



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 27, 2020 – 01:26 am BST

PDB ID : 1MSW  
Title : Structural basis for the transition from initiation to elongation transcription  
in T7 RNA polymerase  
Authors : Yin, Y.W.; Steitz, T.A.  
Deposited on : 2002-09-19  
Resolution : 2.10 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  
Xtrriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

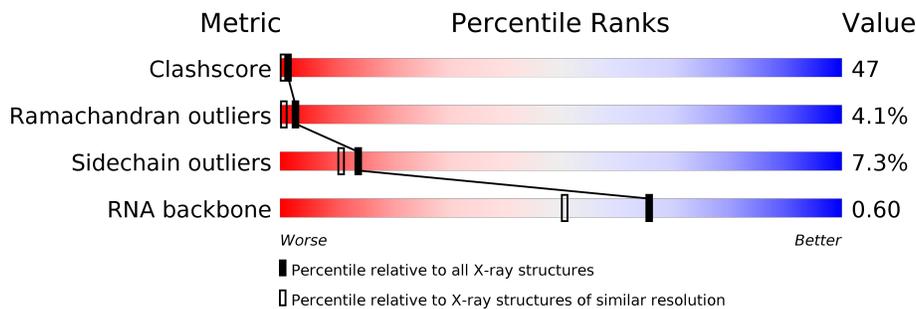
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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))
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RNA backbone	3102	1000 (2.54-1.66)

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  $\leq 5\%$ .

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	T	20	5% 60% 30% 5%
2	N	17	18% 82%
3	R	10	20% 80%
4	D	883	44% 46% 7% .

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 7956 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 DNA chain called Template DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	T	20	405	193	71	122	19	0	0	0

- Molecule 2 is a DNA chain called Non-Template DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	N	17	344	165	60	103	16	0	0	0

- Molecule 3 is a RNA chain called RNA message.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	R	10	215	97	44	65	9	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	863	6802	4333	1182	1251	36	0	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	T	13	Total	O	0	0
			13	13		
5	N	2	Total	O	0	0
			2	2		
5	R	9	Total	O	0	0
			9	9		
5	D	166	Total	O	0	0
			166	166		

### 3 Residue-property plots

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. 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.

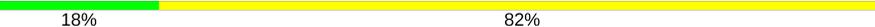
Note EDS was not executed.

- Molecule 1: Template DNA

Chain T: 



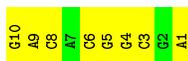
- Molecule 2: Non-Template DNA

Chain N: 



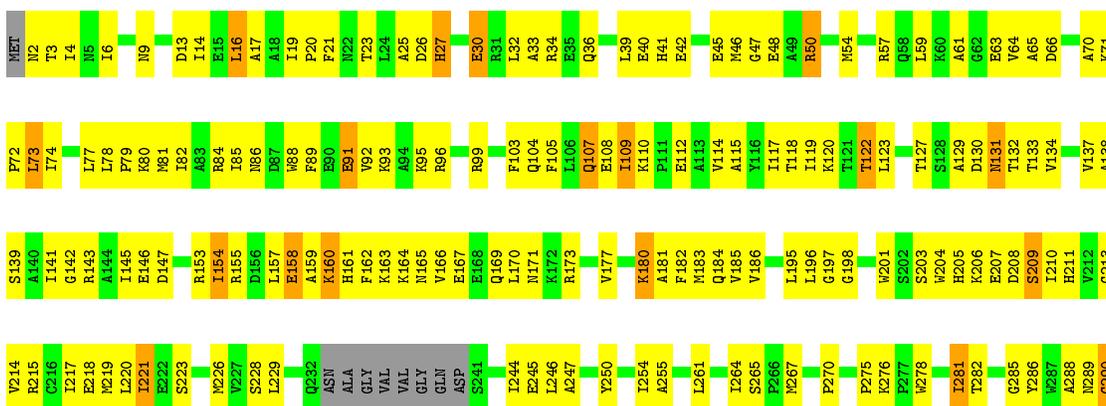
- Molecule 3: RNA message

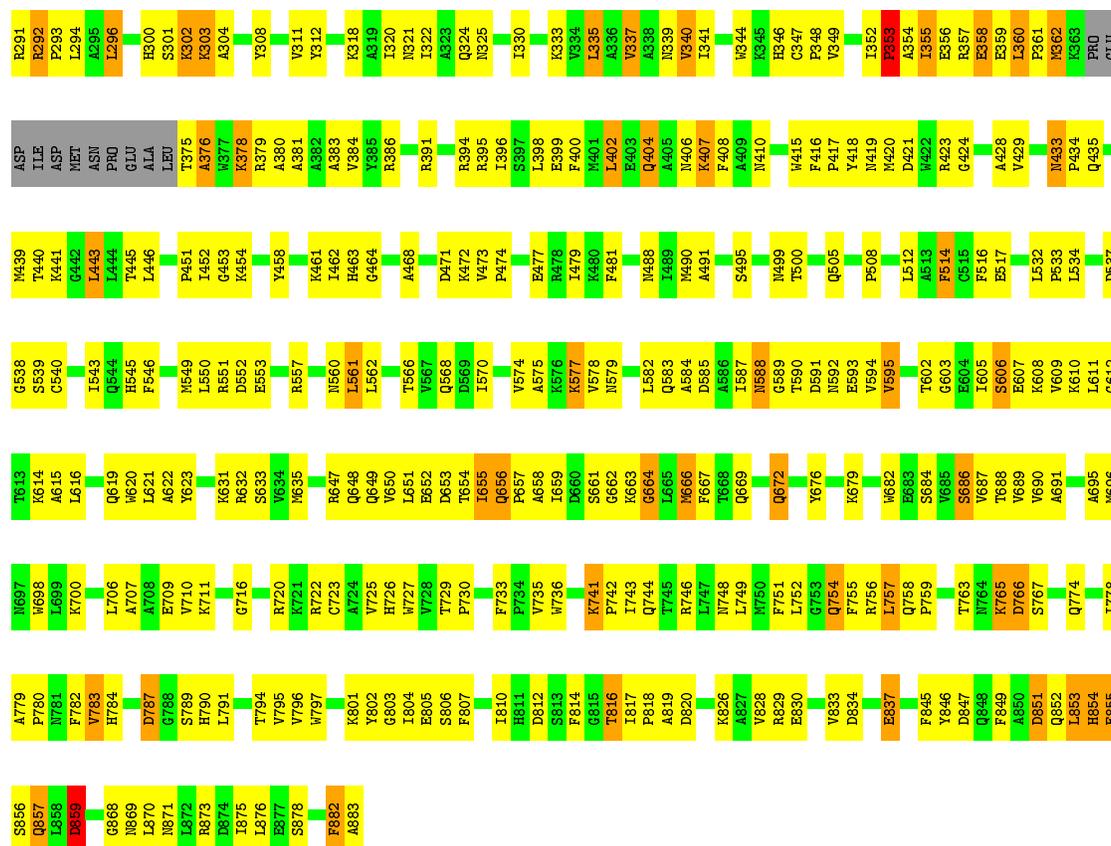
Chain R: 



- Molecule 4: DNA-directed RNA polymerase

Chain D: 





## 4 Data and refinement statistics

Xtrriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	142.91Å 145.46Å 145.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.10	Depositor
% Data completeness (in resolution range)	85.0 (20.00-2.10)	Depositor
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	CNS	Depositor
R, $R_{free}$	0.241 , 0.275	Depositor
Estimated twinning fraction	No twinning to report.	Xtrriage
Total number of atoms	7956	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	75.0	wwPDB-VP

## 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	T	0.46	0/452	0.90	2/696 (0.3%)
2	N	0.31	0/384	0.67	0/591
3	R	0.47	0/241	0.80	0/375
4	D	0.48	1/6955 (0.0%)	0.65	0/9407
All	All	0.47	1/8032 (0.0%)	0.67	2/11069 (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
1	T	0	7

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	407	LYS	CD-CE	5.13	1.64	1.51

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	T	119	DT	N1-C1'-C2'	-5.28	102.57	112.60
1	T	119	DT	C5'-C4'-C3'	-5.10	104.92	114.10

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	T	116	DC	Sidechain
1	T	117	DG	Sidechain
1	T	118	DC	Sidechain

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	T	119	DT	Sidechain
1	T	121	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	T	405	0	227	49	0
2	N	344	0	194	17	0
3	R	215	0	112	15	0
4	D	6802	0	6770	645	0
5	D	166	0	0	246	0
5	N	2	0	0	0	0
5	R	9	0	0	8	0
5	T	13	0	0	27	0
All	All	7956	0	7303	715	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 47.

The worst 5 of 715 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)
4:D:360:LEU:HB2	5:D:968:HOH:O	1.29	1.29
4:D:104:GLN:HB3	5:D:979:HOH:O	1.23	1.27
4:D:276:LYS:HE2	5:D:1032:HOH:O	1.22	1.26
4:D:217:ILE:O	4:D:221:ILE:HD13	1.24	1.24
4:D:428:ALA:HA	5:D:944:HOH:O	1.39	1.22

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
4	D	857/883 (97%)	736 (86%)	86 (10%)	35 (4%)	<b>3</b> <b>1</b>

5 of 35 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	16	LEU
4	D	158	GLU
4	D	302	LYS
4	D	303	LYS
4	D	353	PRO

### 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
4	D	711/729 (98%)	659 (93%)	52 (7%)	<b>14</b> <b>11</b>

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	358	GLU
4	D	443	LEU
4	D	847	ASP
4	D	360	LEU
4	D	378	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such

sidechains are listed below:

Mol	Chain	Res	Type
4	D	568	GLN
4	D	697	ASN
4	D	848	GLN
4	D	649	GLN
4	D	656	GLN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
3	R	9/10 (90%)	1 (11%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
3	R	8	C

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 carbohydrates 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

### 6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.