



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 12, 2024 – 06:32 PM EDT

PDB ID : 4BXZ  
Title : RNA Polymerase II-Bye1 complex  
Authors : Kinkelin, K.; Wozniak, G.G.; Rothbart, S.B.; Lidschreiber, M.; Strahl, B.D.; Cramer, P.  
Deposited on : 2013-07-16  
Resolution : 4.80 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
Xtriage (Phenix)	:	1.20.1
EDS	:	2.36.2
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)	:	2.36.2

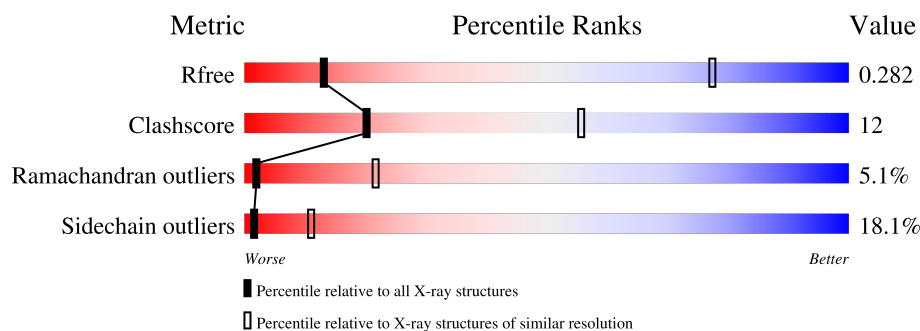
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.80 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	1096 (5.80-3.80)
Clashscore	141614	1170 (5.80-3.80)
Ramachandran outliers	138981	1105 (5.80-3.80)
Sidechain outliers	138945	1085 (5.80-3.80)

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 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	1733	44% 31% 6% • 18%
2	B	1224	51% 32% 6% 10%
3	C	318	47% 31% 5% 16%
4	D	221	50% 28% • 20%
5	E	215	59% 35% 5% •
6	F	155	28% 23% • 46%
7	G	171	67% 31% •

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
8	H	146	<div><div></div><div>47%38%<div><div></div><div></div><div></div></div>9%</div></div>
9	I	122	<div><div></div><div>61%34%<div><div></div><div></div><div></div></div></div></div>
10	J	70	<div><div></div><div>47%37%6%7%<div><div></div><div></div><div></div></div></div></div>
11	K	120	<div><div></div><div>65%24%7%<div><div></div><div></div><div></div></div></div></div>
12	L	70	<div><div></div><div>37%20%7%34%<div><div></div><div></div><div></div></div></div></div>
13	X	594	<div><div></div><div>17%81%<div><div></div><div></div><div></div></div></div></div>

## 2 Entry composition [i](#)

There are 15 unique types of molecules in this entry. The entry contains 31509 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 SUBUNIT RPB1.

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

- Molecule 2 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1097	Total	C	N	O	S	0	0	0
			8720	5526	1523	1617	54			

- Molecule 3 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB3.

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

- Molecule 4 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB4.

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

- Molecule 5 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC1.

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

- Molecule 6 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC2.

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

- Molecule 7 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB7.

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

- Molecule 8 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC3.

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

- Molecule 9 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB9.

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

- Molecule 10 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC5.

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

- Molecule 11 is a protein called DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	115	Total	C	N	O	S	0	0	1
			920	590	157	171	2			

- Molecule 12 is a protein called DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC4.

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

- Molecule 13 is a protein called TRANSCRIPTION FACTOR BYE1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
13	X	113	Total	C	N	O	0	0	0
			564	338	113	113			

- 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		
14	C	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	L	1	Total	Zn	0	0
			1	1		

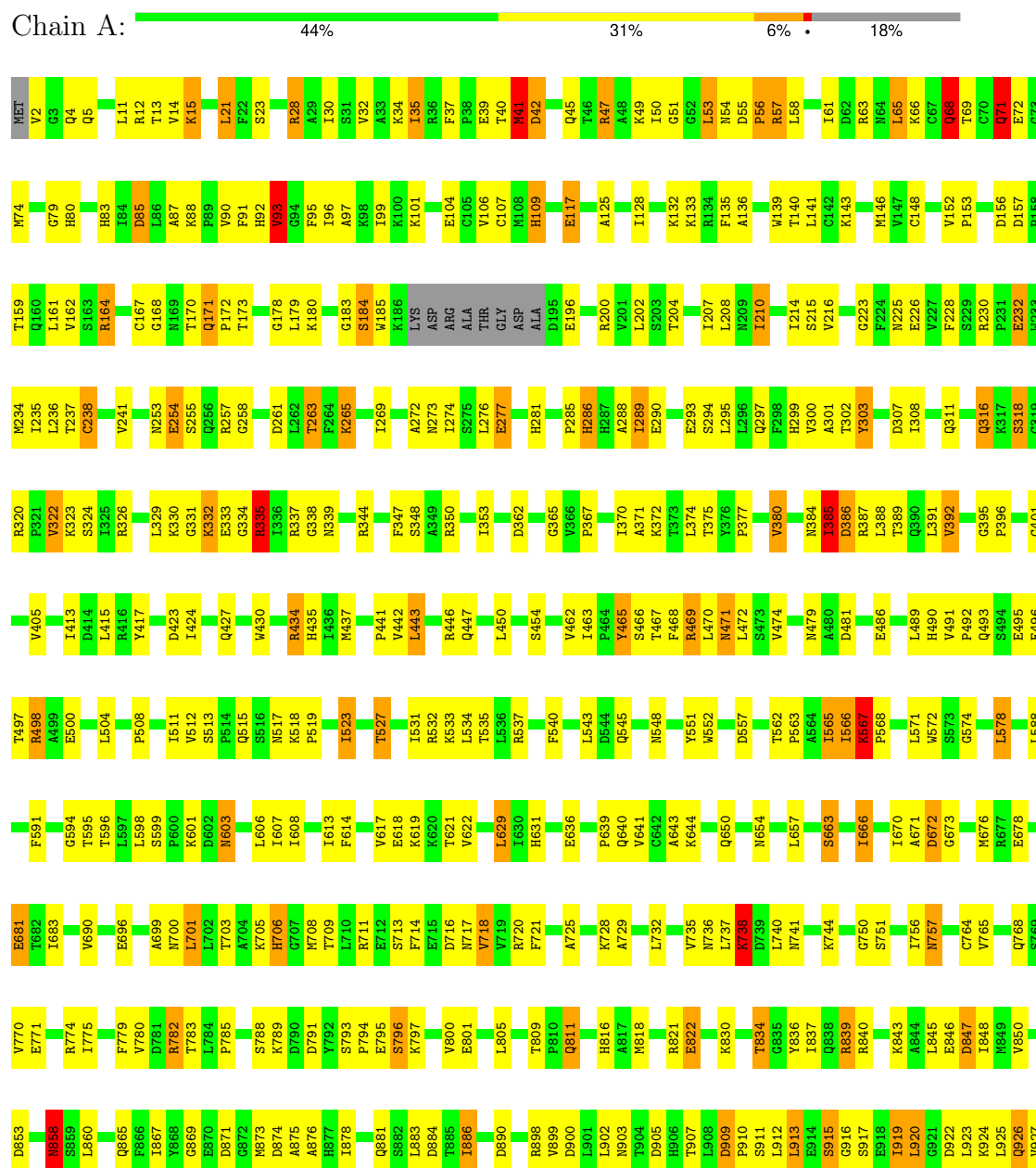
- Molecule 15 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

### 3 Residue-property plots

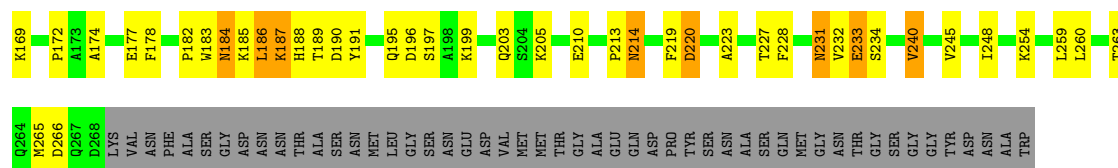
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

#### • Molecule 1: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB1



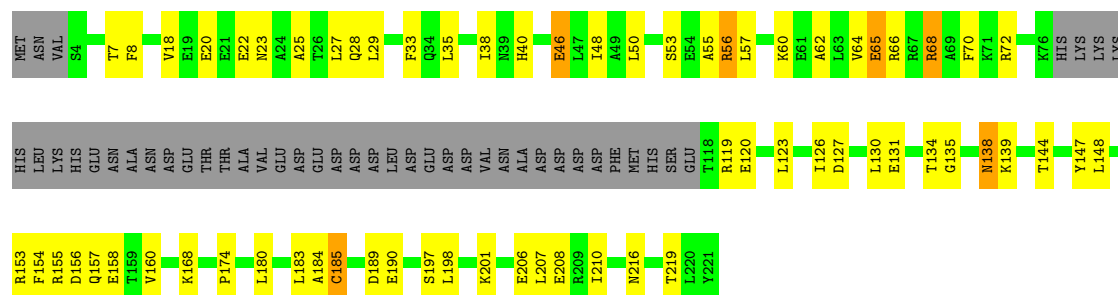






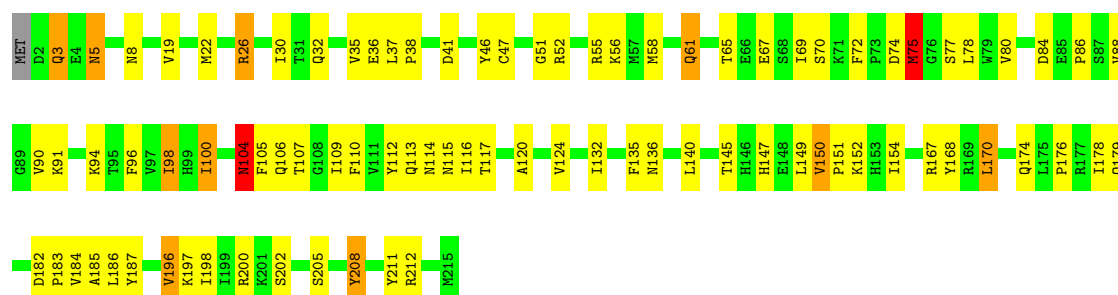
• Molecule 4: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB4

Chain D: 50% 28% 20%



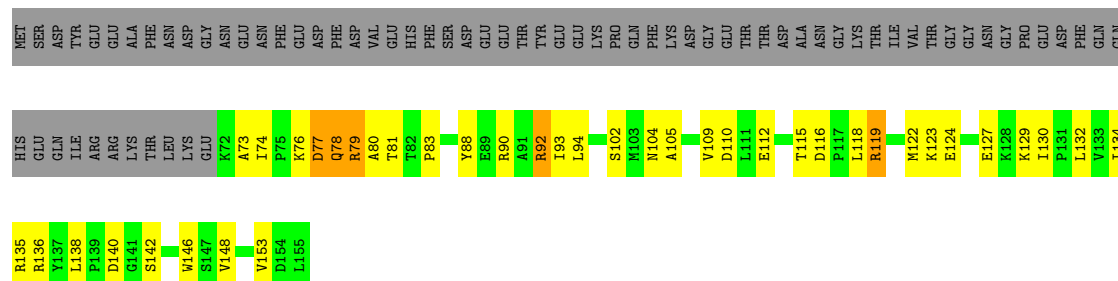
• Molecule 5: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC1

Chain E: 59% 35% 5%



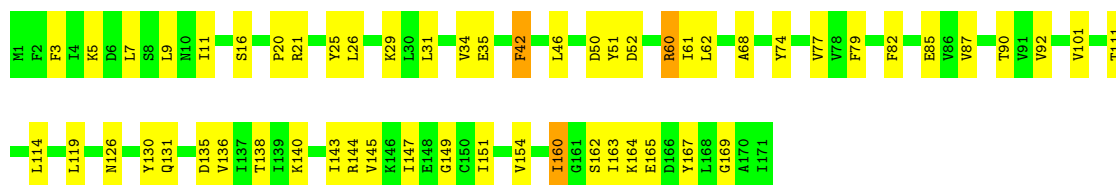
• Molecule 6: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC2

Chain F: 28% 23% 46%



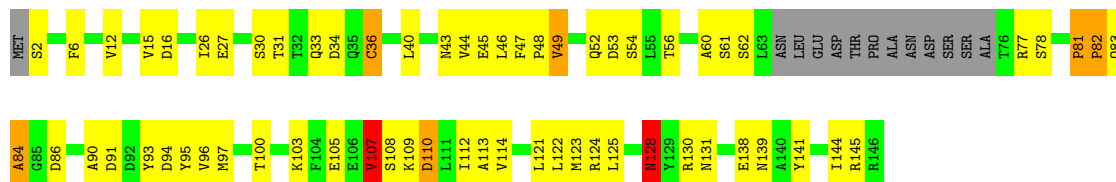
• Molecule 7: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB7

Chain G: 67% 31% 2%



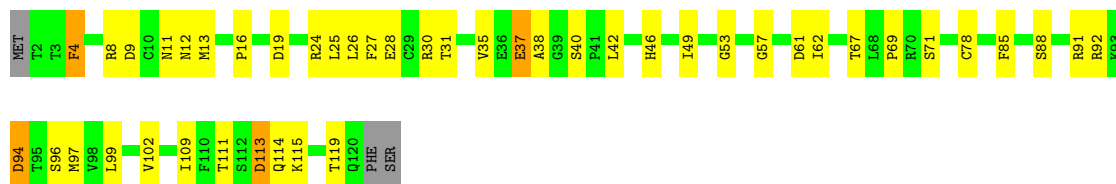
• Molecule 8: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC3

Chain H: 47% 38% 9%



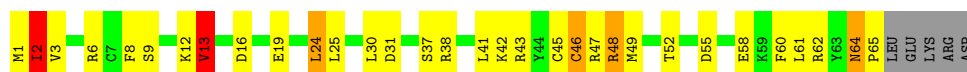
• Molecule 9: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB9

Chain I: 61% 34%



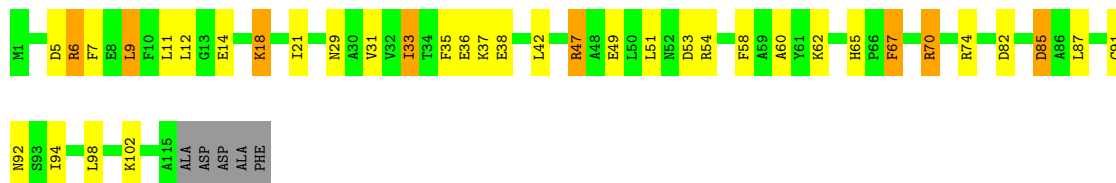
• Molecule 10: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC5

Chain J: 47% 37% 6% 7%



• Molecule 11: DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB11

Chain K: 65% 24% 7%

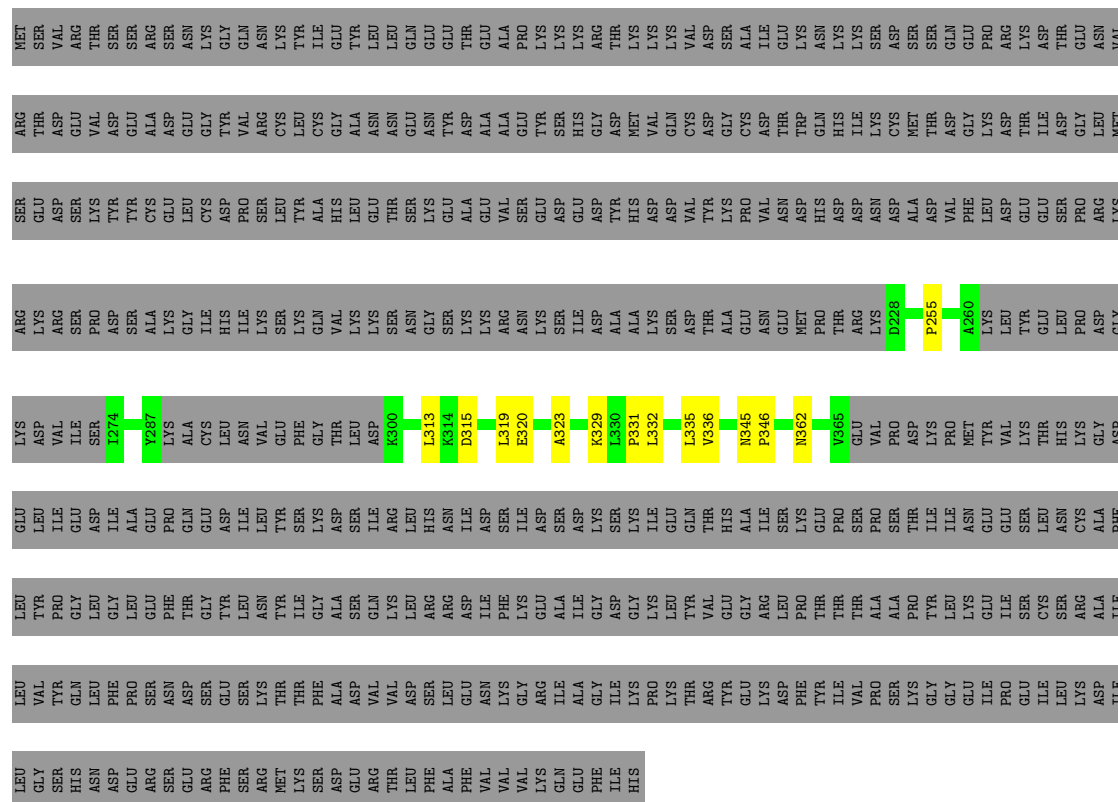


• Molecule 12: DNA-DIRECTED RNA POLYMERASES I, II, AND III SUBUNIT RPABC4

Chain L: 37% 20% 7% 34%



Chain X:  17% 81%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	220.55Å 392.09Å 279.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.63 – 4.80 49.63 – 4.80	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.63-4.80) 100.0 (49.63-4.80)	Depositor EDS
$R_{merge}$	0.41	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.90 (at 4.86Å)	Xtriage
Refinement program	BUSTER 2.11.5	Depositor
R, $R_{free}$	0.191 , 0.253 0.218 , 0.282	Depositor DCC
$R_{free}$ test set	1172 reflections (1.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	151.9	Xtriage
Anisotropy	0.442	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 289.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	0.017 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.039 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	31509	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	198.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.04% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.53	0/11339	0.82	2/15334 (0.0%)
2	B	0.52	0/8889	0.80	2/11987 (0.0%)
3	C	0.48	0/2133	0.77	0/2891
4	D	0.51	0/1365	0.82	0/1837
5	E	0.48	0/1788	0.77	0/2406
6	F	0.54	0/691	0.82	0/933
7	G	0.47	0/1368	0.78	0/1844
8	H	0.53	0/1086	0.85	2/1470 (0.1%)
9	I	0.48	0/989	0.79	0/1331
10	J	0.52	0/541	0.77	0/727
11	K	0.49	0/938	0.75	0/1267
12	L	0.54	0/365	0.94	0/485
13	X	0.62	0/561	0.88	3/780 (0.4%)
All	All	0.52	0/32053	0.81	9/43292 (0.0%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	56	PRO	C-N-CA	6.43	137.77	121.70
13	X	255	PRO	N-CA-CB	6.15	110.67	103.30
13	X	346	PRO	N-CA-CB	5.66	110.09	103.30
13	X	331	PRO	N-CA-CB	5.54	109.94	103.30
1	A	858	ASN	C-N-CA	5.39	135.18	121.70

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	11140	0	11218	359	0
2	B	8720	0	8745	225	0
3	C	2095	0	2051	71	0
4	D	1356	0	1319	31	0
5	E	1752	0	1776	46	0
6	F	679	0	701	21	0
7	G	1340	0	1357	26	0
8	H	1068	0	1040	29	0
9	I	971	0	930	25	0
10	J	532	0	543	16	0
11	K	920	0	929	27	0
12	L	363	0	386	12	0
13	X	564	0	239	4	0
14	A	2	0	0	0	0
14	B	1	0	0	0	0
14	C	1	0	0	0	0
14	I	2	0	0	0	0
14	J	1	0	0	0	0
14	L	1	0	0	0	0
15	A	1	0	0	0	0
All	All	31509	0	31234	783	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 12.

The worst 5 of 783 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:83:SER:HA	3:C:95:CYS:HB2	1.36	1.05
6:F:93:ILE:HD11	6:F:134:ILE:HD11	1.38	1.04
1:A:567:LYS:HB2	1:A:568:PRO:HD3	1.46	0.97
1:A:821:ARG:HG2	2:B:514:LEU:H	1.31	0.95
1:A:93:VAL:HG22	1:A:301:ALA:HA	1.53	0.89

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	1406/1733 (81%)	1111 (79%)	218 (16%)	77 (6%)	2	21
2	B	1075/1224 (88%)	861 (80%)	155 (14%)	59 (6%)	2	21
3	C	264/318 (83%)	208 (79%)	47 (18%)	9 (3%)	3	29
4	D	173/221 (78%)	144 (83%)	18 (10%)	11 (6%)	1	18
5	E	212/215 (99%)	176 (83%)	30 (14%)	6 (3%)	5	32
6	F	82/155 (53%)	69 (84%)	11 (13%)	2 (2%)	6	35
7	G	169/171 (99%)	143 (85%)	22 (13%)	4 (2%)	6	35
8	H	129/146 (88%)	96 (74%)	17 (13%)	16 (12%)	0	5
9	I	117/122 (96%)	96 (82%)	18 (15%)	3 (3%)	5	34
10	J	63/70 (90%)	46 (73%)	14 (22%)	3 (5%)	2	23
11	K	113/120 (94%)	94 (83%)	18 (16%)	1 (1%)	17	56
12	L	44/70 (63%)	25 (57%)	13 (30%)	6 (14%)	0	5
13	X	107/594 (18%)	83 (78%)	20 (19%)	4 (4%)	3	27
All	All	3954/5159 (77%)	3152 (80%)	601 (15%)	201 (5%)	2	22

5 of 201 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	35	ILE
1	A	47	ARG
1	A	57	ARG
1	A	65	LEU
1	A	71	GLN

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar

resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	1239/1520 (82%)	998 (80%)	241 (20%)	1	9
2	B	952/1061 (90%)	773 (81%)	179 (19%)	1	10
3	C	234/274 (85%)	197 (84%)	37 (16%)	2	15
4	D	140/200 (70%)	115 (82%)	25 (18%)	2	11
5	E	196/197 (100%)	164 (84%)	32 (16%)	2	14
6	F	74/137 (54%)	62 (84%)	12 (16%)	2	14
7	G	152/152 (100%)	130 (86%)	22 (14%)	3	17
8	H	117/128 (91%)	101 (86%)	16 (14%)	3	19
9	I	113/116 (97%)	98 (87%)	15 (13%)	4	20
10	J	60/65 (92%)	45 (75%)	15 (25%)	0	4
11	K	99/102 (97%)	82 (83%)	17 (17%)	2	12
12	L	40/57 (70%)	31 (78%)	9 (22%)	1	6
All	All	3416/4009 (85%)	2796 (82%)	620 (18%)	1	11

5 of 620 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	65	GLU
9	I	92	ARG
4	D	185	CYS
4	D	64	VAL
6	F	115	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 76 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	195	GLN
11	K	29	ASN
3	C	231	ASN
7	G	57	GLN
12	L	66	GLN

### 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 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	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	934:LYS	C	935:ARG	N	3.44

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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

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

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

### 6.3 Carbohydrates ⓘ

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

### 6.4 Ligands ⓘ

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

### 6.5 Other polymers ⓘ

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