



Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 08:35 am BST

PDB ID : 5W36
Title : Crystal structure of the RNA polymerase domain (RPD) of Mycobacterium tuberculosis primase DnaG in complex with a double-stranded DNA oligomer with a 6-nucleotide overhang
Authors : Hou, C.; Tsodikov, O.V.
Deposited on : 2017-06-07
Resolution : 2.46 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

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

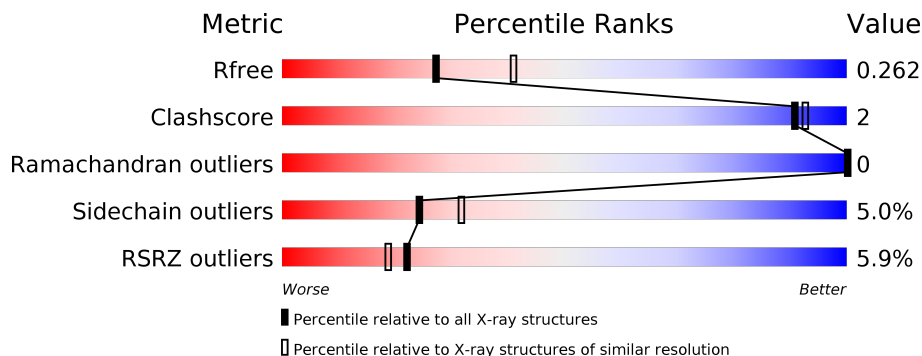
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.46 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	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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 ≥ 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	
1	B	325	
2	G	14	

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 5022 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
			Total	C	N	O	S			
1	A	321	2462	1546	450	449	17	0	0	0
1	B	322	2470	1550	451	452	17	0	0	0

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 5'-TATCGTCCCGCCTC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	G	2	42	20	7	13	2	0	0	0

- Molecule 3 is STRONTIUM ION (three-letter code: SR) (formula: Sr).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	G	1	Total	Sr	0	0
			1	1		

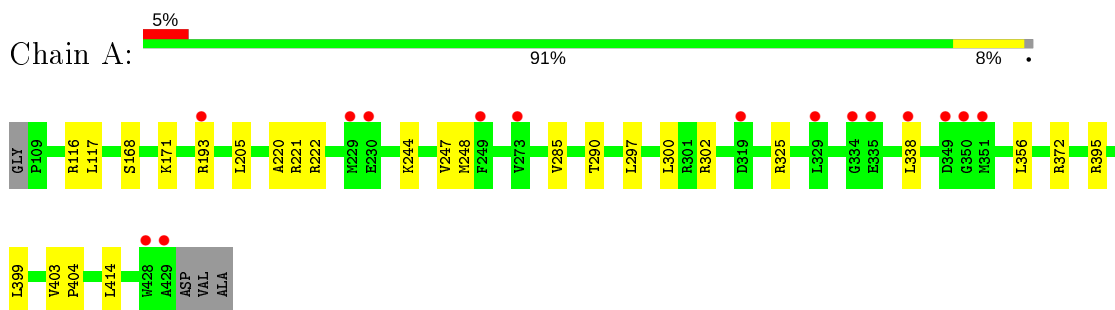
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	26	Total 26	O 26	0	0
4	B	21	Total 21	O 21	0	0

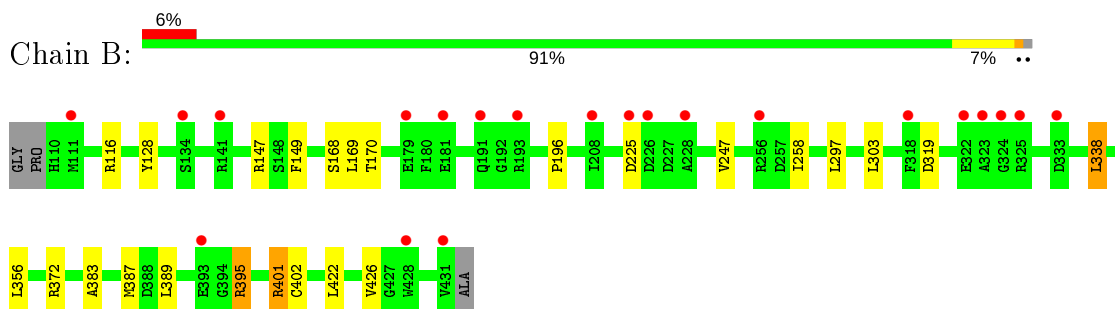
3 Residue-property plots [i](#)

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 5'-TATCGTCCCGCCTC



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	43.84Å 210.94Å 46.05Å 90.00° 106.78° 90.00°	Depositor
Resolution (Å)	35.00 – 2.46 34.10 – 2.46	Depositor EDS
% Data completeness (in resolution range)	95.7 (35.00-2.46) 95.8 (34.10-2.46)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.93 (at 2.45Å)	Xtrriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.215 , 0.269 0.216 , 0.262	Depositor DCC
R_{free} test set	1253 reflections (4.51%)	wwPDB-VP
Wilson B-factor (Å ²)	58.8	Xtrriage
Anisotropy	0.433	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 37.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.037 for l,-k,h	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	5022	wwPDB-VP
Average B, all atoms (Å ²)	69.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.38% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SR

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	0.43	0/2512	0.70	0/3384
1	B	0.42	0/2519	0.69	0/3394
2	G	0.40	0/46	0.87	0/69
All	All	0.42	0/5077	0.70	0/6847

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

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	7	0
1	B	2470	0	2426	10	0
2	G	42	0	24	0	0
3	G	1	0	0	0	0
4	A	26	0	0	0	0
4	B	21	0	0	0	0
All	All	5022	0	4871	17	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 2.

All (17) 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:B:395:ARG:HD2	1:B:426:VAL:O	2.06	0.56
1:B:389:LEU:O	1:B:395:ARG:NE	2.42	0.52
1:A:244:LYS:HD2	1:A:290:THR:HG21	1.93	0.50
1:B:297:LEU:HD13	1:B:338:LEU:HD23	1.94	0.49
1:B:170:THR:HG21	1:B:196:PRO:HB2	1.95	0.49
1:A:221:ARG:NH1	1:A:222:ARG:O	2.46	0.48
1:B:258:ILE:HD11	1:B:303:LEU:CD2	2.46	0.46
1:B:402:CYS:SG	1:B:422:LEU:HD22	2.57	0.45
1:A:205:LEU:HD22	1:A:220:ALA:HB2	1.98	0.44
1:A:297:LEU:HD21	1:A:338:LEU:HD23	1.99	0.44
1:A:403:VAL:HG22	1:A:404:PRO:HD3	1.99	0.44
1:B:258:ILE:HD11	1:B:303:LEU:HD23	2.00	0.42
1:A:248:MET:HE3	1:A:285:VAL:HG21	2.00	0.42
1:B:383:ALA:HB1	1:B:401:ARG:HD2	2.02	0.41
1:B:147:ARG:HG3	1:B:149:PHE:CE2	2.55	0.41
1:A:248:MET:HE1	1:A:300:LEU:HD21	2.03	0.41
1:B:128:TYR:OH	1:B:169:LEU:HD22	2.21	0.40

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	319/325 (98%)	312 (98%)	7 (2%)	0	100	100
1	B	320/325 (98%)	309 (97%)	11 (3%)	0	100	100
All	All	639/650 (98%)	621 (97%)	18 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	239/241 (99%)	226 (95%)	13 (5%)	22	28
1	B	240/241 (100%)	229 (95%)	11 (5%)	27	35
All	All	479/482 (99%)	455 (95%)	24 (5%)	24	32

All (24) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	116	ARG
1	A	117	LEU
1	A	168	SER
1	A	171	LYS
1	A	193	ARG
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	116	ARG
1	B	168	SER
1	B	225	ASP
1	B	247	VAL
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 [i](#)

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](#)

Of 1 ligands modelled in this entry, 1 is 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

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, 95th 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 ²)	Q<0.9
1	A	321/325 (98%)	0.34	15 (4%) 31 29	40, 63, 97, 133	0
1	B	322/325 (99%)	0.35	21 (6%) 18 15	45, 68, 102, 133	0
2	G	2/14 (14%)	2.04	2 (100%) 0 0	118, 118, 118, 118	0
All	All	645/664 (97%)	0.35	38 (5%) 22 19	40, 66, 102, 133	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	350	GLY	5.8
1	A	338	LEU	5.7
1	A	349	ASP	5.1
1	B	228	ALA	4.9
1	B	256	ARG	4.1
1	A	351	MET	4.0
1	A	428	TRP	3.9
1	A	319	ASP	3.7
1	B	431	VAL	3.5
1	A	193	ARG	3.3
1	B	225	ASP	3.3
1	A	229	MET	3.2
1	B	333	ASP	3.0
1	B	226	ASP	2.9
1	A	329	LEU	2.9
1	B	181	GLU	2.7
1	A	429	ALA	2.7
1	A	249	PHE	2.7
1	B	324	GLY	2.6
1	A	230	GLU	2.6
1	B	111	MET	2.5
1	B	322	GLU	2.4
1	B	208	ILE	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	179	GLU	2.4
1	B	141	ARG	2.4
1	B	325	ARG	2.3
1	B	393	GLU	2.2
1	A	334	GLY	2.2
1	B	134	SER	2.2
1	B	193	ARG	2.1
1	B	318	PHE	2.1
1	A	273	VAL	2.1
1	B	191	GLN	2.1
1	A	335	GLU	2.1
2	G	5	DG	2.1
1	B	323	ALA	2.0
1	B	428	TRP	2.0
2	G	6	DT	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](#)

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, 95th 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(Å ²)	Q<0.9
3	SR	G	101	1/1	0.76	0.22	157,157,157,157	0

6.5 Other polymers [i](#)

There are no such residues in this entry.