



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 09:06 am BST

PDB ID : 5X4Z  
Title : RNA Polymerase II from Komagataella Pastoris (Type-1 crystal)  
Authors : Ehara, H.; Umehara, T.; Sekine, S.; Yokoyama, S.  
Deposited on : 2017-02-14  
Resolution : 7.80 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 following versions of software and data (see [references ⓘ](#)) 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

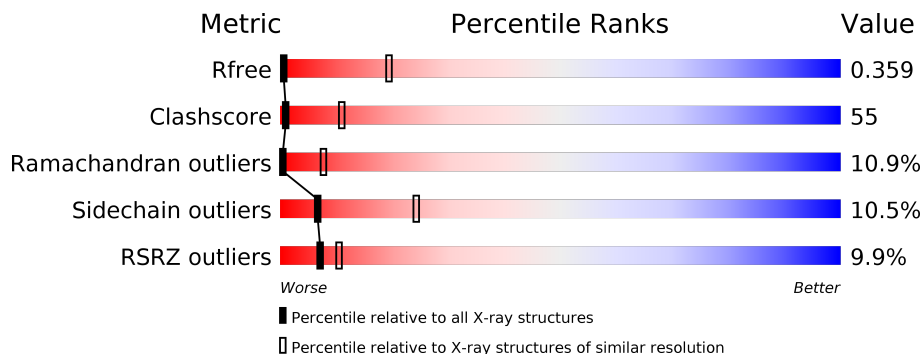
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 7.80 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	1005 (11.50-3.90)
Clashscore	141614	1070 (11.50-3.90)
Ramachandran outliers	138981	1003 (11.50-3.90)
Sidechain outliers	138945	1003 (11.50-3.86)
RSRZ outliers	127900	1004 (9.50-3.80)

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  $\geq 3$ , 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	1743	
1	M	1743	
2	B	1227	
2	N	1227	
3	C	304	
3	O	304	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	186	
4	P	186	
5	E	214	
5	Q	214	
6	F	155	
6	R	155	
7	G	171	
7	S	171	
8	H	145	
8	T	145	
9	I	115	
9	U	115	
10	J	72	
10	V	72	
11	K	118	
11	W	118	
12	L	73	
12	X	73	

## 2 Entry composition

There are 13 unique types of molecules in this entry. The entry contains 56628 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1384	Total	C	N	O	S	0	0	0
			10247	6499	1792	1902	54			
1	M	1387	Total	C	N	O	S	0	0	0
			10269	6512	1796	1907	54			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1070	Total	C	N	O	S	0	0	0
			8153	5184	1417	1506	46			
2	N	1074	Total	C	N	O	S	0	0	0
			8190	5208	1426	1510	46			

- Molecule 3 is a protein called RNA polymerase II third largest subunit B44, part of central core.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	265	Total	C	N	O	S	0	0	0
			1863	1188	310	357	8			
3	O	265	Total	C	N	O	S	0	0	0
			1863	1188	310	357	8			

- Molecule 4 is a protein called RNA polymerase II subunit B32.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	159	Total	C	N	O	S	0	0	0
			1082	695	185	200	2			
4	P	164	Total	C	N	O	S	0	0	0
			1113	712	191	208	2			

- Molecule 5 is a protein called RNA polymerase subunit ABC27, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	214	Total	C	N	O	S	0	0	0
			1638	1045	288	297	8			
5	Q	214	Total	C	N	O	S	0	0	0
			1638	1045	288	297	8			

- Molecule 6 is a protein called RNA polymerase subunit ABC23, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	84	Total	C	N	O	S	0	0	0
			637	406	109	119	3			
6	R	84	Total	C	N	O	S	0	0	0
			637	406	109	119	3			

- Molecule 7 is a protein called RNA polymerase II subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	171	Total	C	N	O	S	0	0	0
			1187	775	192	217	3			
7	S	171	Total	C	N	O	S	0	0	0
			1187	775	192	217	3			

- Molecule 8 is a protein called RNA polymerase subunit ABC14.5, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	130	Total	C	N	O	S	0	0	0
			959	613	154	189	3			
8	T	131	Total	C	N	O	S	0	0	0
			965	616	155	191	3			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	I	113	Total	C	N	O	S	0	0	0
			854	535	150	158	11			
9	U	113	Total	C	N	O	S	0	0	0
			854	535	150	158	11			

- Molecule 10 is a protein called RNA polymerase subunit ABC10-beta, common to RNA polymerases I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	J	62	Total	C	N	O	S	0	0	0
			487	320	85	76	6			
10	V	62	Total	C	N	O	S	0	0	0
			487	320	85	76	6			

- Molecule 11 is a protein called RNA polymerase II subunit B12.5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	114	Total	C	N	O	S	0	0	0
			832	543	139	148	2			
11	W	114	Total	C	N	O	S	0	0	0
			832	543	139	148	2			

- Molecule 12 is a protein called RNA polymerase subunit, found in RNA polymerase complexes I, II, and III.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	L	46	Total	C	N	O	S	0	0	0
			319	196	64	55	4			
12	X	46	Total	C	N	O	S	0	0	0
			319	196	64	55	4			

- Molecule 13 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
13	J	1	Total	Zn	0	0
			1	1		
13	B	1	Total	Zn	0	0
			1	1		
13	I	2	Total	Zn	0	0
			2	2		
13	C	1	Total	Zn	0	0
			1	1		
13	V	1	Total	Zn	0	0
			1	1		
13	A	2	Total	Zn	0	0
			2	2		
13	N	1	Total	Zn	0	0
			1	1		
13	U	2	Total	Zn	0	0
			2	2		
13	X	1	Total	Zn	0	0
			1	1		

*Continued on next page...*

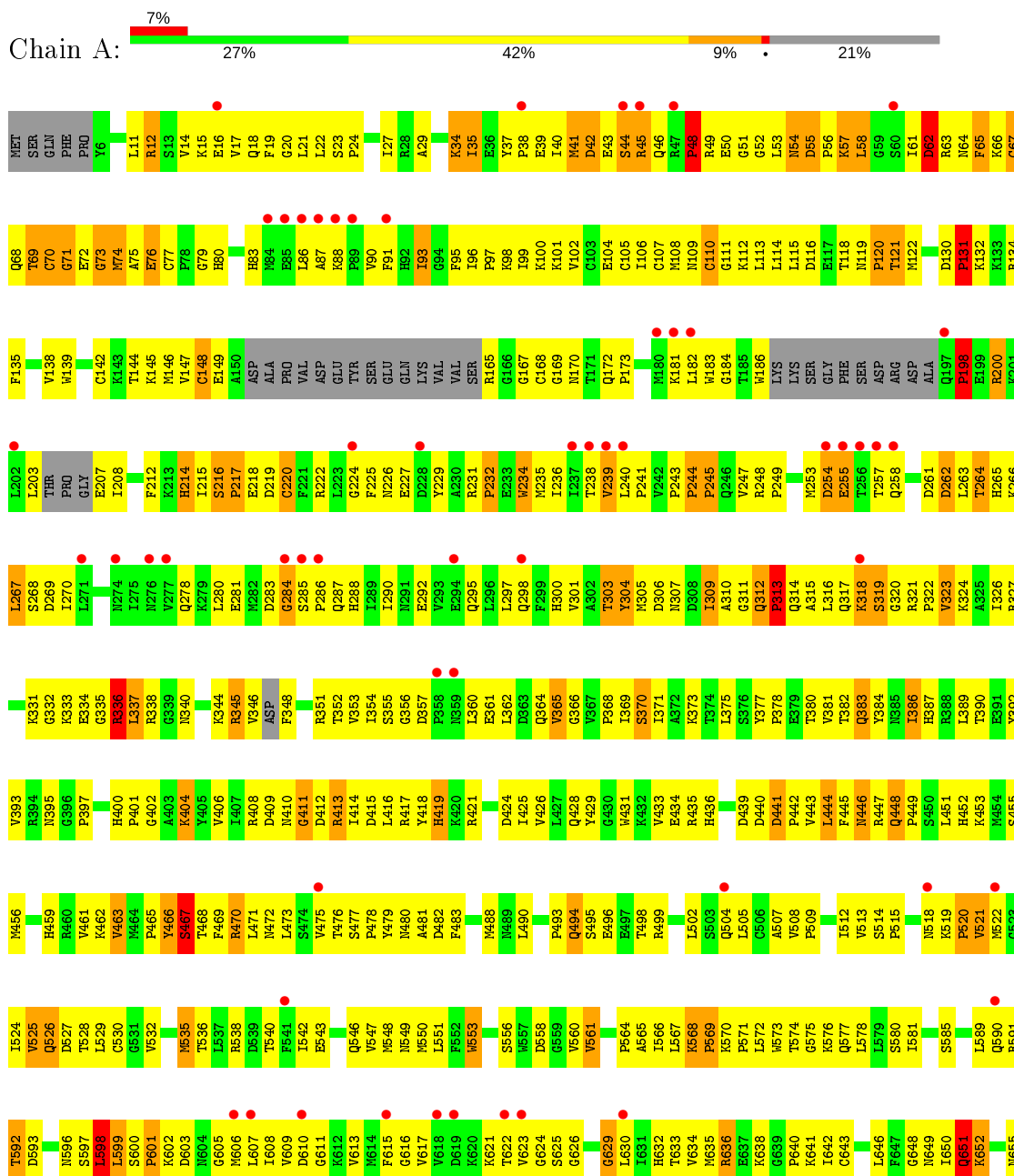
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
13	O	1	Total 1	Zn 1	0	0
13	L	1	Total 1	Zn 1	0	0
13	M	2	Total 2	Zn 2	0	0

### 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 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





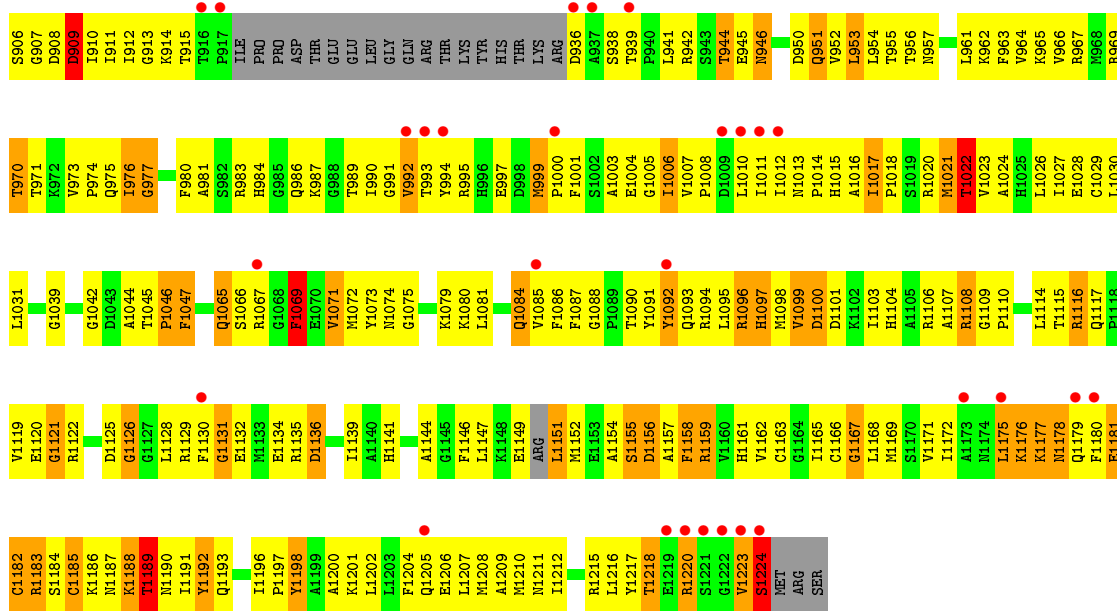




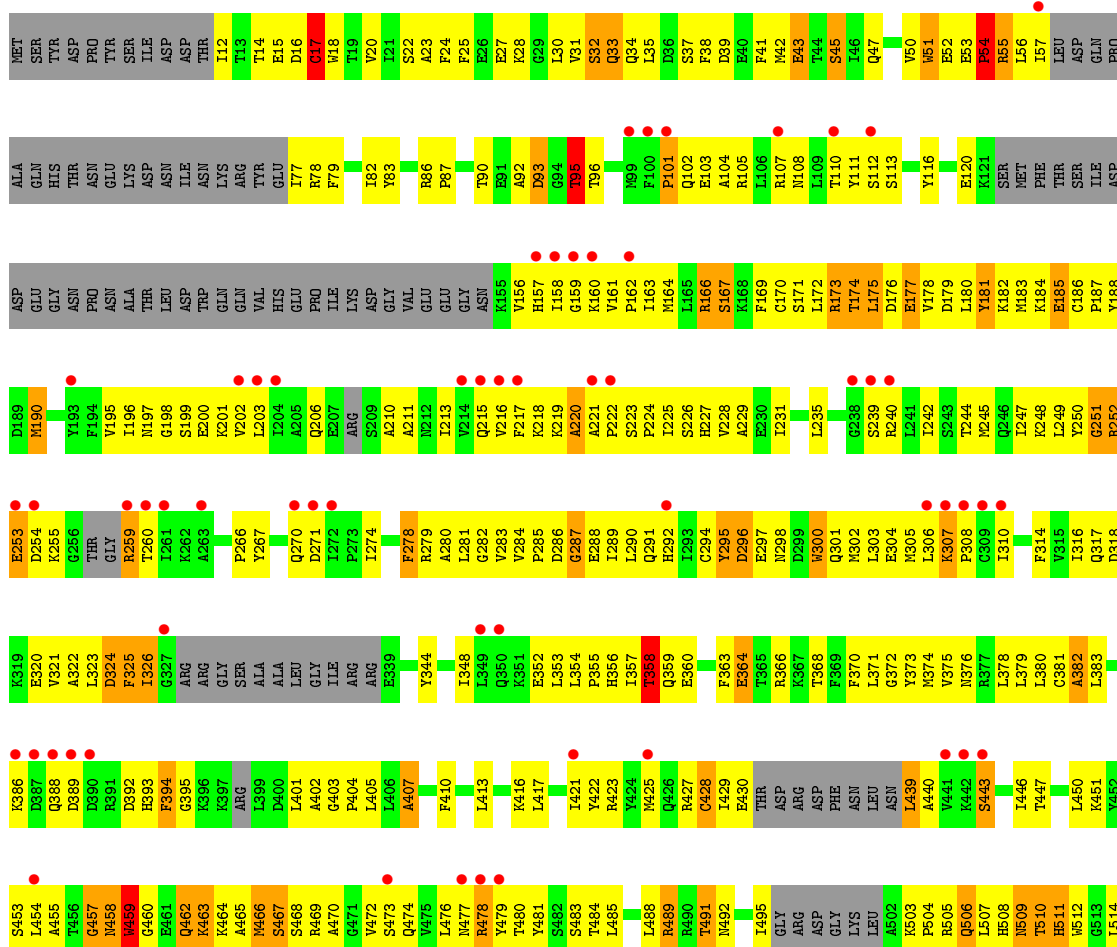




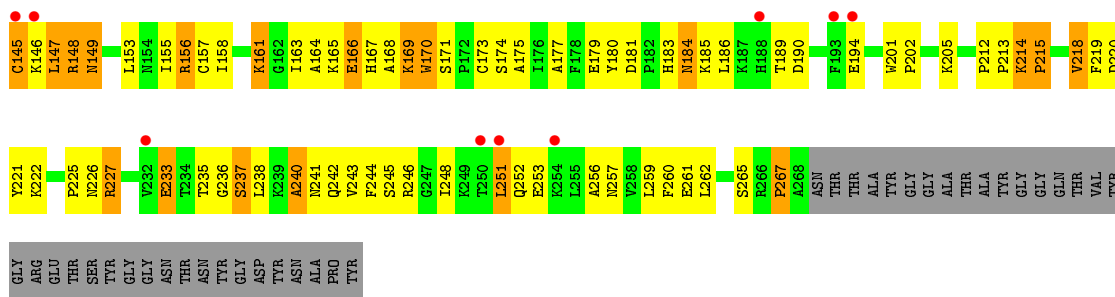
MET	ALA	ASP	D254	A322	D389	T456	P517	L589	H651	L8U	G774	M839
SER	GLN	GLU	K255	L323	D392	G457	A518	V590	I652	GLU	K715	I840
TYR	HIS	GLY	K256	D394	D393	G458	E519	S591	R653	GLN	Q776	I841
ASP	THR	ASN	THR	F325	H393	V459	N458	T592	R654	LYS	A777	M842
PRO	ASN	PRO	GLY	I326	F394	G460	G461	M593	I655	GLU	Q843	M843
THR	ASN	ASN	R259	R327	G395	E461	Q524	R594	L656	ILE	M778	S844
SER	LYS	ALA	T260	ARG	K396	Q462	Q525	D595	G657	ASP	F781	S845
ILE	ASP	THR	S199	ARG	LYS	K463	L528	L596	GLY	LEU	L782	I846
ASP	ASP	LEU	E200	GLY	ARG	K464	V529	R597	TYR	ASN	L783	D847
ASP	E200	LEU	P266	THR	L399	K465	R598	R598	ASP	THR	L784	R848
ASP	ASP	ASP	Y267	SER	D400	A465	K530	R598	ASP	ASP	M785	R849
THR	ASN	TRP	K201	ALA	M466	M466	M331	S599	ASP	ASP	I722	G849
THR	ASN	TRP	V202	ALA	L401	S467	L532	G600	ASP	ASP	A723	L850
LYS	GLN	GLN	L203	LEU	A402	S468	L533	A601	ALA	ALA	K724	F851
I12	GLN	GLN	L204	LEU	A403	S469	S536	I602	MET	MET	R725	V787
I13	VAL	VAL	A205	ILE	P404	R469	G537	I602	ASN	ASN	R726	R788
I14	HIS	HIS	Q206	ARG	L405	A470	G538	S603	ASP	ASP	I727	M789
E15	GLU	GLU	E207	ARG	A406	G471	I539	P604	ASP	ASP	K727	L790
D16	LEU	PRO	E207	ARG	A407	S473	I540	V606	SER	SER	P728	F855
G17	PRO	PRO	S209	LYS	L407	Q474	G541	S607	GLU	GLU	GLU	F856
M18	LYS	ILE	S209	LYS	F410	Y475	S542	I608	GLU	SER	MET	M792
M19	LYS	ILE	A210	LYS	F411	L476	S543	I608	GLU	THR	THR	L796
V20	ASP	ASP	A211	ASP	L413	N477	S544	R610	GLN	THR	THR	Y797
I21	GLY	GLY	M212	VAL	L414	R478	E545	D611	GLN	SER	SER	Y798
A23	GLU	GLU	V214	GLU	K416	Y479	E546	R612	ASP	HIS	HIS	Q800
F24	GLY	GLY	Q215	GLU	L417	T480	I547	R613	ASP	HIS	HIS	K801
F25	GLY	GLY	F217	ASN	L418	Y481	I548	E614	ASP	HIS	HIS	K802
E26	ASN	ASN	F217	ASN	L419	S482	I549	R615	ASP	HIS	HIS	L803
E27	ASN	ASN	K355	ASN	I421	S483	I550	R616	ASP	HIS	HIS	L804
K28	ASN	ASN	K218	ASN	Y422	G484	F551	B616	ASP	HIS	HIS	A804
K29	ASN	ASN	K219	ASN	Y423	T485	F552	K617	ASP	HIS	HIS	G805
G30	ASN	ASN	A221	ASN	Y424	L485	F553	F618	ASP	HIS	HIS	R805
L30	ASN	ASN	G94	ASN	M425	L486	F554	R619	ASP	HIS	HIS	T806
G31	ASN	ASN	T355	ASN	Q426	L487	F555	P620	ASP	HIS	HIS	A807
V31	ASN	ASN	T96	ASN	R427	L488	F556	P621	ASP	HIS	HIS	M809
S32	ASN	ASN	T96	ASN	N428	R489	F557	D622	ASP	HIS	HIS	E810
S33	ASN	ASN	T96	ASN	E430	R490	F558	L626	ASP	HIS	HIS	E811
Q34	ASN	ASN	P101	ASN	E431	R491	F559	V626	ASP	HIS	HIS	Y811
I35	ASN	ASN	Q102	ASN	E432	N492	F560	V627	ASP	HIS	HIS	L812
D36	ASN	ASN	I103	ASN	E433	N493	F561	R628	ASP	HIS	HIS	K813
D37	ASN	ASN	A104	ASN	THR	P494	F562	P629	ASP	HIS	HIS	F814
F38	ASN	ASN	R105	ASN	ASP	I495	Y562	R628	ASP	HIS	HIS	R815
D39	ASN	ASN	R106	ASN	ASP	L496	Y563	P629	ASP	HIS	HIS	E816
E40	ASN	ASN	L107	ASN	ASP	G502	Q566	L630	ASP	HIS	HIS	R816
F41	ASN	ASN	M108	ASN	PHE	ARG	Q567	F631	ASP	HIS	HIS	L817
M42	ASN	ASN	M109	ASN	ASP	GLY	T568	I632	ASP	HIS	HIS	I755
E43	ASN	ASN	T110	ASN	LEU	LEU	T569	V633	ASP	HIS	HIS	I756
T44	ASN	ASN	L112	ASN	LEU	LEU	T570	R634	ASP	HIS	HIS	I757
S45	ASN	ASN	S112	ASN	LEU	LEU	R572	D635	ASP	HIS	HIS	F758
I46	ASN	ASN	S113	ASN	LEU	LEU	I573	D636	ASP	HIS	HIS	F759
Q47	ASN	ASN	L116	ASN	LEU	LEU	F574	D637	ASP	HIS	HIS	P759
V50	ASN	ASN	L117	ASN	LEU	LEU	R575	D638	ASP	HIS	HIS	D760
M51	ASN	ASN	D118	ASN	LEU	LEU	V576	S638	ASP	HIS	HIS	H761
E52	ASN	ASN	M119	ASN	LEU	LEU	M576	K639	ASP	HIS	HIS	N762
E53	ASN	ASN	E120	ASN	LEU	LEU	M577	D640	ASP	HIS	HIS	A701
P54	ASN	ASN	M245	ASN	LEU	LEU	M578	M641	ASP	HIS	HIS	S764
R55	ASN	ASN	Y181	ASN	LEU	LEU	T580	L642	ASP	HIS	HIS	C829
L56	ASN	ASN	K182	ASN	LEU	LEU	G581	GLY	ASP	HIS	HIS	C829
L57	ASN	ASN	M183	ASN	LEU	LEU	I582	E644	ASP	HIS	HIS	P704
LEU	ASN	ASN	K184	ASN	LEU	LEU	I583	E645	ASP	HIS	HIS	R705
ASP	ASN	ASN	E185	ASN	LEU	LEU	H511	L645	ASP	HIS	HIS	E706
GLN	ASN	ASN	C186	ASN	LEU	LEU	H512	R646	ASP	HIS	HIS	D706
ASP	ASN	ASN	Y187	ASN	LEU	LEU	G513	I647	ASP	HIS	HIS	L707
ASP	ASN	ASN	Y188	ASN	LEU	LEU	L514	T648	ASP	HIS	HIS	GLN
ASP	ASN	ASN	D189	ASN	LEU	LEU	V515	R649	ASP	HIS	HIS	THR
							M588	E650				ARG
												SER



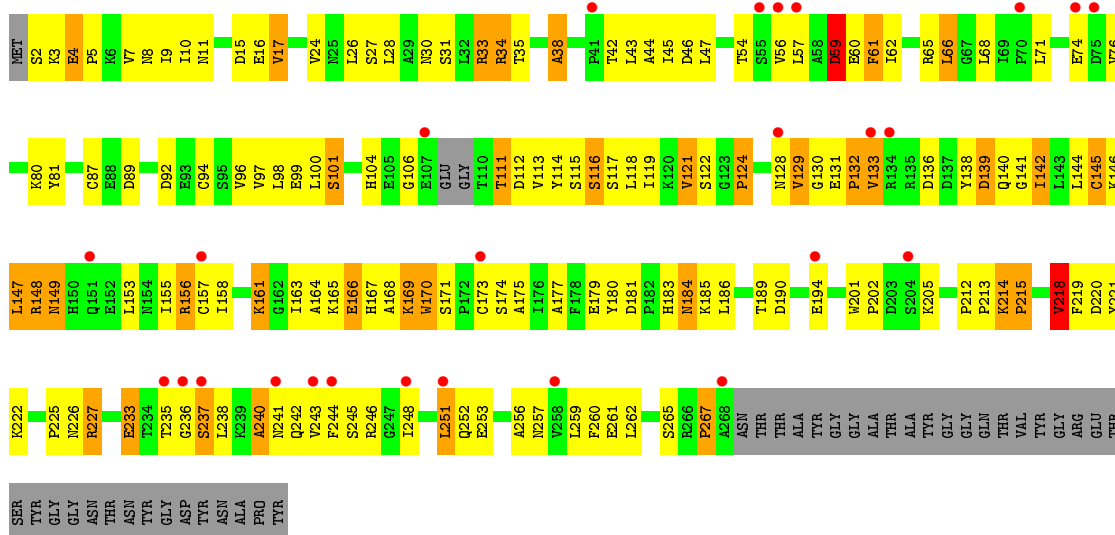
• Molecule 2: DNA-directed RNA polymerase subunit beta



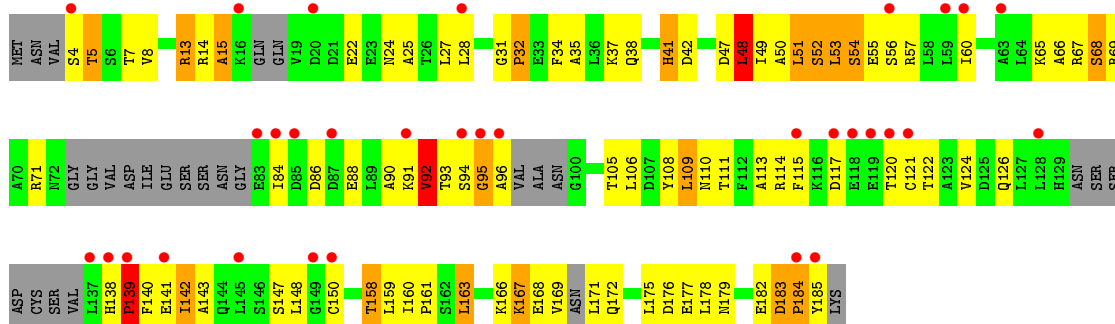




• Molecule 3: RNA polymerase II third largest subunit B44, part of central core

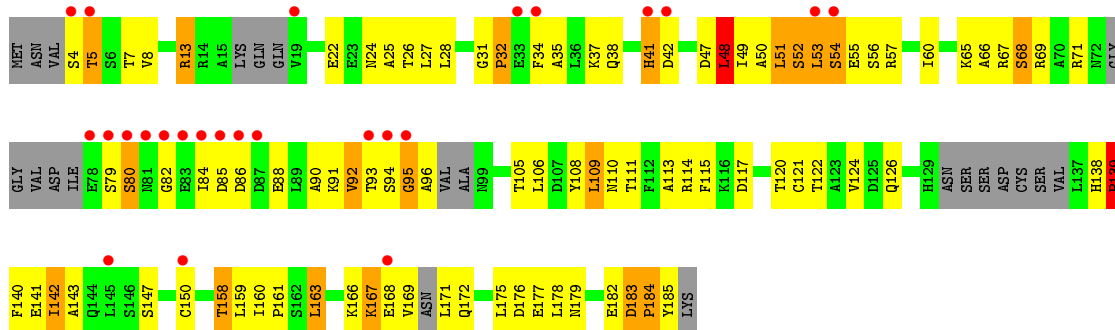


• Molecule 4: RNA polymerase II subunit B32

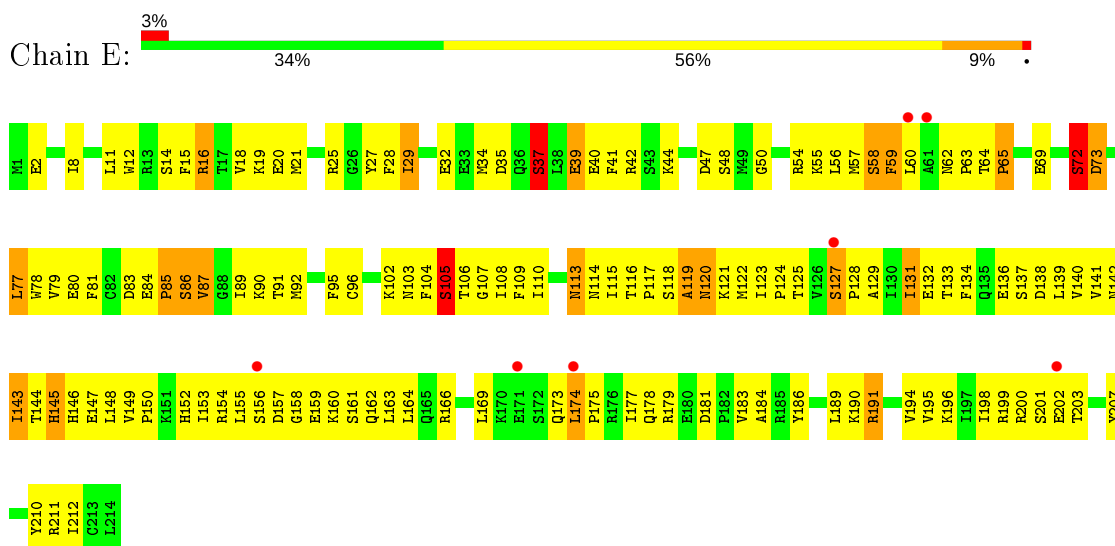


• Molecule 4: RNA polymerase II subunit B32

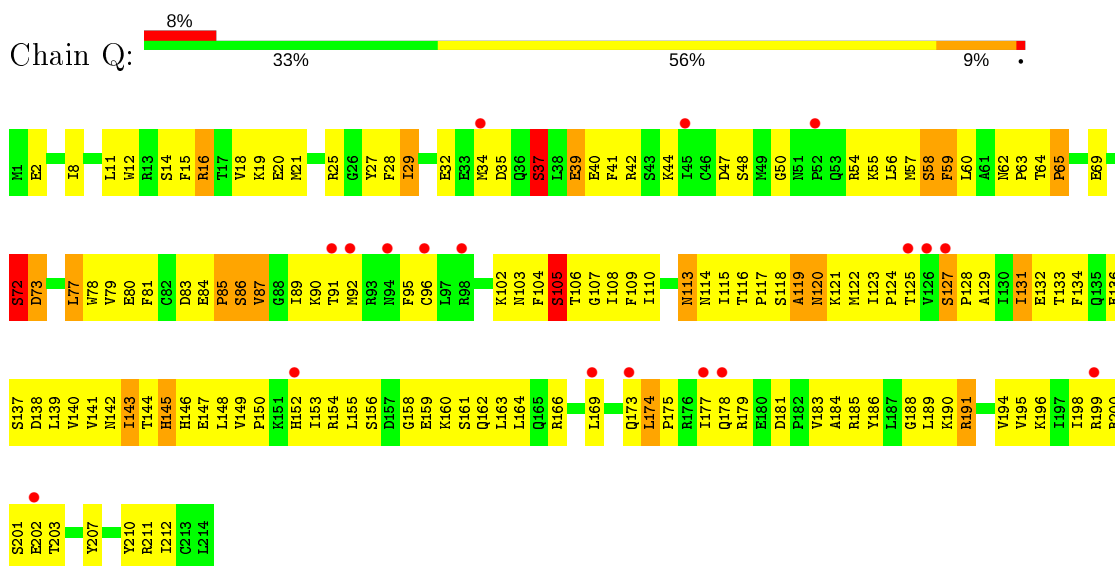




- Molecule 5: RNA polymerase subunit ABC27, common to RNA polymerases I, II, and III



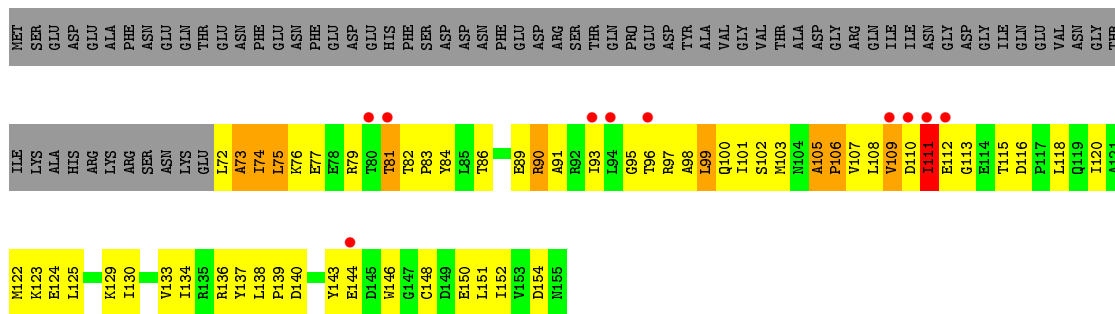
- Molecule 5: RNA polymerase subunit ABC27, common to RNA polymerases I, II, and III



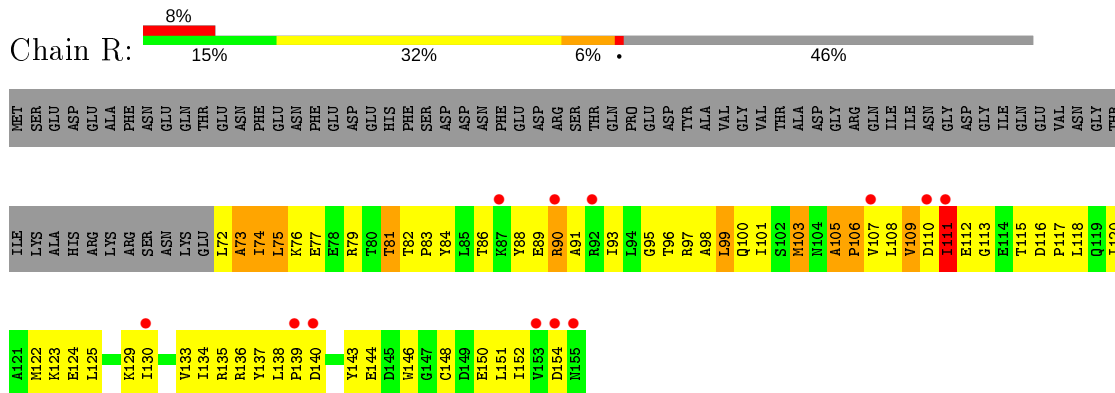
- Molecule 6: RNA polymerase subunit ABC23, common to RNA polymerases I, II, and III



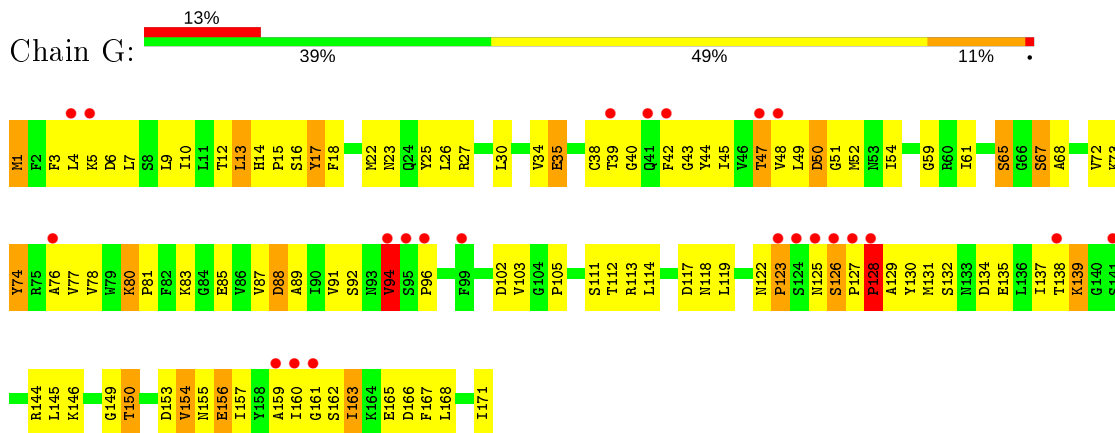




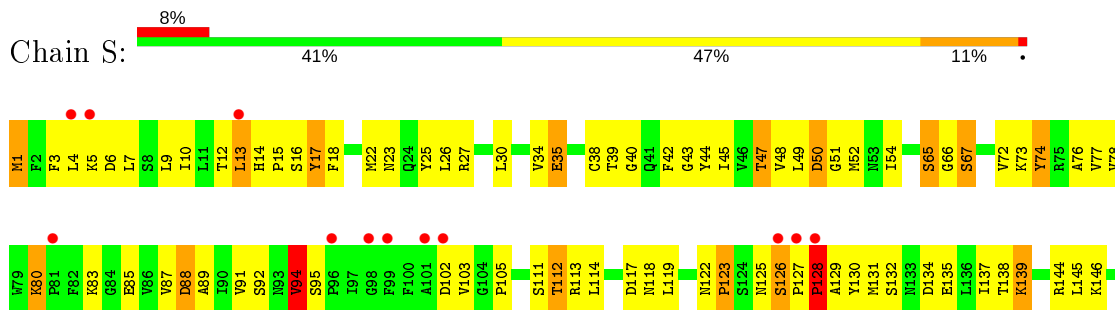
• Molecule 6: RNA polymerase subunit ABC23, common to RNA polymerases I, II, and III

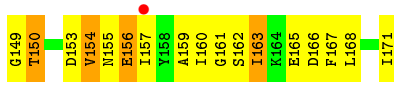


• Molecule 7: RNA polymerase II subunit

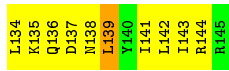
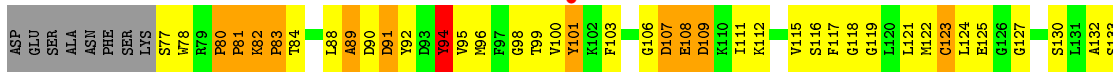
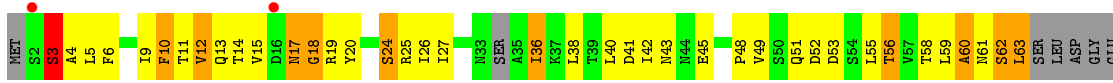


• Molecule 7: RNA polymerase II subunit

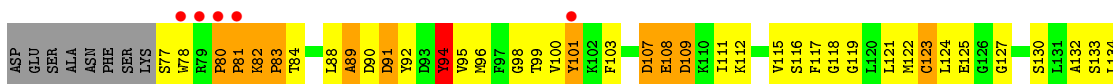
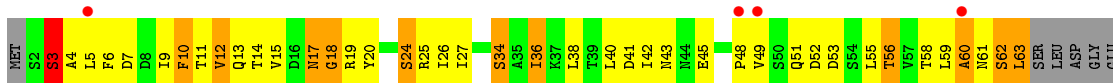




- Molecule 8: RNA polymerase subunit ABC14.5, common to RNA polymerases I, II, and III

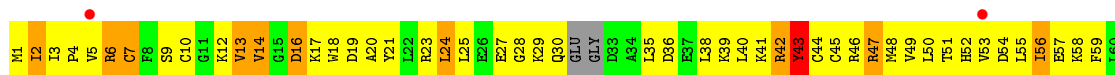
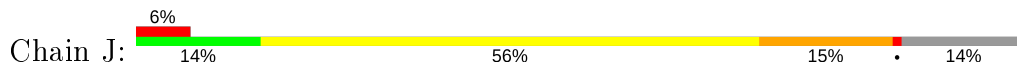


- Molecule 8: RNA polymerase subunit ABC14.5, common to RNA polymerases I, II, and III

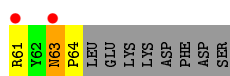
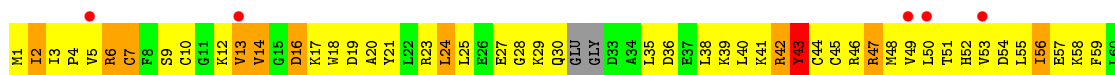
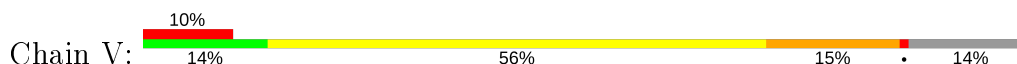




- Molecule 10: RNA polymerase subunit ABC10-beta, common to RNA polymerases I, II, and III



- Molecule 10: RNA polymerase subunit ABC10-beta, common to RNA polymerases I, II, and III



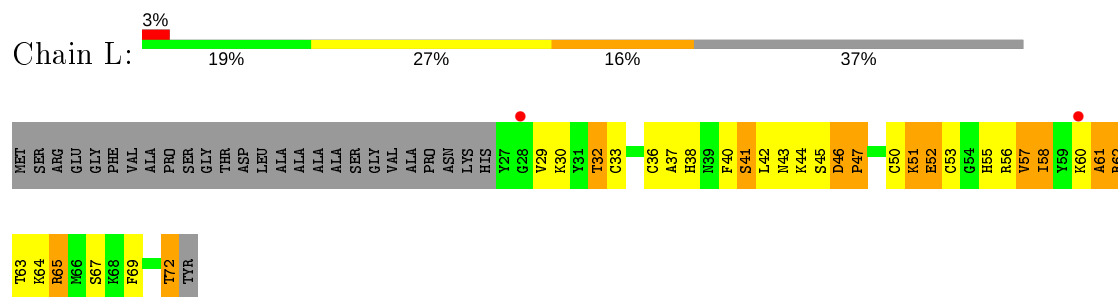
- Molecule 11: RNA polymerase II subunit B12.5



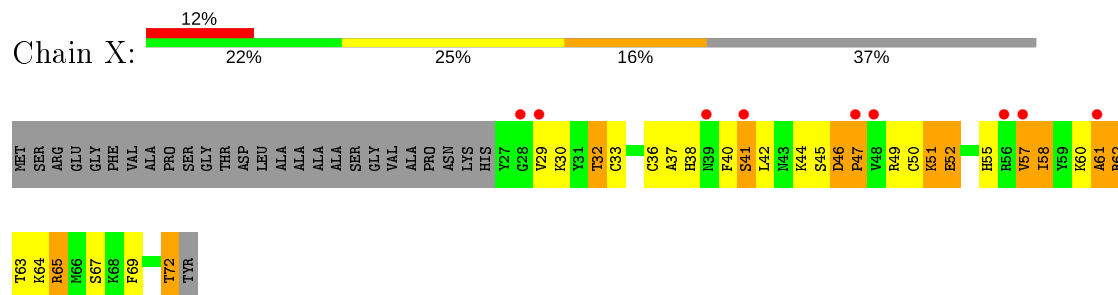
- Molecule 11: RNA polymerase II subunit B12.5



- Molecule 12: RNA polymerase subunit, found in RNA polymerase complexes I, II, and III



- Molecule 12: RNA polymerase subunit, found in RNA polymerase complexes I, II, and III



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	155.05Å 160.35Å 254.35Å 90.00° 105.43° 90.00°	Depositor
Resolution (Å)	49.82 – 7.80 49.82 – 7.73	Depositor EDS
% Data completeness (in resolution range)	99.7 (49.82-7.80) 80.4 (49.82-7.73)	Depositor EDS
$R_{merge}$	0.38	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.13 (at 7.37Å)	Xtrriage
Refinement program	PHENIX dev_2614	Depositor
R, $R_{free}$	0.353 , 0.357 0.354 , 0.359	Depositor DCC
$R_{free}$ test set	1389 reflections (9.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	174.0	Xtrriage
Anisotropy	0.345	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 329.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.41$ , $\langle L^2 \rangle = 0.24$	Xtrriage
Estimated twinning fraction	0.136 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.58	EDS
Total number of atoms	56628	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	186.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.35% 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:  
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	0.68	28/10399 (0.3%)	0.81	17/14038 (0.1%)
1	M	0.68	27/10421 (0.3%)	0.81	19/14067 (0.1%)
2	B	0.69	15/8294 (0.2%)	0.77	4/11176 (0.0%)
2	N	0.69	15/8331 (0.2%)	0.77	4/11223 (0.0%)
3	C	0.70	6/1888 (0.3%)	0.83	8/2558 (0.3%)
3	O	0.70	6/1888 (0.3%)	0.83	8/2558 (0.3%)
4	D	0.84	5/1086 (0.5%)	0.76	4/1460 (0.3%)
4	P	0.92	7/1117 (0.6%)	0.76	4/1501 (0.3%)
5	E	0.73	7/1668 (0.4%)	0.69	4/2245 (0.2%)
5	Q	0.72	7/1668 (0.4%)	0.70	4/2245 (0.2%)
6	F	0.84	1/646 (0.2%)	0.82	2/873 (0.2%)
6	R	0.84	1/646 (0.2%)	0.82	2/873 (0.2%)
7	G	1.07	7/1207 (0.6%)	0.80	1/1629 (0.1%)
7	S	1.08	8/1207 (0.7%)	0.80	1/1629 (0.1%)
8	H	1.37	7/973 (0.7%)	0.75	2/1313 (0.2%)
8	T	1.38	8/980 (0.8%)	0.74	2/1324 (0.2%)
9	I	0.92	4/868 (0.5%)	0.70	0/1169
9	U	0.92	4/868 (0.5%)	0.70	0/1169
10	J	0.55	0/495	0.79	0/664
10	V	0.55	0/495	0.80	0/664
11	K	0.65	3/848 (0.4%)	0.85	5/1147 (0.4%)
11	W	0.65	2/848 (0.2%)	0.85	5/1147 (0.4%)
12	L	0.83	3/321 (0.9%)	0.96	2/425 (0.5%)
12	X	0.83	3/321 (0.9%)	0.96	2/425 (0.5%)
All	All	0.76	174/57483 (0.3%)	0.79	100/77522 (0.1%)

The worst 5 of 174 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	H	24	SER	CB-OG	30.78	1.82	1.42
8	T	24	SER	CB-OG	30.77	1.82	1.42

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	N	17	CYS	CB-SG	-23.57	1.42	1.82
2	B	17	CYS	CB-SG	-23.55	1.42	1.82
2	N	428	CYS	CB-SG	-22.62	1.43	1.82

The worst 5 of 100 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
12	L	29	VAL	CA-CB-CG1	9.64	125.37	110.90
12	X	29	VAL	CA-CB-CG1	9.61	125.32	110.90
2	N	682	VAL	CA-CB-CG1	7.32	121.89	110.90
2	B	682	VAL	CA-CB-CG2	7.30	121.85	110.90
11	K	54	PRO	N-CA-CB	6.99	111.69	103.30

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts

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	10247	0	9725	1250	8
1	M	10269	0	9748	1200	0
2	B	8153	0	7831	1005	1
2	N	8190	0	7881	978	1
3	C	1863	0	1645	196	0
3	O	1863	0	1645	193	8
4	D	1082	0	945	110	1
4	P	1113	0	970	102	1
5	E	1638	0	1551	160	0
5	Q	1638	0	1551	163	0
6	F	637	0	620	59	1
6	R	637	0	620	78	1
7	G	1187	0	1092	161	2
7	S	1187	0	1092	148	1
8	H	959	0	858	127	2
8	T	965	0	864	123	2
9	I	854	0	763	84	0

Continued on next page...

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
9	U	854	0	763	82	1
10	J	487	0	492	116	0
10	V	487	0	492	115	0
11	K	832	0	727	89	0
11	W	832	0	727	90	0
12	L	319	0	287	30	1
12	X	319	0	287	30	1
13	A	2	0	0	0	0
13	B	1	0	0	1	0
13	C	1	0	0	0	0
13	I	2	0	0	0	0
13	J	1	0	0	0	0
13	L	1	0	0	0	0
13	M	2	0	0	0	0
13	N	1	0	0	0	0
13	O	1	0	0	0	0
13	U	2	0	0	0	0
13	V	1	0	0	0	0
13	X	1	0	0	0	0
All	All	56628	0	53176	6068	16

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 55.

The worst 5 of 6068 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)
2:B:51:TRP:CB	2:B:51:TRP:CG	1.81	1.63
1:A:521:VAL:CG1	1:A:521:VAL:CB	1.76	1.61
5:Q:123:ILE:CG1	5:Q:124:PRO:HD3	1.31	1.60
2:N:51:TRP:CG	2:N:51:TRP:CB	1.81	1.60
1:M:521:VAL:CB	1:M:521:VAL:CG1	1.76	1.58

The worst 5 of 16 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1178:ILE:CB	3:O:3:LYS:CB[2_555]	1.48	0.72
1:A:1178:ILE:CA	3:O:3:LYS:CB[2_555]	1.51	0.69
1:A:1178:ILE:O	3:O:3:LYS:CG[2_555]	1.58	0.62

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:869:SER:O	6:F:129:LYS:NZ[2_646]	1.65	0.55
1:A:1178:ILE:C	3:O:3:LYS:CG[2_555]	1.66	0.54

## 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	1358/1743 (78%)	945 (70%)	279 (20%)	134 (10%)	0	9
1	M	1361/1743 (78%)	948 (70%)	277 (20%)	136 (10%)	0	9
2	B	1040/1227 (85%)	708 (68%)	210 (20%)	122 (12%)	0	6
2	N	1044/1227 (85%)	712 (68%)	211 (20%)	121 (12%)	0	6
3	C	261/304 (86%)	182 (70%)	56 (22%)	23 (9%)	1	11
3	O	261/304 (86%)	182 (70%)	56 (22%)	23 (9%)	1	11
4	D	147/186 (79%)	101 (69%)	27 (18%)	19 (13%)	0	5
4	P	152/186 (82%)	106 (70%)	27 (18%)	19 (12%)	0	5
5	E	212/214 (99%)	148 (70%)	42 (20%)	22 (10%)	0	8
5	Q	212/214 (99%)	148 (70%)	43 (20%)	21 (10%)	0	9
6	F	82/155 (53%)	57 (70%)	15 (18%)	10 (12%)	0	6
6	R	82/155 (53%)	57 (70%)	15 (18%)	10 (12%)	0	6
7	G	169/171 (99%)	129 (76%)	27 (16%)	13 (8%)	1	13
7	S	169/171 (99%)	129 (76%)	27 (16%)	13 (8%)	1	13
8	H	124/145 (86%)	92 (74%)	17 (14%)	15 (12%)	0	6
8	T	127/145 (88%)	95 (75%)	17 (13%)	15 (12%)	0	6
9	I	109/115 (95%)	70 (64%)	25 (23%)	14 (13%)	0	5
9	U	109/115 (95%)	70 (64%)	25 (23%)	14 (13%)	0	5
10	J	58/72 (81%)	34 (59%)	15 (26%)	9 (16%)	0	3

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
10	V	58/72 (81%)	34 (59%)	15 (26%)	9 (16%)	0	3
11	K	112/118 (95%)	84 (75%)	17 (15%)	11 (10%)	0	10
11	W	112/118 (95%)	83 (74%)	18 (16%)	11 (10%)	0	10
12	L	44/73 (60%)	19 (43%)	12 (27%)	13 (30%)	0	0
12	X	44/73 (60%)	19 (43%)	12 (27%)	13 (30%)	0	0
All	All	7447/9046 (82%)	5152 (69%)	1485 (20%)	810 (11%)	0	8

5 of 810 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	42	ASP
1	A	48	PRO
1	A	54	ASN
1	A	57	LYS
1	A	67	CYS

### 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	972/1528 (64%)	877 (90%)	95 (10%)	8	26
1	M	974/1528 (64%)	879 (90%)	95 (10%)	8	26
2	B	813/1077 (76%)	730 (90%)	83 (10%)	7	25
2	N	817/1077 (76%)	734 (90%)	83 (10%)	7	25
3	C	157/264 (60%)	135 (86%)	22 (14%)	3	17
3	O	156/264 (59%)	134 (86%)	22 (14%)	3	16
4	D	78/160 (49%)	66 (85%)	12 (15%)	2	14
4	P	79/160 (49%)	68 (86%)	11 (14%)	3	17
5	E	155/197 (79%)	141 (91%)	14 (9%)	9	30
5	Q	155/197 (79%)	141 (91%)	14 (9%)	9	30
6	F	60/137 (44%)	54 (90%)	6 (10%)	7	26

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
6	R	60/137 (44%)	54 (90%)	6 (10%)	7	26
7	G	102/148 (69%)	93 (91%)	9 (9%)	10	31
7	S	104/148 (70%)	95 (91%)	9 (9%)	10	31
8	H	89/130 (68%)	82 (92%)	7 (8%)	12	35
8	T	91/130 (70%)	84 (92%)	7 (8%)	13	37
9	I	81/109 (74%)	68 (84%)	13 (16%)	2	13
9	U	81/109 (74%)	70 (86%)	11 (14%)	3	17
10	J	47/66 (71%)	40 (85%)	7 (15%)	3	15
10	V	47/66 (71%)	40 (85%)	7 (15%)	3	15
11	K	67/109 (62%)	56 (84%)	11 (16%)	2	12
11	W	68/109 (62%)	56 (82%)	12 (18%)	2	11
12	L	26/58 (45%)	25 (96%)	1 (4%)	33	57
12	X	26/58 (45%)	25 (96%)	1 (4%)	33	57
All	All	5305/7966 (67%)	4747 (90%)	558 (10%)	7	24

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

Mol	Chain	Res	Type
9	I	15	TYR
1	M	446	ASN
7	S	88	ASP
9	I	106	CYS
1	M	41	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 138 such sidechains are listed below:

Mol	Chain	Res	Type
9	I	46	HIS
1	M	494	GLN
6	R	104	ASN
9	I	90	GLN
1	M	119	ASN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 16 ligands modelled in this entry, 16 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

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

There are no chain breaks in this entry.

## 6 Fit of model and data [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	1384/1743 (79%)	0.50	114 (8%) 11 14	125, 172, 223, 262	0
1	M	1387/1743 (79%)	0.50	108 (7%) 13 15	134, 181, 232, 271	0
2	B	1070/1227 (87%)	0.66	122 (11%) 5 9	123, 180, 233, 267	0
2	N	1074/1227 (87%)	0.73	127 (11%) 4 8	133, 190, 242, 276	0
3	C	265/304 (87%)	0.56	19 (7%) 15 16	135, 169, 204, 228	0
3	O	265/304 (87%)	0.70	26 (9%) 7 10	144, 178, 213, 237	0
4	D	159/186 (85%)	1.04	32 (20%) 1 3	153, 191, 228, 234	0
4	P	164/186 (88%)	0.76	25 (15%) 2 5	162, 200, 239, 243	0
5	E	214/214 (100%)	0.37	7 (3%) 46 41	152, 214, 258, 267	0
5	Q	214/214 (100%)	0.50	18 (8%) 11 13	161, 223, 267, 276	0
6	F	84/155 (54%)	0.72	10 (11%) 4 8	129, 155, 185, 192	0
6	R	84/155 (54%)	0.73	12 (14%) 2 6	138, 164, 194, 201	0
7	G	171/171 (100%)	0.83	23 (13%) 3 6	157, 176, 212, 222	0
7	S	171/171 (100%)	0.66	13 (7%) 13 15	166, 185, 221, 231	0
8	H	130/145 (89%)	0.44	3 (2%) 60 54	178, 210, 238, 254	0
8	T	131/145 (90%)	0.61	11 (8%) 11 13	187, 219, 247, 263	0
9	I	113/115 (98%)	0.97	29 (25%) 0 2	170, 210, 241, 250	0
9	U	113/115 (98%)	0.85	22 (19%) 1 3	179, 219, 250, 259	0
10	J	62/72 (86%)	0.57	4 (6%) 18 18	131, 160, 200, 214	0
10	V	62/72 (86%)	0.62	7 (11%) 5 9	140, 169, 209, 223	0
11	K	114/118 (96%)	0.31	6 (5%) 26 26	142, 170, 194, 220	0
11	W	114/118 (96%)	0.53	7 (6%) 21 20	152, 179, 203, 230	0
12	L	46/73 (63%)	0.40	2 (4%) 35 32	166, 238, 253, 255	0
12	X	46/73 (63%)	1.02	9 (19%) 1 3	175, 248, 262, 265	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
All	All	7637/9046 (84%)	0.61	756 (9%) <b>7</b> <b>10</b>	123, 183, 240, 276	0

The worst 5 of 756 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1115	THR	7.8
2	N	388	GLN	7.0
2	N	389	ASP	6.7
2	N	216	VAL	6.7
2	N	721	ASP	6.5

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

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

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. 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
13	ZN	L	101	1/1	0.56	0.21	194,194,194,194	0
13	ZN	O	401	1/1	0.78	0.17	163,163,163,163	0
13	ZN	U	201	1/1	0.79	0.21	191,191,191,191	0
13	ZN	X	101	1/1	0.81	0.24	203,203,203,203	0
13	ZN	U	202	1/1	0.81	0.15	239,239,239,239	0
13	ZN	B	1301	1/1	0.83	0.17	152,152,152,152	0
13	ZN	M	1802	1/1	0.84	0.20	149,149,149,149	0
13	ZN	I	201	1/1	0.85	0.13	182,182,182,182	0
13	ZN	C	401	1/1	0.86	0.16	154,154,154,154	0
13	ZN	M	1801	1/1	0.89	0.10	174,174,174,174	0
13	ZN	A	1801	1/1	0.91	0.06	165,165,165,165	0
13	ZN	N	1301	1/1	0.91	0.10	161,161,161,161	0
13	ZN	A	1802	1/1	0.92	0.09	140,140,140,140	0

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
13	ZN	J	101	1/1	0.94	0.13	158,158,158,158	0
13	ZN	I	202	1/1	0.95	0.09	230,230,230,230	0
13	ZN	V	101	1/1	0.95	0.10	167,167,167,167	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.