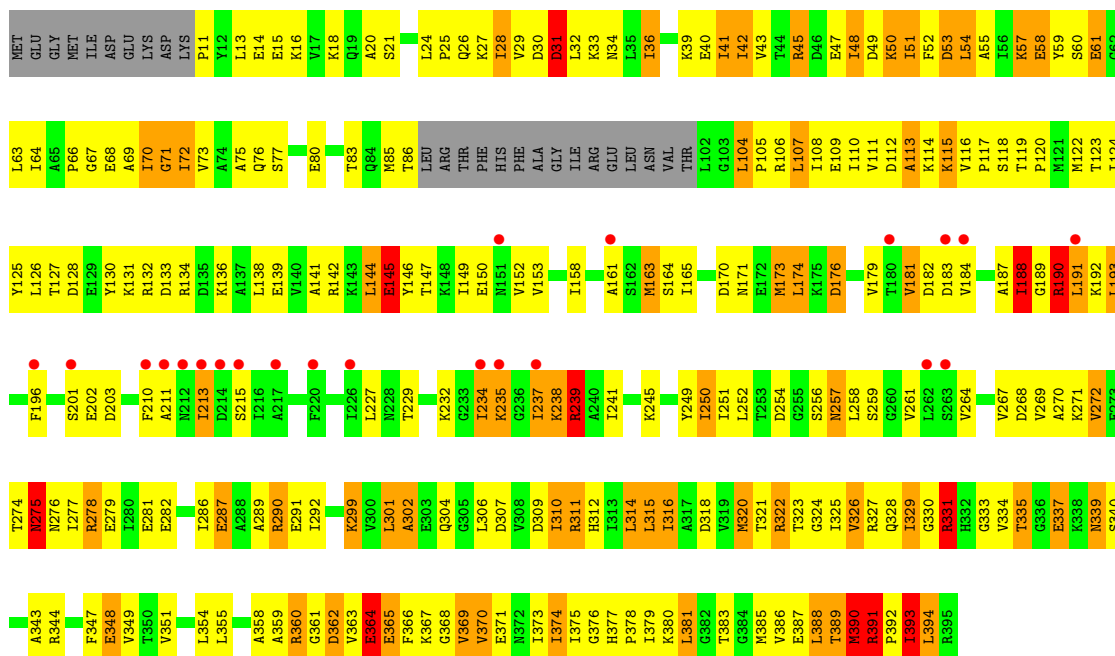


• Molecule 1: DNA-directed RNA polymerase subunit A'

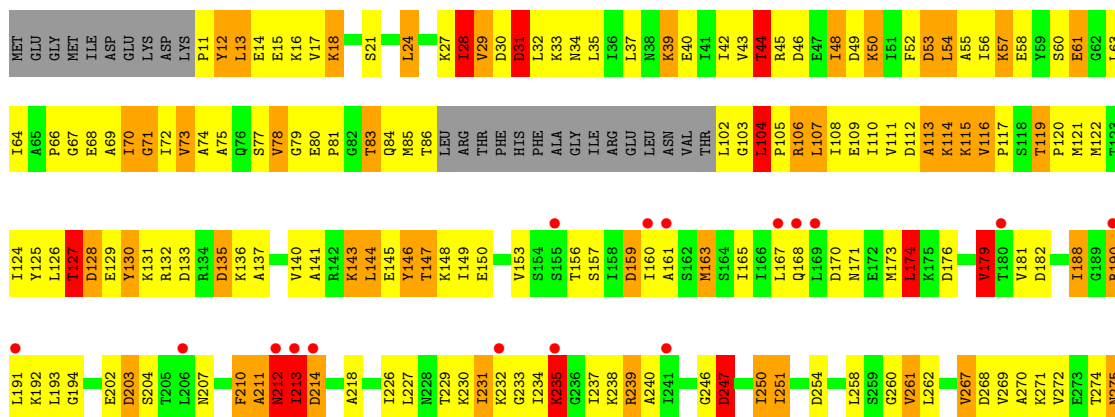


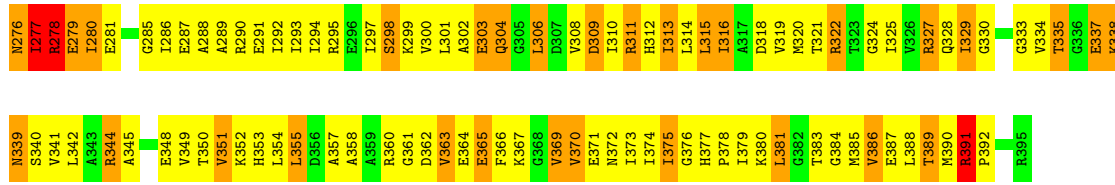


• Molecule 2: DNA-directed RNA polymerase subunit A''

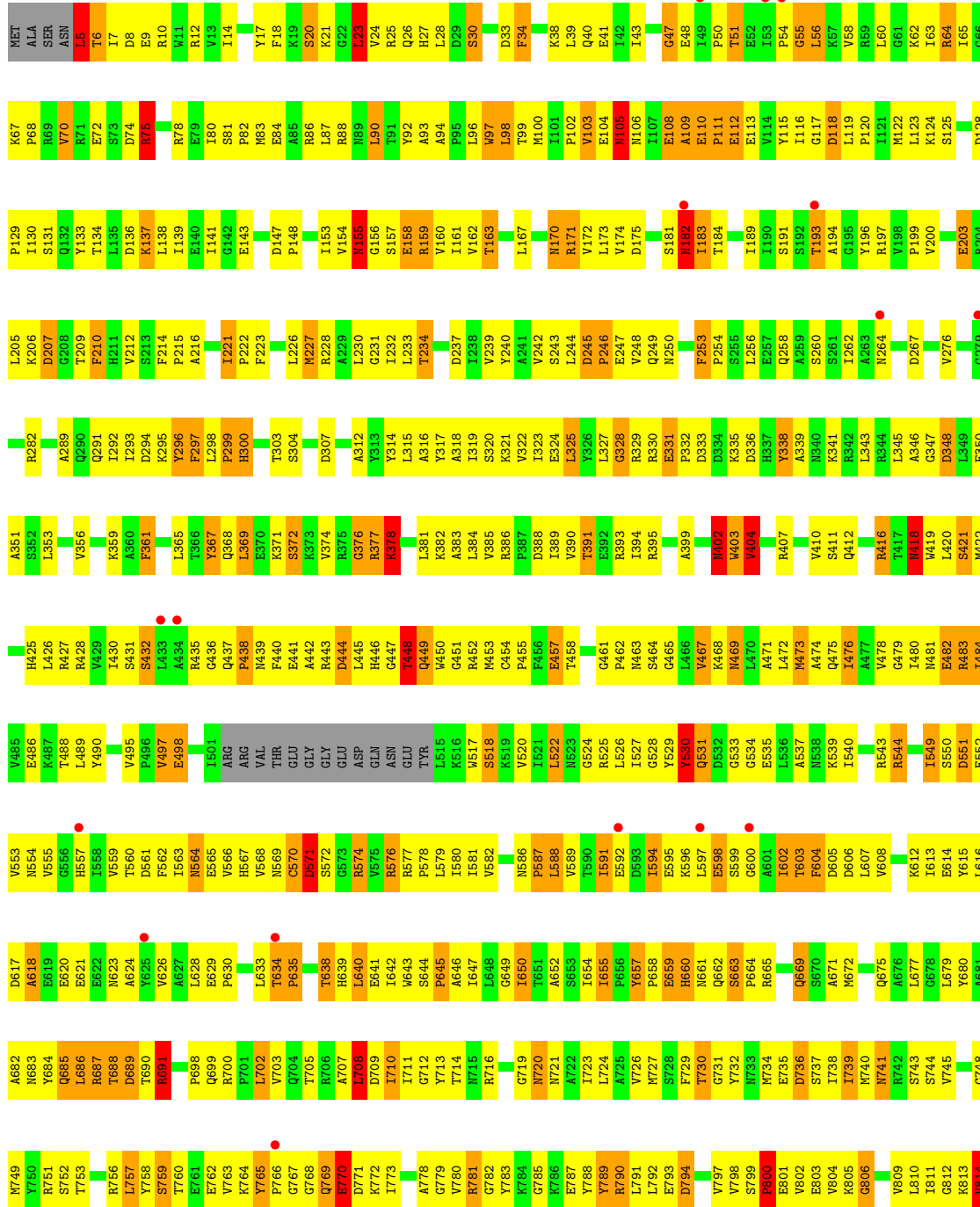


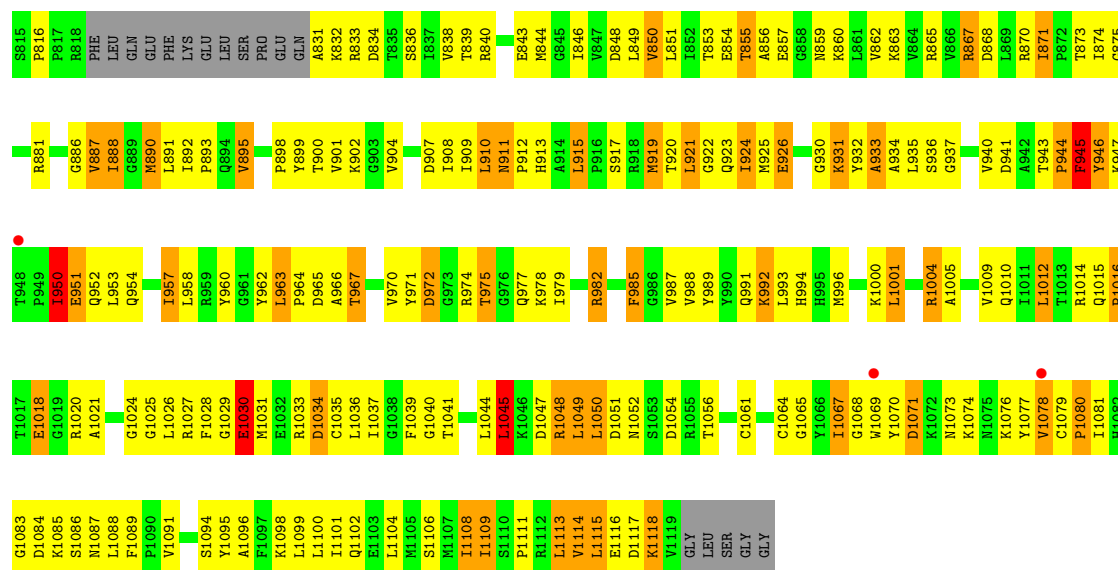
• Molecule 2: DNA-directed RNA polymerase subunit A''



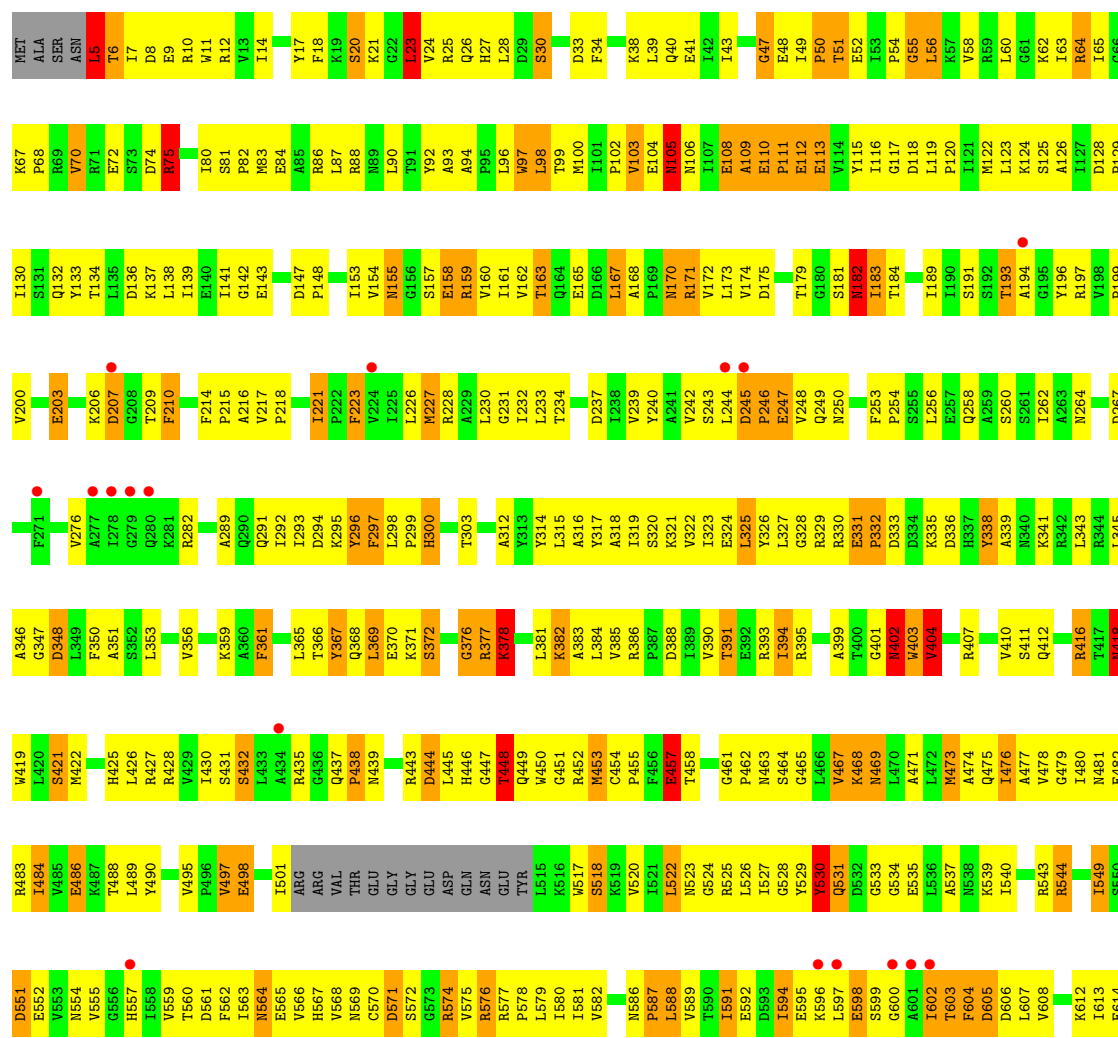


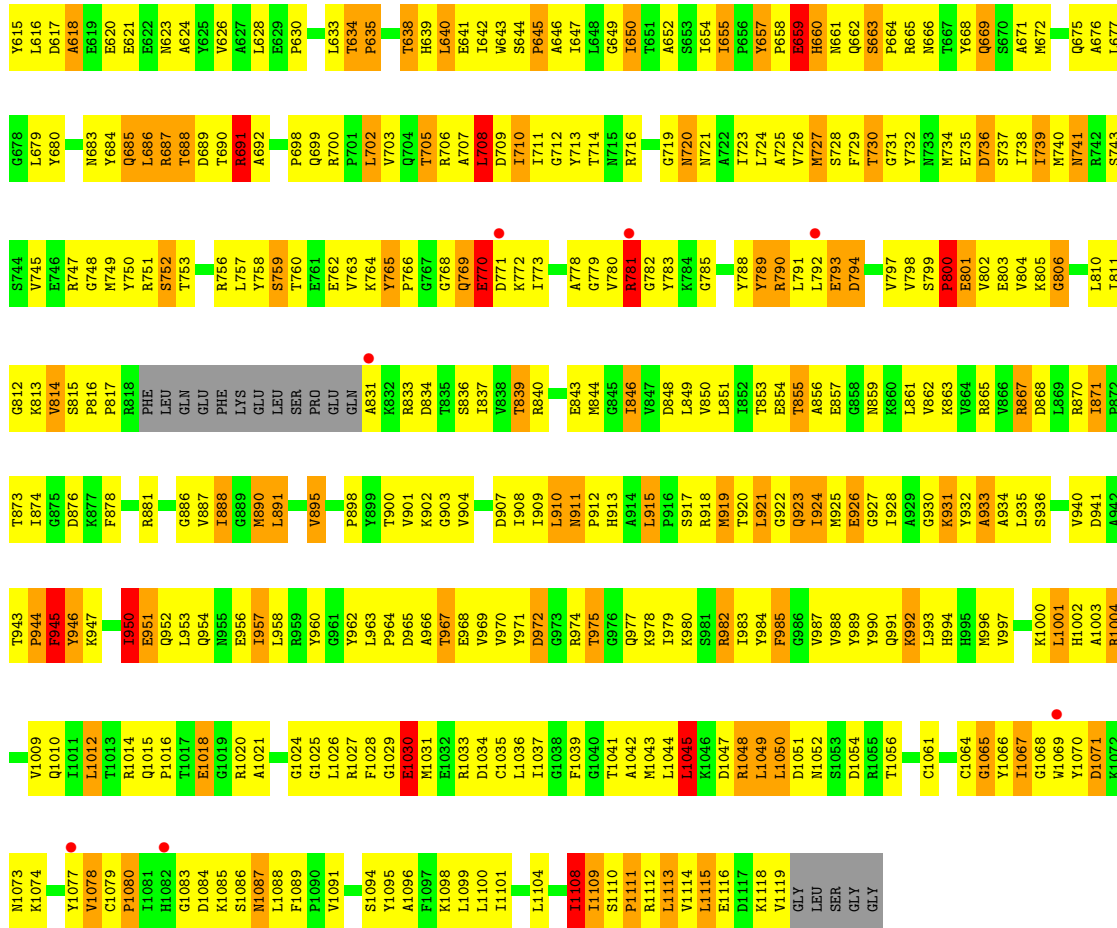
● Molecule 3: DNA-directed RNA polymerase subunit B



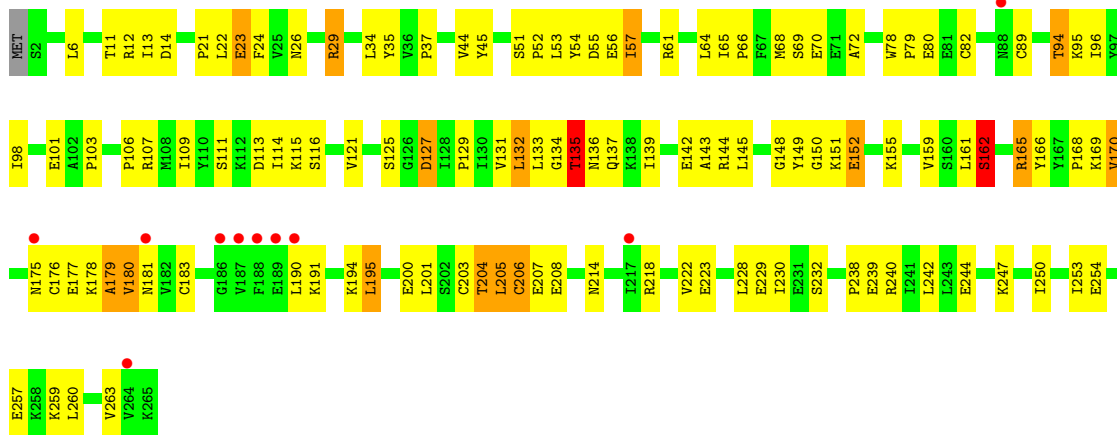


• Molecule 3: DNA-directed RNA polymerase subunit B



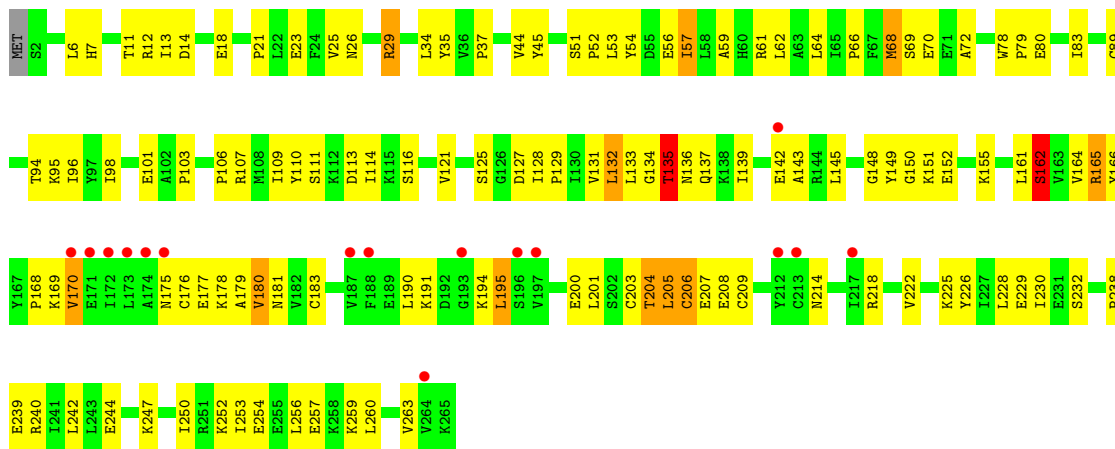


• Molecule 4: DNA-directed RNA polymerase subunit D

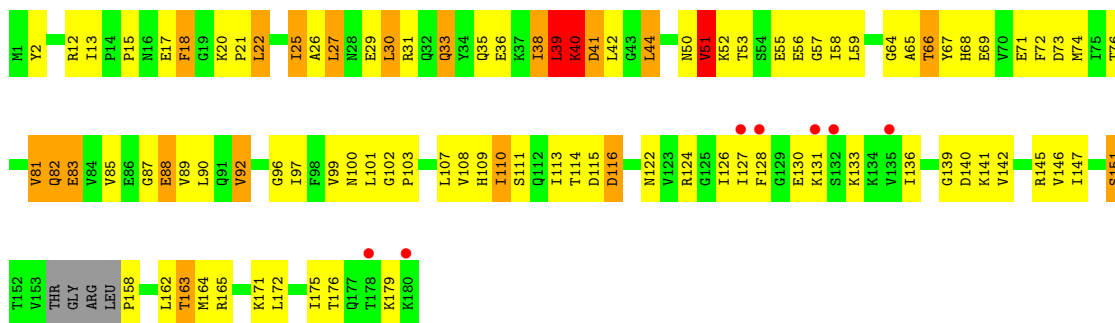


• Molecule 4: DNA-directed RNA polymerase subunit D

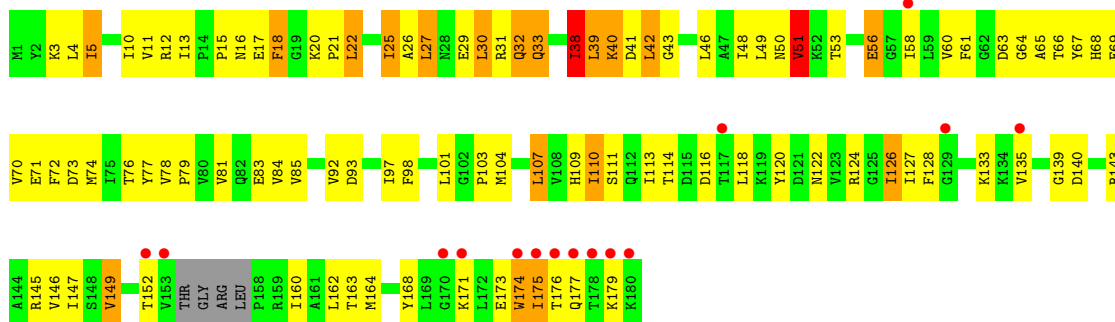
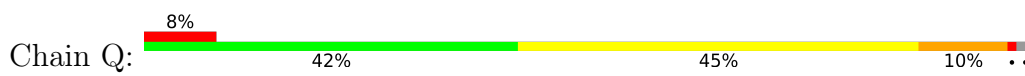




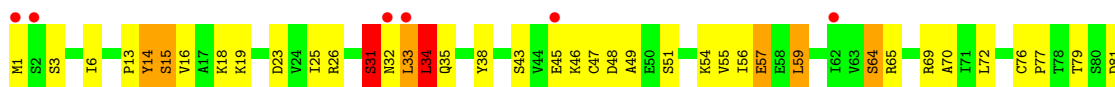
• Molecule 5: DNA-directed RNA polymerase, subunit E' (RpoE1)



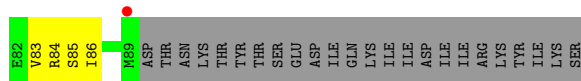
• Molecule 5: DNA-directed RNA polymerase, subunit E' (RpoE1)



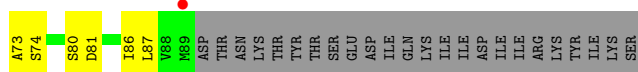
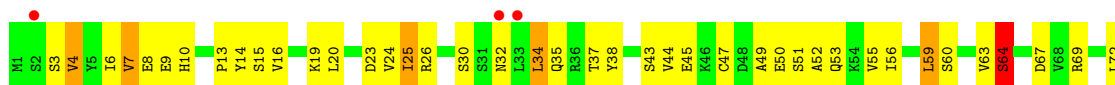
• Molecule 6: DNA-directed RNA polymerase, subunit F (RpoF)



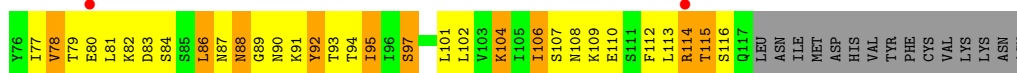




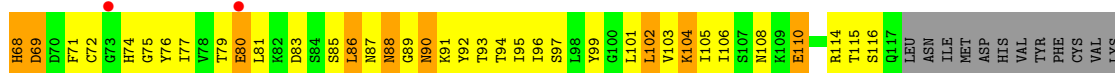
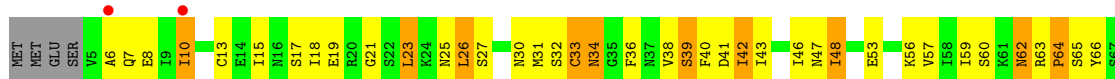
- Molecule 6: DNA-directed RNA polymerase, subunit F (RpoF)



- Molecule 7: DNA-directed RNA polymerase, subunit G (RpoG)



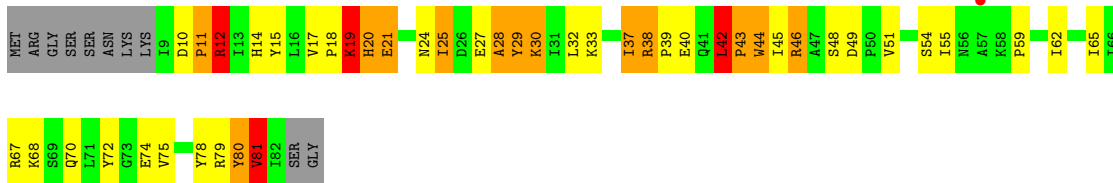
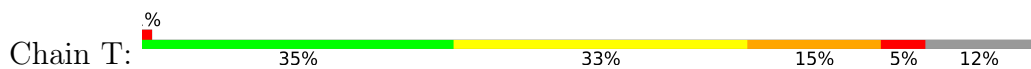
- Molecule 7: DNA-directed RNA polymerase, subunit G (RpoG)



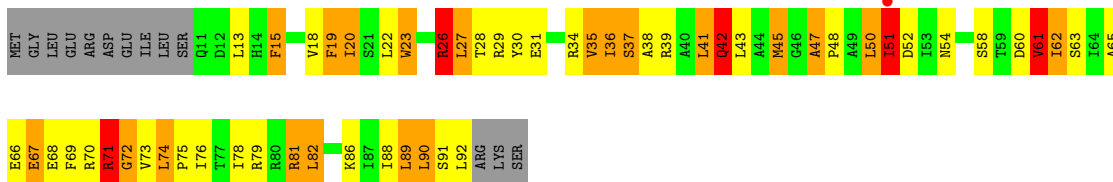
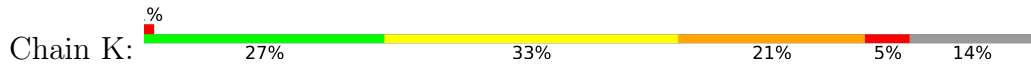
- Molecule 8: DNA-directed RNA polymerase subunit H



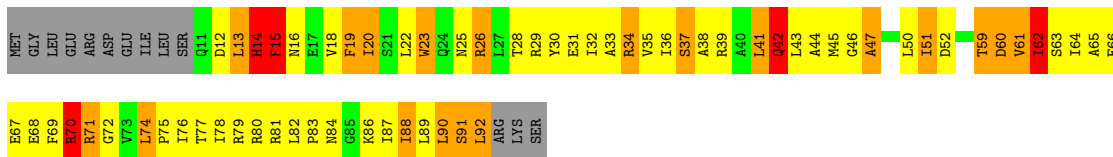
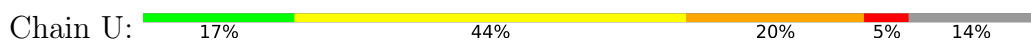
- Molecule 8: DNA-directed RNA polymerase subunit H



- Molecule 9: DNA-directed RNA polymerase subunit K



- Molecule 9: DNA-directed RNA polymerase subunit K



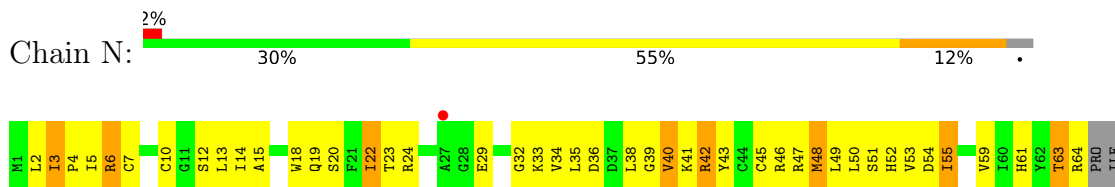
- Molecule 10: DNA-directed RNA polymerase subunit L



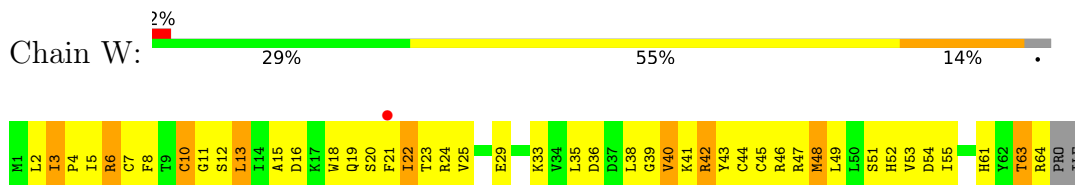
- Molecule 10: DNA-directed RNA polymerase subunit L



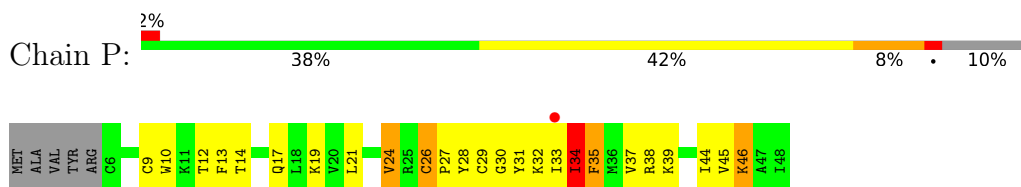
- Molecule 11: DNA-directed RNA polymerase subunit N



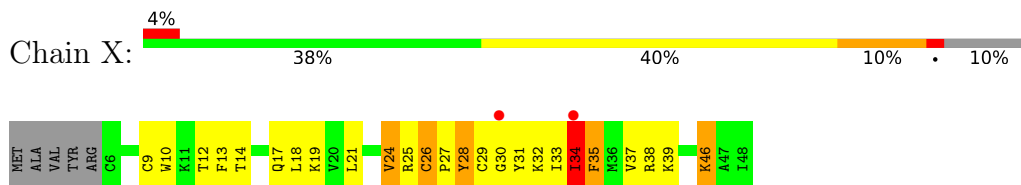
- Molecule 11: DNA-directed RNA polymerase subunit N



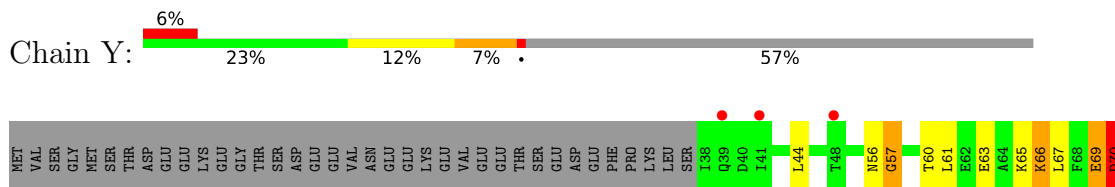
- Molecule 12: DNA-directed RNA polymerase subunit P



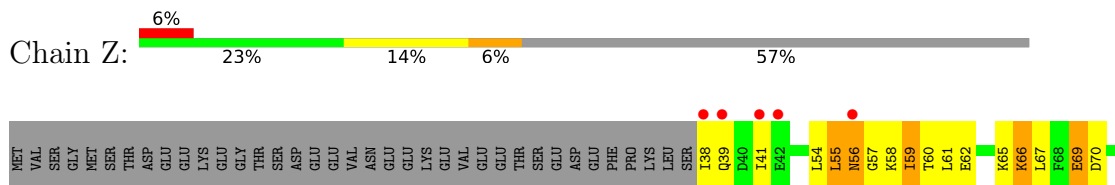
- Molecule 12: DNA-directed RNA polymerase subunit P

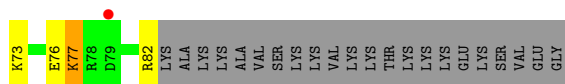


- Molecule 13: DNA-directed RNA polymerase subunit 13



- Molecule 13: DNA-directed RNA polymerase subunit 13





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	125.82Å 201.24Å 196.05Å 90.00° 100.92° 90.00°	Depositor
Resolution (Å)	40.00 – 3.40 39.79 – 3.40	Depositor EDS
% Data completeness (in resolution range)	80.3 (40.00-3.40) 80.3 (39.79-3.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.69 (at 3.40Å)	Xtrriage
Refinement program	REFMAC 5.5.0072	Depositor
R, $R_{free}$	0.265 , 0.341 0.264 , 0.337	Depositor DCC
$R_{free}$ test set	5323 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	76.2	Xtrriage
Anisotropy	0.062	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 71.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.41$ , $\langle L^2 \rangle = 0.24$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	53072	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	78.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.70% 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: F3S, MG, 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	0.60	0/6815	0.76	4/9219 (0.0%)
1	I	0.62	0/6815	0.77	5/9219 (0.1%)
2	C	0.64	0/2892	0.85	1/3891 (0.0%)
2	M	0.66	0/2892	0.84	2/3891 (0.1%)
3	B	0.60	0/8810	0.78	7/11921 (0.1%)
3	J	0.60	0/8810	0.79	6/11921 (0.1%)
4	D	0.49	0/2152	0.66	0/2911
4	O	0.52	0/2152	0.66	0/2911
5	E	0.53	0/1423	0.72	0/1919
5	Q	0.52	0/1423	0.72	0/1919
6	F	0.50	0/701	0.66	1/949 (0.1%)
6	R	0.52	0/701	0.70	0/949
7	G	0.63	0/895	0.78	0/1203
7	S	0.68	0/895	0.75	1/1203 (0.1%)
8	H	0.54	0/625	0.77	0/848
8	T	0.61	0/625	0.83	1/848 (0.1%)
9	K	0.61	0/667	0.87	0/903
9	U	0.70	0/667	0.95	1/903 (0.1%)
10	L	0.51	0/733	0.71	1/986 (0.1%)
10	V	0.57	0/733	0.74	0/986
11	N	0.51	0/523	0.74	0/705
11	W	0.52	0/523	0.74	0/705
12	P	0.63	0/354	0.73	0/475
12	X	0.65	0/354	0.70	0/475
13	Y	0.71	0/395	0.73	0/527
13	Z	0.67	0/395	0.75	0/527
All	All	0.60	0/53970	0.77	30/72914 (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	M	0	1
3	B	0	1
3	J	0	1
7	G	0	1
9	K	0	1
9	U	0	1
All	All	0	6

There are no bond length outliers.

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	464	LEU	CA-CB-CG	6.85	131.04	115.30
1	I	239	LEU	CA-CB-CG	6.55	130.36	115.30
8	T	42	LEU	CA-CB-CG	6.31	129.80	115.30
1	A	239	LEU	CA-CB-CG	6.30	129.79	115.30
1	I	464	LEU	CA-CB-CG	6.19	129.55	115.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	B	945	PHE	Peptide
7	G	40	PHE	Peptide
3	J	945	PHE	Peptide
9	K	26	ARG	Peptide
2	M	27	LYS	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	6673	0	6763	612	0
1	I	6673	0	6763	619	0
2	C	2868	0	3035	277	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	M	2868	0	3035	331	0
3	B	8645	0	8780	817	0
3	J	8645	0	8780	862	0
4	D	2114	0	2146	120	0
4	O	2114	0	2146	122	0
5	E	1402	0	1467	80	0
5	Q	1402	0	1467	97	0
6	F	694	0	705	37	0
6	R	694	0	705	29	0
7	G	884	0	888	91	0
7	S	884	0	888	86	0
8	H	611	0	641	49	0
8	T	611	0	641	51	0
9	K	658	0	692	81	0
9	U	658	0	692	92	0
10	L	723	0	749	54	0
10	V	723	0	749	57	0
11	N	514	0	528	66	0
11	W	514	0	528	69	0
12	P	346	0	375	28	0
12	X	346	0	375	28	0
13	Y	391	0	389	27	0
13	Z	391	0	389	13	0
14	A	2	0	0	0	0
14	B	1	0	0	2	0
14	I	2	0	0	0	0
14	J	1	0	0	2	0
14	N	1	0	0	0	0
14	P	1	0	0	0	0
14	W	1	0	0	0	0
14	X	1	0	0	0	0
15	A	1	0	0	0	0
15	I	1	0	0	0	0
16	D	7	0	0	1	0
16	O	7	0	0	6	0
All	All	53072	0	54316	4172	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 39.

The worst 5 of 4172 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)
8:H:29:TYR:OH	13:Y:67:LEU:HD22	1.31	1.25
4:D:183:CYS:SG	16:D:1001:F3S:S2	2.42	1.17
6:R:72:LEU:HD21	6:R:86:ILE:HD12	1.30	1.14
5:Q:147:ILE:HD11	5:Q:163:THR:HB	1.21	1.14
3:B:581:ILE:HD11	3:B:614:GLU:HB2	1.16	1.13

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	828/880 (94%)	597 (72%)	154 (19%)	77 (9%)	0	4
1	I	828/880 (94%)	594 (72%)	157 (19%)	77 (9%)	0	4
2	C	366/395 (93%)	225 (62%)	92 (25%)	49 (13%)	0	1
2	M	366/395 (93%)	232 (63%)	84 (23%)	50 (14%)	0	1
3	B	1084/1124 (96%)	752 (69%)	215 (20%)	117 (11%)	0	3
3	J	1084/1124 (96%)	752 (69%)	217 (20%)	115 (11%)	0	3
4	D	262/265 (99%)	199 (76%)	50 (19%)	13 (5%)	2	14
4	O	262/265 (99%)	198 (76%)	53 (20%)	11 (4%)	3	18
5	E	172/180 (96%)	125 (73%)	35 (20%)	12 (7%)	1	7
5	Q	172/180 (96%)	128 (74%)	34 (20%)	10 (6%)	1	11
6	F	87/113 (77%)	57 (66%)	25 (29%)	5 (6%)	1	12
6	R	87/113 (77%)	59 (68%)	19 (22%)	9 (10%)	0	3
7	G	111/132 (84%)	82 (74%)	22 (20%)	7 (6%)	1	9
7	S	111/132 (84%)	84 (76%)	16 (14%)	11 (10%)	0	4
8	H	72/84 (86%)	46 (64%)	12 (17%)	14 (19%)	0	0
8	T	72/84 (86%)	39 (54%)	20 (28%)	13 (18%)	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
9	K	80/95 (84%)	50 (62%)	21 (26%)	9 (11%)	0	3
9	U	80/95 (84%)	43 (54%)	18 (22%)	19 (24%)	0	0
10	L	90/92 (98%)	66 (73%)	16 (18%)	8 (9%)	1	5
10	V	90/92 (98%)	58 (64%)	28 (31%)	4 (4%)	2	16
11	N	62/66 (94%)	41 (66%)	17 (27%)	4 (6%)	1	8
11	W	62/66 (94%)	43 (69%)	15 (24%)	4 (6%)	1	8
12	P	41/48 (85%)	29 (71%)	8 (20%)	4 (10%)	0	4
12	X	41/48 (85%)	28 (68%)	8 (20%)	5 (12%)	0	2
13	Y	43/104 (41%)	26 (60%)	13 (30%)	4 (9%)	0	4
13	Z	43/104 (41%)	30 (70%)	11 (26%)	2 (5%)	2	15
All	All	6596/7156 (92%)	4583 (70%)	1360 (21%)	653 (10%)	0	4

5 of 653 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	47	PRO
1	A	56	GLN
1	A	58	CYS
1	A	64	THR
1	A	145	VAL

### 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	729/766 (95%)	601 (82%)	128 (18%)	2	6
1	I	729/766 (95%)	604 (83%)	125 (17%)	2	8
2	C	318/340 (94%)	241 (76%)	77 (24%)	0	2
2	M	318/340 (94%)	226 (71%)	92 (29%)	0	1
3	B	937/965 (97%)	772 (82%)	165 (18%)	2	6
3	J	937/965 (97%)	771 (82%)	166 (18%)	2	6

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	D	241/242 (100%)	220 (91%)	21 (9%)	10	34
4	O	241/242 (100%)	221 (92%)	20 (8%)	11	36
5	E	156/159 (98%)	128 (82%)	28 (18%)	2	6
5	Q	156/159 (98%)	123 (79%)	33 (21%)	1	3
6	F	82/106 (77%)	72 (88%)	10 (12%)	5	18
6	R	82/106 (77%)	72 (88%)	10 (12%)	5	18
7	G	105/125 (84%)	87 (83%)	18 (17%)	2	8
7	S	105/125 (84%)	82 (78%)	23 (22%)	1	3
8	H	67/75 (89%)	49 (73%)	18 (27%)	0	1
8	T	67/75 (89%)	54 (81%)	13 (19%)	1	4
9	K	72/84 (86%)	48 (67%)	24 (33%)	0	1
9	U	72/84 (86%)	50 (69%)	22 (31%)	0	1
10	L	81/81 (100%)	66 (82%)	15 (18%)	1	5
10	V	81/81 (100%)	63 (78%)	18 (22%)	1	2
11	N	58/60 (97%)	48 (83%)	10 (17%)	2	8
11	W	58/60 (97%)	48 (83%)	10 (17%)	2	8
12	P	39/43 (91%)	32 (82%)	7 (18%)	2	6
12	X	39/43 (91%)	32 (82%)	7 (18%)	2	6
13	Y	43/97 (44%)	34 (79%)	9 (21%)	1	3
13	Z	43/97 (44%)	34 (79%)	9 (21%)	1	3
All	All	5856/6286 (93%)	4778 (82%)	1078 (18%)	1	5

5 of 1078 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	O	170	VAL
5	Q	85	VAL
4	O	165	ARG
10	V	38	VAL
4	D	204	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 134 such sidechains are listed below:

Mol	Chain	Res	Type
3	J	913	HIS
5	Q	33	GLN
9	U	54	ASN
3	B	991	GLN
3	B	913	HIS

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 14 ligands modelled in this entry, 12 are monoatomic - leaving 2 for Mogul analysis.

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
16	F3S	O	1001	-	0,9,9	-	-	-		
16	F3S	D	1001	4	0,9,9	-	-	-		

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
16	F3S	O	1001	-	-	-	0/3/3/3
16	F3S	D	1001	4	-	-	0/3/3/3

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.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
16	O	1001	F3S	6	0
16	D	1001	F3S	1	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

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

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 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	836/880 (95%)	-0.16	10 (1%) 79 77	28, 71, 105, 133	0
1	I	836/880 (95%)	-0.17	15 (1%) 68 67	28, 71, 106, 133	0
2	C	370/395 (93%)	0.18	22 (5%) 22 23	37, 78, 130, 138	0
2	M	370/395 (93%)	0.07	16 (4%) 35 35	38, 79, 130, 138	0
3	B	1090/1124 (96%)	-0.14	19 (1%) 70 68	33, 66, 117, 149	0
3	J	1090/1124 (96%)	-0.13	24 (2%) 62 60	33, 67, 117, 149	0
4	D	264/265 (99%)	0.09	10 (3%) 40 39	57, 86, 121, 139	0
4	O	264/265 (99%)	0.19	16 (6%) 21 22	57, 86, 121, 139	0
5	E	176/180 (97%)	0.17	7 (3%) 38 37	58, 98, 140, 143	0
5	Q	176/180 (97%)	0.23	15 (8%) 10 12	58, 98, 139, 142	0
6	F	89/113 (78%)	0.23	7 (7%) 12 14	100, 127, 132, 133	0
6	R	89/113 (78%)	0.21	4 (4%) 33 33	100, 127, 132, 133	0
7	G	113/132 (85%)	0.17	4 (3%) 44 43	65, 88, 108, 112	0
7	S	113/132 (85%)	0.09	4 (3%) 44 43	66, 88, 108, 112	0
8	H	74/84 (88%)	-0.14	0 100 100	66, 81, 99, 102	0
8	T	74/84 (88%)	-0.19	1 (1%) 75 74	66, 81, 99, 102	0
9	K	82/95 (86%)	-0.40	1 (1%) 79 77	41, 57, 86, 94	0
9	U	82/95 (86%)	-0.33	0 100 100	43, 58, 87, 93	0
10	L	92/92 (100%)	-0.13	1 (1%) 80 79	53, 73, 103, 107	0
10	V	92/92 (100%)	-0.03	1 (1%) 80 79	54, 73, 103, 107	0
11	N	64/66 (96%)	-0.12	1 (1%) 72 70	62, 80, 103, 124	0
11	W	64/66 (96%)	-0.05	1 (1%) 72 70	62, 80, 103, 124	0
12	P	43/48 (89%)	0.07	1 (2%) 60 59	66, 89, 103, 110	0
12	X	43/48 (89%)	-0.11	2 (4%) 31 31	66, 89, 103, 110	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
13	Y	45/104 (43%)	0.61	6 (13%) 3 4	98, 108, 124, 125	0
13	Z	45/104 (43%)	0.77	6 (13%) 3 4	98, 108, 124, 124	0
All	All	6676/7156 (93%)	-0.05	194 (2%) 51 50	28, 76, 127, 149	0

The worst 5 of 194 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	J	278	ILE	7.2
6	R	89	MET	5.5
2	C	213	ILE	5.0
5	E	135	VAL	4.9
1	A	60	THR	4.7

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 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
14	ZN	I	1002	1/1	0.96	0.06	78,78,78,78	0
14	ZN	A	1002	1/1	0.98	0.06	69,69,69,69	0
14	ZN	B	2001	1/1	0.98	0.03	66,66,66,66	0
14	ZN	P	1001	1/1	0.98	0.04	33,33,33,33	0
14	ZN	A	1001	1/1	0.98	0.04	46,46,46,46	0
14	ZN	X	1001	1/1	0.98	0.08	54,54,54,54	0
15	MG	A	1003	1/1	0.98	0.04	2,2,2,2	0
15	MG	I	1003	1/1	0.98	0.03	2,2,2,2	0
14	ZN	W	1001	1/1	0.99	0.10	63,63,63,63	0
14	ZN	I	1001	1/1	0.99	0.12	57,57,57,57	0

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
14	ZN	N	1001	1/1	0.99	0.08	56,56,56,56	0
14	ZN	J	2001	1/1	0.99	0.03	124,124,124,124	0
16	F3S	D	1001	7/7	0.99	0.11	79,81,82,84	0
16	F3S	O	1001	7/7	0.99	0.07	104,105,106,106	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.