



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 30, 2020 – 04:49 am BST

PDB ID : 6P70  
Title : X-ray crystal structure of bacterial RNA polymerase and pyrBI promoter complex  
Authors : Shin, Y.; Murakami, K.S.  
Deposited on : 2019-06-04  
Resolution : 3.05 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

PERCENTILES INFOmissingINFO

# 1 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 28421 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 alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	226	Total	C	N	O	S	0	0	0
			1782	1138	310	332	2			
1	B	222	Total	C	N	O	S	0	0	0
			1750	1118	304	326	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1111	Total	C	N	O	S	0	0	0
			8764	5545	1561	1634	24			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1486	Total	C	N	O	S	0	0	0
			11738	7440	2067	2195	36			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	94	Total	C	N	O	S	0	0	0
			761	486	132	139	4			

- Molecule 5 is a protein called RNA polymerase sigma factor SigA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	346	Total	C	N	O	S	0	0	0
			2807	1770	509	524	4			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	46	THR	ALA	conflict	UNP Q72L95

- Molecule 6 is a DNA chain called DNA (5'-D(P\*TP\*CP\*CP\*CP\*GP\*GP\*CP\*AP\*AP\*AP\*TP\*TP\*GP\*TP\*CP\*CP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	18	Total	C	N	O	P	0	0	0
			364	173	64	109	18			

- Molecule 7 is a DNA chain called DNA (5'-D(\*TP\*AP\*TP\*AP\*AP\*TP\*CP\*GP\*AP\*TP\*CP\*TP\*TP\*TP\*GP\*CP\*CP\*GP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	22	Total	C	N	O	P	0	0	0
			450	216	81	132	21			

- Molecule 8 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	G	1	Total	Mg	0	0
			1	1		
8	B	1	Total	Mg	0	0
			1	1		
8	D	1	Total	Mg	0	0
			1	1		

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	2	Total	Zn	0	0
			2	2		

SEQUENCE-PLOTS INFOmissingINFO

## 2 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	185.18Å 100.84Å 294.87Å 90.00° 98.81° 90.00°	Depositor
Resolution (Å)	46.06 – 3.05 46.06 – 3.05	Depositor EDS
% Data completeness (in resolution range)	99.1 (46.06-3.05) 99.1 (46.06-3.05)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.70 (at 3.06Å)	Xtriage
Refinement program	PHENIX (1.14_3260)	Depositor
R, $R_{free}$	0.207 , 0.252 (Not available) , (Not available)	Depositor DCC
$R_{free}$ test set	2000 reflections (1.97%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	87.8	Xtriage
Anisotropy	0.519	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 51.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	28421	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	100.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.89% 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.

## 3 Model quality ⓘ

### 3.1 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	0/1814	0.75	0/2466
1	B	0.54	0/1782	0.75	0/2424
2	C	0.53	0/8931	0.78	4/12080 (0.0%)
3	D	0.57	2/11944 (0.0%)	0.82	6/16148 (0.0%)
4	E	0.49	0/775	0.76	0/1045
5	F	0.55	0/2852	0.83	5/3837 (0.1%)
6	G	1.22	1/406 (0.2%)	1.03	2/623 (0.3%)
7	H	1.42	1/503 (0.2%)	1.13	1/773 (0.1%)
All	All	0.59	4/29007 (0.0%)	0.81	18/39396 (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
3	D	0	1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	H	23	DG	O3'-P	-17.21	1.40	1.61
6	G	19	DC	C1'-N1	7.03	1.58	1.49
3	D	156	GLU	CG-CD	6.07	1.61	1.51
3	D	734	GLU	CG-CD	5.23	1.59	1.51

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	H	23	DG	P-O3'-C3'	8.35	129.72	119.70

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	944	LEU	CB-CG-CD2	-7.53	98.20	111.00
6	G	15	DT	O4'-C4'-C3'	-7.04	101.68	104.50
2	C	775	ARG	CB-CG-CD	-6.94	93.55	111.60
5	F	405	LEU	CA-CB-CG	6.88	131.12	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	D	620	GLY	Mainchain

### 3.2 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	1782	0	1834	37	0
1	B	1750	0	1797	30	0
2	C	8764	0	8863	211	1
3	D	11738	0	11969	249	0
4	E	761	0	778	9	0
5	F	2807	0	2882	81	1
6	G	364	0	203	12	0
7	H	450	0	252	13	0
8	B	1	0	0	0	0
8	D	1	0	0	0	0
8	G	1	0	0	0	0
9	D	2	0	0	0	0
All	All	28421	0	28578	571	1

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 10.

The worst 5 of 571 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:C:775:ARG:HE	2:C:782:ALA:HB2	1.28	0.98

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:F:372:ARG:HG2	5:F:386:VAL:HG21	1.47	0.97
3:D:218:LYS:HG2	3:D:338:GLU:HG2	1.44	0.97
5:F:193:ARG:HB3	7:H:7:DC:H5"	1.52	0.92
2:C:709:GLU:OE2	2:C:824:ARG:NH1	2.07	0.88

All (1) 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)
2:C:68:PHE:O	5:F:396:ARG:NH2[1_545]	1.84	0.36

### 3.3 Torsion angles [i](#)

#### 3.3.1 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	224/315 (71%)	222 (99%)	2 (1%)	0	100	100
1	B	220/315 (70%)	209 (95%)	10 (4%)	1 (0%)	29	60
2	C	1107/1119 (99%)	1077 (97%)	27 (2%)	3 (0%)	41	70
3	D	1482/1524 (97%)	1444 (97%)	35 (2%)	3 (0%)	47	77
4	E	92/99 (93%)	89 (97%)	2 (2%)	1 (1%)	14	42
5	F	344/423 (81%)	327 (95%)	14 (4%)	3 (1%)	17	47
All	All	3469/3795 (91%)	3368 (97%)	90 (3%)	11 (0%)	41	70

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	1130	ARG
1	B	154	GLU
2	C	764	GLU
3	D	1236	LEU

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
5	F	356	LYS

### 3.3.2 Protein sidechains ⓘ

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	199/273 (73%)	192 (96%)	7 (4%)	36	66
1	B	195/273 (71%)	190 (97%)	5 (3%)	46	72
2	C	935/941 (99%)	891 (95%)	44 (5%)	26	56
3	D	1253/1279 (98%)	1187 (95%)	66 (5%)	22	51
4	E	83/88 (94%)	81 (98%)	2 (2%)	49	74
5	F	301/371 (81%)	294 (98%)	7 (2%)	50	75
All	All	2966/3225 (92%)	2835 (96%)	131 (4%)	28	58

5 of 131 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	81	THR
3	D	276	ASP
3	D	1470	ARG
3	D	141	ILE
3	D	191	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	714	GLN
3	D	724	GLN
3	D	1445	HIS
3	D	636	GLN
3	D	1359	GLN

### 3.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

### 3.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

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

Of 5 ligands modelled in this entry, 5 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.

### 3.7 Other polymers [i](#)

There are no such residues in this entry.

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

There are no chain breaks in this entry.

## 4 Fit of model and data ⓘ

### 4.1 Protein, DNA and RNA chains ⓘ

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	226/315 (71%)	-0.29	0	100	100	70, 93, 114, 122	0
1	B	222/315 (70%)	-0.32	0	100	100	68, 96, 123, 140	0
2	C	1111/1119 (99%)	-0.08	34 (3%)	49	25	53, 93, 163, 194	0
3	D	1486/1524 (97%)	-0.18	24 (1%)	72	49	49, 87, 150, 181	0
4	E	94/99 (94%)	-0.27	0	100	100	66, 95, 132, 147	0
5	F	346/423 (81%)	0.47	49 (14%)	2	1	62, 107, 194, 203	0
6	G	18/21 (85%)	0.01	1 (5%)	24	10	98, 119, 167, 170	0
7	H	22/27 (81%)	-0.32	2 (9%)	9	3	88, 126, 166, 182	0
All	All	3525/3843 (91%)	-0.10	110 (3%)	49	25	49, 93, 161, 203	0

The worst 5 of 110 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	F	382	THR	9.6
5	F	376	ILE	9.3
2	C	770	GLU	9.2
5	F	377	ASP	9.1
5	F	381	HIS	9.0

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

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

### 4.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 4.4 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( $\text{\AA}^2$ )	Q<0.9
8	MG	G	101	1/1	0.51	0.34	113,113,113,113	0
9	ZN	D	2001	1/1	0.97	0.18	96,96,96,96	0
8	MG	D	2003	1/1	0.98	0.36	50,50,50,50	0
8	MG	B	401	1/1	0.99	0.39	69,69,69,69	0
9	ZN	D	2002	1/1	1.00	0.08	124,124,124,124	0

## 4.5 Other polymers [i](#)

There are no such residues in this entry.