



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 08:26 PM UTC

PDB ID : 5C44 / pdb\_00005c44  
Title : Crystal structure of a transcribing RNA Polymerase II complex reveals a complete transcription bubble  
Authors : Barnes, C.O.; Calero, M.; Malik, I.; Spahr, H.; Zhang, Q.; Pullara, F.; Kaplan, C.D.; Calero, G.  
Deposited on : 2015-06-17  
Resolution : 3.95 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 : 4-5-2 with Phenix2.0  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.49

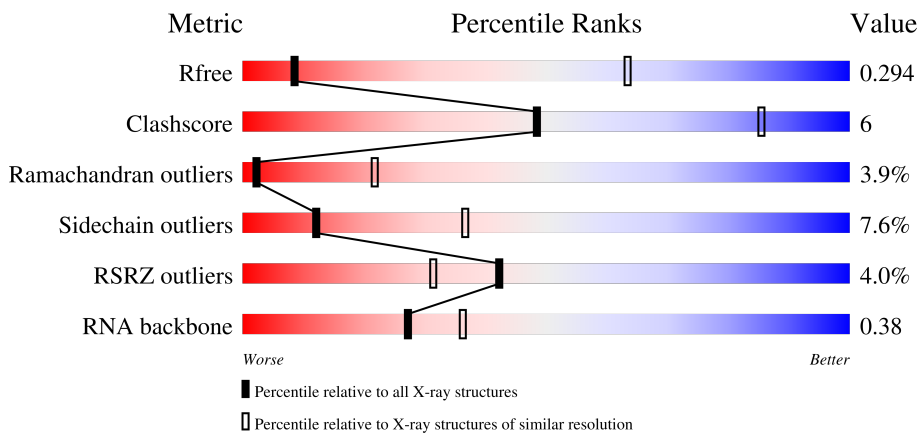
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.95 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))
$R_{free}$	180053	1046 (4.16-3.76)
Clashscore	190562	1019 (4.14-3.78)
Ramachandran outliers	187476	1031 (4.16-3.76)
Sidechain outliers	187428	1024 (4.16-3.76)
RSRZ outliers	180081	1046 (4.16-3.76)
RNA backbone	3983	1016 (4.80-3.10)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	1733	 2% (Poor fit), 64% (0 outliers), 16% (1 outlier), 17% (2+ outliers)
2	B	1224	 3% (Poor fit), 70% (0 outliers), 22% (1 outlier), 5% (2+ outliers)
3	C	318	 2% (Poor fit), 66% (0 outliers), 14% (1 outlier), 17% (2+ outliers)
4	D	221	 2% (Poor fit), 64% (0 outliers), 15% (1 outlier), 19% (2+ outliers)

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	E	215	
6	F	155	
7	G	179	
8	H	146	
9	I	120	
10	J	70	
11	K	120	
12	L	70	
13	R	9	
14	S	53	
15	U	53	

## 2 Entry composition [i](#)

There are 15 unique types of molecules in this entry. The entry contains 32540 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	1434	11249	7083	1967	2137	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	1158	9175	5795	1603	1721	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	178	1417	875	254	286	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	171	1339	861	222	248	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	921	568	165	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	65	532	339	93	94	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
			Total	C	N	O	S			
12	L	44	346	214	67	61	4	0	0	0

- Molecule 13 is a RNA chain called Synthetic RNA.

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

- Molecule 14 is a DNA chain called Synthetic DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
14	S	13	268	128	46	81	13	0	0	0

- Molecule 15 is a DNA chain called Synthetic DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
15	U	27	549	261	102	159	27	0	0	0











## 4 Data and refinement statistics i

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	220.22Å 391.84Å 282.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	79.00 – 3.95 79.00 – 3.95	Depositor EDS
% Data completeness (in resolution range)	94.3 (79.00-3.95) 94.5 (79.00-3.95)	Depositor EDS
$R_{merge}$	0.23	Depositor
$R_{sym}$	0.23	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.62 (at 4.01Å)	Xtrriage
Refinement program	BUSTER 2.10.2	Depositor
R, $R_{free}$	0.215 , 0.236 0.272 , 0.294	Depositor DCC
$R_{free}$ test set	3013 reflections (2.83%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	114.2	Xtrriage
Anisotropy	0.078	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 232.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.32$ , $\langle L^2 \rangle = 0.15$	Xtrriage
Estimated twinning fraction	0.178 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.186 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtrriage
$F_o, F_c$ correlation	0.86	EDS
Total number of atoms	32540	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	197.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.15% 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

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	0.79	1/11452 (0.0%)	1.36	35/15492 (0.2%)
2	B	0.76	3/9347 (0.0%)	1.27	22/12601 (0.2%)
3	C	0.77	0/2124	1.26	2/2879 (0.1%)
4	D	0.71	0/1427	1.35	3/1911 (0.2%)
5	E	0.72	0/1788	1.25	4/2406 (0.2%)
6	F	0.74	0/717	1.28	0/967
7	G	0.76	0/1367	1.20	1/1844 (0.1%)
8	H	0.73	0/1097	1.10	2/1484 (0.1%)
9	I	0.72	0/939	1.21	0/1266
10	J	0.70	0/541	1.28	0/727
11	K	0.70	0/942	1.31	0/1272
12	L	0.75	0/348	1.21	0/461
13	R	0.56	0/221	0.73	1/343 (0.3%)
14	S	1.42	2/299 (0.7%)	0.85	0/460
15	U	1.05	0/615	0.63	0/945
All	All	0.77	6/33224 (0.0%)	1.28	70/45058 (0.2%)

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
14	S	0	1
15	U	0	1
All	All	0	2

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	260	GLY	C-N	7.77	1.47	1.33
2	B	265	SER	C-N	7.64	1.44	1.33
2	B	932	HIS	CG-ND1	5.63	1.44	1.38

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
14	S	27	DT	O5'-C5'	5.54	1.59	1.42
1	A	1107	VAL	CA-C	5.37	1.59	1.52

The worst 5 of 70 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	260	GLY	O-C-N	-10.39	109.19	122.70
2	B	79	THR	CA-C-N	8.20	136.46	121.70
2	B	79	THR	C-N-CA	8.20	136.46	121.70
2	B	136	THR	CA-C-N	7.63	135.43	121.70
2	B	136	THR	C-N-CA	7.63	135.43	121.70

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
14	S	32	DA	Sidechain
15	U	8	DA	Sidechain

## 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	11249	0	11277	121	0
2	B	9175	0	9138	128	0
3	C	2086	0	2049	24	0
4	D	1417	0	1428	10	0
5	E	1752	0	1776	11	0
6	F	705	0	731	13	0
7	G	1339	0	1357	13	0
8	H	1080	0	1049	9	0
9	I	921	0	877	10	0
10	J	532	0	546	5	0
11	K	924	0	934	11	0
12	L	346	0	367	3	0
13	R	197	0	97	4	0
14	S	268	0	149	25	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
15	U	549	0	303	33	0
All	All	32540	0	32078	354	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 6.

The worst 5 of 354 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)
14:S:38:DA:N6	15:U:2:DC:N4	2.15	0.93
14:S:38:DA:N6	15:U:2:DC:C4	2.43	0.86
1:A:567:LYS:HB3	1:A:568:PRO:CD	2.07	0.85
1:A:67:CYS:HG	1:A:77:CYS:HG	1.12	0.85
14:S:38:DA:H62	15:U:2:DC:N4	1.74	0.83

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	1426/1733 (82%)	1227 (86%)	135 (10%)	64 (4%)	2	20
2	B	1134/1224 (93%)	967 (85%)	121 (11%)	46 (4%)	2	21
3	C	263/318 (83%)	228 (87%)	26 (10%)	9 (3%)	3	24
4	D	174/221 (79%)	152 (87%)	12 (7%)	10 (6%)	1	17
5	E	212/215 (99%)	198 (93%)	10 (5%)	4 (2%)	6	33
6	F	85/155 (55%)	79 (93%)	3 (4%)	3 (4%)	3	23
7	G	169/179 (94%)	143 (85%)	21 (12%)	5 (3%)	3	26
8	H	129/146 (88%)	111 (86%)	14 (11%)	4 (3%)	3	25

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
9	I	112/120 (93%)	98 (88%)	13 (12%)	1 (1%)	14	48
10	J	63/70 (90%)	56 (89%)	3 (5%)	4 (6%)	1	15
11	K	113/120 (94%)	110 (97%)	3 (3%)	0	100	100
12	L	42/70 (60%)	31 (74%)	8 (19%)	3 (7%)	1	14
All	All	3922/4571 (86%)	3400 (87%)	369 (9%)	153 (4%)	2	22

5 of 153 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	49	LYS
1	A	66	LYS
1	A	319	GLY
1	A	385	ILE
1	A	567	LYS

### 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%)	1139 (92%)	105 (8%)	10	33
2	B	996/1061 (94%)	917 (92%)	79 (8%)	11	35
3	C	233/274 (85%)	215 (92%)	18 (8%)	12	36
4	D	156/200 (78%)	145 (93%)	11 (7%)	13	39
5	E	196/197 (100%)	184 (94%)	12 (6%)	17	42
6	F	77/137 (56%)	74 (96%)	3 (4%)	28	51
7	G	152/160 (95%)	143 (94%)	9 (6%)	18	43
8	H	118/128 (92%)	114 (97%)	4 (3%)	32	55
9	I	107/114 (94%)	105 (98%)	2 (2%)	50	67
10	J	60/65 (92%)	53 (88%)	7 (12%)	5	21
11	K	99/102 (97%)	92 (93%)	7 (7%)	13	39
12	L	38/57 (67%)	31 (82%)	7 (18%)	1	11

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	3476/4015 (87%)	3212 (92%)	264 (8%)	12 37

5 of 264 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	G	57	GLN
8	H	111	LEU
12	L	40	LEU
1	A	1371	LEU
1	A	1356	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 66 such sidechains are listed below:

Mol	Chain	Res	Type
8	H	139	ASN
9	I	51	ASN
11	K	96	ASN
2	B	115	GLN
2	B	103	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
13	R	7/9 (77%)	2 (28%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
13	R	3	C
13	R	6	G

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 oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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	1434/1733 (82%)	0.25	43 (2%) 52 37	75, 183, 269, 300	0
2	B	1158/1224 (94%)	0.36	39 (3%) 48 34	84, 195, 283, 300	0
3	C	265/318 (83%)	0.30	6 (2%) 61 43	119, 190, 255, 297	0
4	D	178/221 (80%)	0.32	5 (2%) 55 39	115, 201, 273, 300	0
5	E	214/215 (99%)	0.36	11 (5%) 33 26	131, 218, 277, 294	0
6	F	87/155 (56%)	0.01	2 (2%) 61 43	91, 156, 211, 261	0
7	G	171/179 (95%)	0.24	6 (3%) 47 33	117, 184, 232, 300	0
8	H	135/146 (92%)	0.36	5 (3%) 45 33	150, 236, 291, 300	0
9	I	114/120 (95%)	0.57	7 (6%) 27 23	156, 234, 282, 300	0
10	J	65/70 (92%)	0.31	2 (3%) 51 36	114, 197, 266, 287	0
11	K	115/120 (95%)	0.28	2 (1%) 69 49	113, 189, 252, 273	0
12	L	44/70 (62%)	0.77	2 (4%) 38 29	155, 219, 273, 300	0
13	R	9/9 (100%)	1.11	2 (22%) 2 4	230, 267, 300, 300	0
14	S	13/53 (24%)	3.66	13 (100%) 0 0	20, 167, 300, 300	0
15	U	27/53 (50%)	2.48	18 (66%) 0 1	160, 198, 300, 300	0
All	All	4029/4686 (85%)	0.34	163 (4%) 42 31	20, 193, 276, 300	0

The worst 5 of 163 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1081	LEU	5.1
15	U	10	DA	5.1
15	U	5	DC	4.9
14	S	28	DG	4.8
14	S	38	DA	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 oligosaccharides in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.