



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 05:38 am BST

PDB ID : 5W35  
Title : Crystal structure of the RNA polymerase domain (RPD) of Mycobacterium tuberculosis primase DnaG in complex with a double-stranded DNA oligomer with a 1-nucleotide overhang  
Authors : Hou, C.; Tsodikov, O.V.  
Deposited on : 2017-06-07  
Resolution : 3.31 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

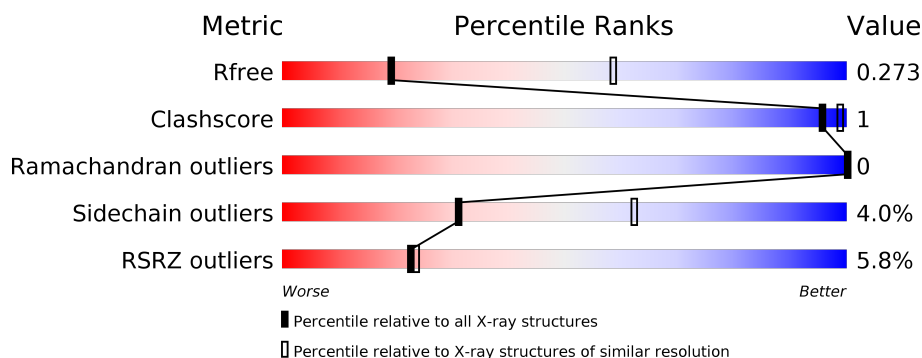
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.31 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	1089 (3.36-3.28)
Clashscore	141614	1137 (3.36-3.28)
Ramachandran outliers	138981	1115 (3.36-3.28)
Sidechain outliers	138945	1114 (3.36-3.28)
RSRZ outliers	127900	1059 (3.36-3.28)

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  $\leq 5\%$ . 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	325	<div> <div>6%</div> <div>94%</div> <div>5%</div> </div>
1	B	325	<div> <div>5%</div> <div>95%</div> <div>•</div> </div>
2	C	12	<div> <div>17%</div> <div>25%</div> <div>75%</div> </div>
3	D	13	<div> <div>15%</div> <div>23%</div> <div>62%</div> </div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 5010 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 primase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	321	Total	C	N	O	S	0	0	0
			2462	1546	450	449	17			
1	B	322	Total	C	N	O	S	0	0	0
			2470	1550	451	452	17			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	108	GLY	-	expression tag	UNP P9WNW1
A	109	PRO	-	expression tag	UNP P9WNW1
A	110	HIS	-	expression tag	UNP P9WNW1
A	111	MET	-	expression tag	UNP P9WNW1
B	108	GLY	-	expression tag	UNP P9WNW1
B	109	PRO	-	expression tag	UNP P9WNW1
B	110	HIS	-	expression tag	UNP P9WNW1
B	111	MET	-	expression tag	UNP P9WNW1

- Molecule 2 is a DNA chain called synthetic DNA oligomer 5'-CCACTTCCGGTC.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	3	Total	C	O	P	0	0	0
			22	10	10	2			

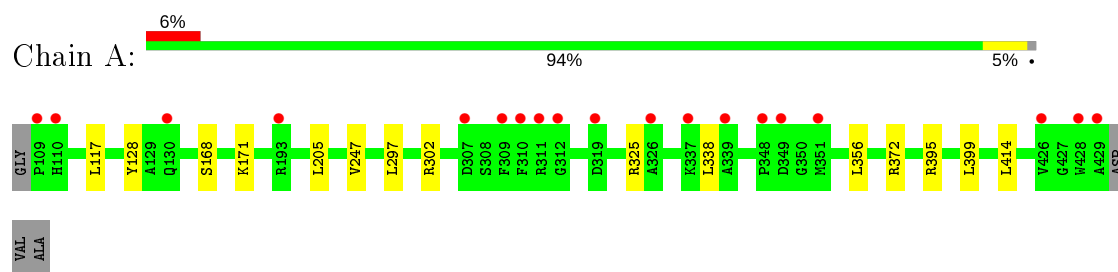
- Molecule 3 is a DNA chain called synthetic DNA oligomer 5'-TGACCGGAAGTGG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	5	Total	C	N	O	P	0	0	0
			56	24	3	24	5			

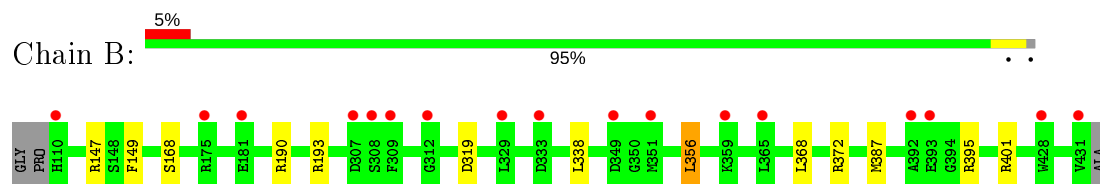
### 3 Residue-property plots [i](#)

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 and electron density. 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. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). 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 primase



- Molecule 1: DNA primase



- Molecule 2: synthetic DNA oligomer 5'-CCACTTCCGGTC



- Molecule 3: synthetic DNA oligomer 5'-TGACCGGAAGTGG



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	43.99 Å 210.41 Å 47.75 Å 90.00° 108.60° 90.00°	Depositor
Resolution (Å)	35.00 – 3.31 36.57 – 3.31	Depositor EDS
% Data completeness (in resolution range)	99.2 (35.00-3.31) 99.3 (36.57-3.31)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.95 (at 3.32 Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, $R_{free}$	0.224 , 0.279 0.224 , 0.273	Depositor DCC
$R_{free}$ test set	620 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	111.6	Xtriage
Anisotropy	0.421	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 72.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5010	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	131.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.50% 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	0.38	0/2512	0.64	0/3384
1	B	0.38	0/2519	0.62	0/3394
2	C	0.30	0/23	0.87	0/34
3	D	0.47	0/60	1.00	0/89
All	All	0.38	0/5114	0.64	0/6901

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	2462	0	2421	2	0
1	B	2470	0	2426	2	4
2	C	22	0	10	0	0
3	D	56	0	29	7	4
All	All	5010	0	4886	11	4

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 1.

All (11) 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:D:4:DC:C2'	3:D:5:DG:H5''	1.71	1.21
3:D:4:DC:H2'	3:D:5:DG:H5''	1.42	0.97
3:D:4:DC:H2''	3:D:5:DG:H5''	1.55	0.84
3:D:4:DC:H2'	3:D:5:DG:C5'	2.10	0.81
3:D:4:DC:C2'	3:D:5:DG:C5'	2.58	0.78
3:D:4:DC:C3'	3:D:5:DG:H5''	2.35	0.55
3:D:4:DC:C3'	3:D:5:DG:C5'	2.91	0.47
1:A:297:LEU:HD21	1:A:338:LEU:HD23	1.96	0.46
1:B:356:LEU:HD12	1:B:368:LEU:HD22	2.01	0.42
1:A:128:TYR:CD1	1:A:205:LEU:HD12	2.55	0.41
1:B:147:ARG:HG3	1:B:149:PHE:CE2	2.57	0.40

All (4) 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)
1:B:190:ARG:NH1	3:D:3:DC:OP2[2_656]	1.60	0.60
1:B:190:ARG:NH2	3:D:3:DC:OP2[2_656]	1.78	0.42
1:B:190:ARG:CZ	3:D:3:DC:OP2[2_656]	1.88	0.32
1:B:193:ARG:N	3:D:4:DC:O2[2_656]	2.06	0.14

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	319/325 (98%)	307 (96%)	12 (4%)	0	100	100
1	B	320/325 (98%)	307 (96%)	13 (4%)	0	100	100
All	All	639/650 (98%)	614 (96%)	25 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 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	239/241 (99%)	228 (95%)	11 (5%)	27	60
1	B	240/241 (100%)	232 (97%)	8 (3%)	38	67
All	All	479/482 (99%)	460 (96%)	19 (4%)	31	62

All (19) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	117	LEU
1	A	168	SER
1	A	171	LYS
1	A	247	VAL
1	A	302	ARG
1	A	325	ARG
1	A	356	LEU
1	A	372	ARG
1	A	395	ARG
1	A	399	LEU
1	A	414	LEU
1	B	168	SER
1	B	319	ASP
1	B	338	LEU
1	B	356	LEU
1	B	372	ARG
1	B	387	MET
1	B	395	ARG
1	B	401	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

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 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

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	321/325 (98%)	0.26	19 (5%) 22 23	89, 126, 170, 209	0
1	B	322/325 (99%)	0.24	17 (5%) 26 26	92, 128, 173, 206	0
2	C	3/12 (25%)	2.13	2 (66%) 0 0	140, 140, 175, 179	0
3	D	5/13 (38%)	0.14	0 100 100	159, 165, 178, 178	0
All	All	651/675 (96%)	0.26	38 (5%) 23 24	89, 127, 175, 209	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	428	TRP	6.5
1	A	429	ALA	6.1
1	A	428	TRP	5.0
1	B	308	SER	5.0
1	A	109	PRO	4.7
1	B	431	VAL	4.4
1	B	309	PHE	3.6
1	B	349	ASP	3.6
1	A	312	GLY	3.3
1	A	110	HIS	3.1
1	A	309	PHE	3.1
1	B	393	GLU	3.0
1	A	130	GLN	3.0
1	B	307	ASP	3.0
1	A	337	LYS	2.7
1	B	329	LEU	2.6
1	B	392	ALA	2.6
1	B	110	HIS	2.6
2	C	20	DC	2.6
1	A	339	ALA	2.5
1	A	351	MET	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	193	ARG	2.5
2	C	19	DC	2.4
1	A	326	ALA	2.4
1	A	348	PRO	2.4
1	A	319	ASP	2.3
1	B	175	ARG	2.3
1	A	310	PHE	2.2
1	A	307	ASP	2.2
1	B	365	LEU	2.2
1	B	351	MET	2.2
1	B	359	LYS	2.1
1	A	311	ARG	2.1
1	B	333	ASP	2.1
1	A	426	VAL	2.1
1	A	349	ASP	2.1
1	B	181	GLU	2.0
1	B	312	GLY	2.0

## 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 carbohydrates 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.