



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 18, 2023 – 11:44 PM EDT

PDB ID : 5C4A
Title : Crystal structure of a transcribing RNA Polymerase II complex reveals a complete transcription bubble
Authors : Barnes, C.O.; Calero, M.; Malik, I.; Saphr, H.; Zhang, Q.; Pullara, F.; Kaplan, C.D.; Calero, G.
Deposited on : 2015-06-17
Resolution : 4.20 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ 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) : 1.13
EDS : 2.35.1
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

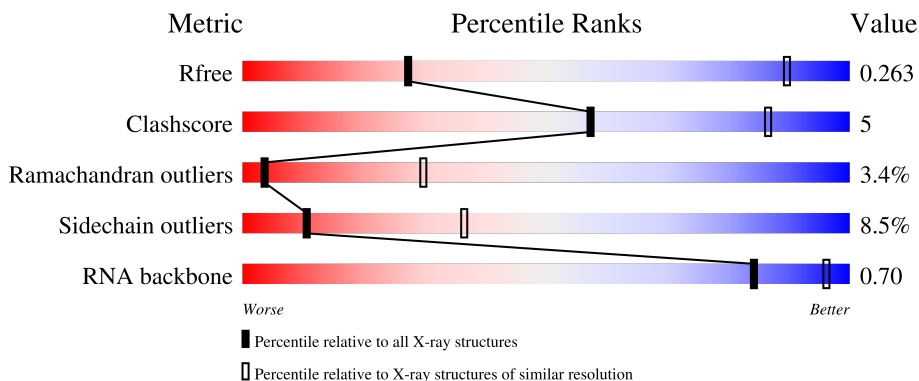
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.20 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 (4.62-3.78)
Clashscore	141614	1044 (4.60-3.80)
Ramachandran outliers	138981	1000 (4.60-3.80)
Sidechain outliers	138945	1007 (4.62-3.78)
RNA backbone	3102	1049 (5.04-3.00)

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 ≥ 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	64% (0 types), 16% (1 type), 18% (2 types), 2% (3 types), 0% (4 types), 0% (not modelled)
2	B	1224	71% (0 types), 21% (1 type), 5% (2 types), 3% (3 types), 0% (4 types), 0% (not modelled)
3	C	318	67% (0 types), 14% (1 type), 17% (2 types), 2% (3 types), 0% (4 types), 0% (not modelled)
4	D	221	65% (0 types), 11% (1 type), 24% (2 types), 0% (3 types), 0% (4 types), 0% (not modelled)
5	E	215	85% (0 types), 14% (1 type), 0% (2 types), 0% (3 types), 0% (4 types), 0% (not modelled)
6	F	155	48% (0 types), 8% (1 type), 44% (2 types), 0% (3 types), 0% (4 types), 0% (not modelled)

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	G	179	75% 18% •• 5%
8	H	146	71% 21% • 8%
9	I	122	80% 12% • 7%
10	J	70	71% 20% • 6%
11	K	120	87% 8% ••
12	L	70	47% 16% 37%
13	R	9	67% 33%
14	S	56	23% 5% 71%
15	U	56	38% 14% 48%

2 Entry composition [i](#)

There are 17 unique types of molecules in this entry. The entry contains 32574 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
			Total	C	N	O	S			
1	A	1425	11206	7057	1960	2127	62	0	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerase II subunit RPB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1162	9215	5816	1617	1726	56	0	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	265	2086	1312	347	414	13	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerase II subunit RPB4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	168	1331	822	237	270	2	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

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 subunit RPABC2.

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 subunit RPB7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	G	170	1331	857	220	246	8	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	172	LEU	-	expression tag	UNP P34087
G	173	GLU	-	expression tag	UNP P34087
G	174	HIS	-	expression tag	UNP P34087
G	175	HIS	-	expression tag	UNP P34087
G	176	HIS	-	expression tag	UNP P34087
G	177	HIS	-	expression tag	UNP P34087
G	178	HIS	-	expression tag	UNP P34087
G	179	HIS	-	expression tag	UNP P34087

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	H	135	1080	679	182	214	5	0	0	0

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
9	I	114	927	571	168	178	10	0	0	0

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
10	J	66	540	345	94	95	6	0	0	0

- Molecule 11 is a protein called DNA-directed RNA polymerase II subunit RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
11	K	115	924	593	157	172	2	0	0	0

- 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	44	Total	C	N	O	S	0	0	0
			352	217	70	61	4			

- Molecule 13 is a RNA chain called RNA (5'-R(P*UP*CP*GP*AP*GP*AP*GP*GP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	R	9	Total	C	N	O	P	0	0	0
			197	88	40	60	9			

- Molecule 14 is a DNA chain called Scaffold 2 Non-template Strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
14	S	16	Total	C	N	O	P	0	0	0
			331	158	58	99	16			

- Molecule 15 is a DNA chain called Scaffold 2 Template Strand.

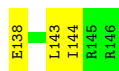
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
15	U	29	Total	C	N	O	P	0	0	0
			587	280	107	171	29			

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

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

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
17	A	2	Total	Mg	0	0
			2	2		



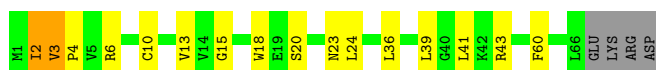
- Molecule 9: DNA-directed RNA polymerase II subunit RPB9

Chain I: 80% 12% 7%



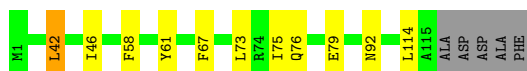
- Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J: 71% 20% 6%



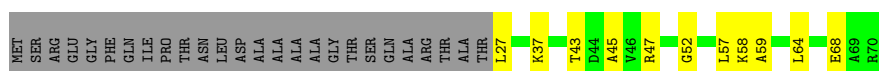
- Molecule 11: DNA-directed RNA polymerase II subunit RPB11

Chain K: 87% 8% 2%



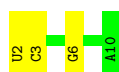
- Molecule 12: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L: 47% 16% 37%



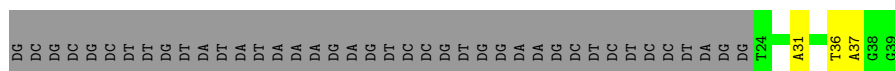
- Molecule 13: RNA (5'-R(P*UP*CP*GP*AP*GP*AP*GP*GP*A)-3')

Chain R: 67% 33%




- Molecule 14: Scaffold 2 Non-template Strand

Chain S: 23% 5% 71%



- Molecule 15: Scaffold 2 Template Strand

Chain U:  38% 14% 48%



4 Data and refinement statistics i

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	219.83Å 396.71Å 273.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.91 – 4.20 39.91 – 4.20	Depositor EDS
% Data completeness (in resolution range)	99.7 (39.91-4.20) 99.7 (39.91-4.20)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.63 (at 4.13Å)	Xtrriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.224 , 0.242 0.246 , 0.263	Depositor DCC
R_{free} test set	6719 reflections (7.74%)	wwPDB-VP
Wilson B-factor (Å ²)	100.3	Xtrriage
Anisotropy	0.017	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 121.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	0.019 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.026 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtrriage
F_o, F_c correlation	0.85	EDS
Total number of atoms	32574	wwPDB-VP
Average B, all atoms (Å ²)	139.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.36% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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, 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	0.52	0/11407	0.66	0/15428
2	B	0.52	1/9390 (0.0%)	0.67	0/12662
3	C	0.49	0/2124	0.61	0/2879
4	D	0.50	0/1339	0.63	0/1793
5	E	0.51	0/1788	0.60	0/2406
6	F	0.51	0/717	0.64	0/967
7	G	0.50	0/1358	0.66	0/1830
8	H	0.47	0/1097	0.63	0/1484
9	I	0.48	0/945	0.62	0/1273
10	J	0.49	0/549	0.65	0/738
11	K	0.48	0/942	0.60	0/1272
12	L	0.51	0/354	0.65	0/468
13	R	0.78	0/221	0.80	0/343
14	S	0.95	0/370	0.95	0/570
15	U	1.04	0/657	0.95	0/1009
All	All	0.53	1/33258 (0.0%)	0.67	0/45122

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

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	264	SER	C-N	7.45	1.51	1.34

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	265	SER	Mainchain

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	11206	0	11265	134	0
2	B	9215	0	9210	111	0
3	C	2086	0	2045	23	0
4	D	1331	0	1345	8	0
5	E	1752	0	1776	9	0
6	F	705	0	731	5	0
7	G	1331	0	1350	19	0
8	H	1080	0	1049	10	0
9	I	927	0	883	7	0
10	J	540	0	555	9	0
11	K	924	0	934	3	0
12	L	352	0	377	2	0
13	R	197	0	96	4	0
14	S	331	0	183	2	0
15	U	587	0	326	6	0
16	A	2	0	0	0	0
16	B	1	0	0	0	0
16	C	1	0	0	0	0
16	I	2	0	0	0	0
16	J	1	0	0	0	0
16	L	1	0	0	0	0
17	A	2	0	0	0	0
All	All	32574	0	32125	316	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 5.

The worst 5 of 316 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:869:GLY:HA3	1:A:1366:ARG:HG2	1.46	0.98
3:C:39:ALA:HA	3:C:164:ALA:HB3	1.64	0.80
1:A:768:GLN:HG2	1:A:816:HIS:HA	1.64	0.79
1:A:311:GLN:HB2	1:A:312:PRO:HD3	1.67	0.77
1:A:1107:VAL:HA	1:A:1108:ALA:HB3	1.68	0.75

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	1417/1733 (82%)	1222 (86%)	147 (10%)	48 (3%)	3	30
2	B	1146/1224 (94%)	977 (85%)	124 (11%)	45 (4%)	3	27
3	C	263/318 (83%)	235 (89%)	20 (8%)	8 (3%)	4	32
4	D	164/221 (74%)	150 (92%)	8 (5%)	6 (4%)	3	28
5	E	212/215 (99%)	197 (93%)	9 (4%)	6 (3%)	5	33
6	F	85/155 (55%)	74 (87%)	10 (12%)	1 (1%)	13	50
7	G	166/179 (93%)	147 (89%)	14 (8%)	5 (3%)	4	32
8	H	129/146 (88%)	111 (86%)	10 (8%)	8 (6%)	1	20
9	I	112/122 (92%)	100 (89%)	10 (9%)	2 (2%)	8	42
10	J	64/70 (91%)	57 (89%)	6 (9%)	1 (2%)	9	45
11	K	113/120 (94%)	109 (96%)	4 (4%)	0	100	100
12	L	42/70 (60%)	31 (74%)	9 (21%)	2 (5%)	2	24
All	All	3913/4573 (86%)	3410 (87%)	371 (10%)	132 (3%)	3	30

5 of 132 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	49	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	66	LYS
1	A	76	GLU
1	A	194	ALA
1	A	286	HIS

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%)	1128 (91%)	116 (9%)	9	31
2	B	1002/1061 (94%)	900 (90%)	102 (10%)	7	28
3	C	233/274 (85%)	219 (94%)	14 (6%)	19	47
4	D	146/200 (73%)	138 (94%)	8 (6%)	21	49
5	E	196/197 (100%)	188 (96%)	8 (4%)	30	56
6	F	77/137 (56%)	73 (95%)	4 (5%)	23	50
7	G	151/160 (94%)	141 (93%)	10 (7%)	16	44
8	H	118/128 (92%)	108 (92%)	10 (8%)	10	36
9	I	108/116 (93%)	103 (95%)	5 (5%)	27	54
10	J	61/65 (94%)	57 (93%)	4 (7%)	16	44
11	K	99/102 (97%)	92 (93%)	7 (7%)	14	42
12	L	39/57 (68%)	33 (85%)	6 (15%)	2	16
All	All	3474/4017 (86%)	3180 (92%)	294 (8%)	10	36

5 of 294 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	260	LEU
11	K	92	ASN
4	D	214	LEU
7	G	143	ILE
1	A	1186	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	109	HIS
1	A	611	GLN
1	A	1078	GLN
4	D	41	GLN
8	H	131	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
13	R	7/9 (77%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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 10 ligands modelled in this entry, 10 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

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.