



# Full wwPDB X-ray Structure Validation Report ⓘ

May 19, 2020 – 02:10 am BST

PDB ID : 5HMX  
Title : Dengue serotype 3 RNA-dependent RNA polymerase bound to compound 10  
Authors : Noble, C.G.  
Deposited on : 2016-01-17  
Resolution : 2.40 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
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

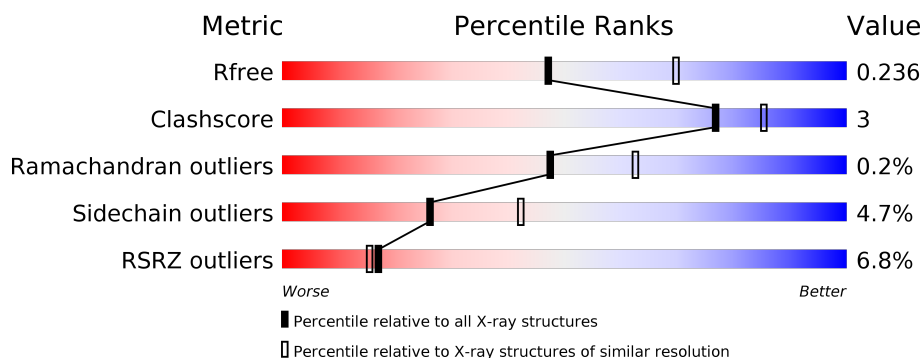
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	635	<div> <div>6%</div> <div>82%</div> <div>10%</div> <div>7%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5087 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 RNA-directed RNA polymerase NS5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	590	Total	C	N	O	S	0	2	0
			4815	3039	865	879	32			

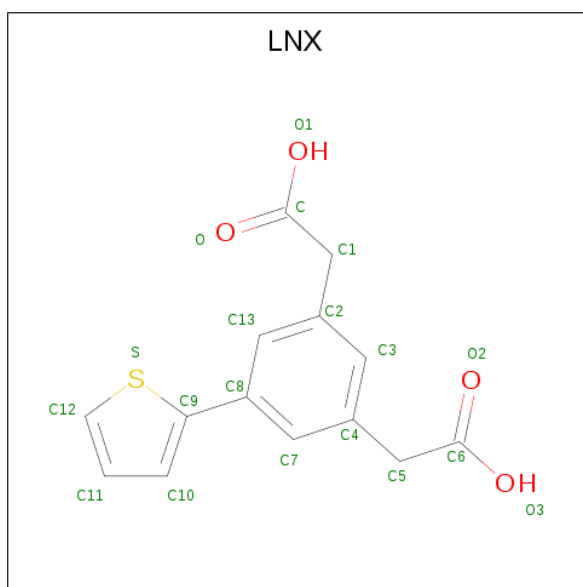
There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	266	GLY	-	expression tag	UNP Q6DLV0
A	267	SER	-	expression tag	UNP Q6DLV0
A	268	HIS	-	expression tag	UNP Q6DLV0
A	269	MET	-	expression tag	UNP Q6DLV0
A	270	LEU	-	expression tag	UNP Q6DLV0
A	271	ASP	-	expression tag	UNP Q6DLV0
A	374	GLU	GLY	variant	UNP Q6DLV0

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Zn	0	0
			2	2		

- Molecule 3 is 2,2'-(5-(thiophen-2-yl)-1,3-phenylene)diacetic acid (three-letter code: LNX) (formula: C<sub>14</sub>H<sub>12</sub>O<sub>4</sub>S).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	O	S	0	0
			19	14	4	1		
3	A	1	Total	C	O	S	0	0
			19	14	4	1		

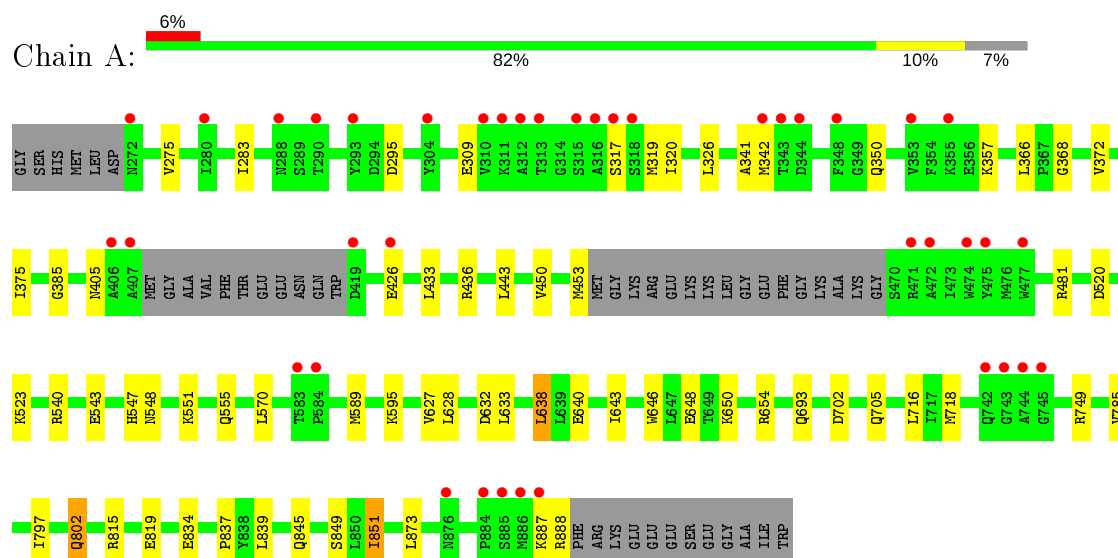
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	232	Total	O	0	0
			232	232		

### 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: RNA-directed RNA polymerase NS5



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	161.34Å 176.30Å 57.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	51.44 – 2.40 51.44 – 2.40	Depositor EDS
% Data completeness (in resolution range)	98.6 (51.44-2.40) 98.6 (51.44-2.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.15 (at 2.39Å)	Xtriage
Refinement program	BUSTER 2.11.5	Depositor
R, $R_{free}$	0.186 , 0.230 0.193 , 0.236	Depositor DCC
$R_{free}$ test set	1638 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	35.3	Xtriage
Anisotropy	0.523	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 56.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5087	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.66% 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

### 5.1 Standard geometry

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

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	0.49	0/4943	0.67	0/6693

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	4815	0	4712	29	0
2	A	2	0	0	0	0
3	A	38	0	0	0	0
4	A	232	0	0	0	0
All	All	5087	0	4712	29	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 3.

All (29) 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:A:802:GLN:H	1:A:802:GLN:HE21	1.37	0.73
1:A:372:VAL:HG11	1:A:628:LEU:HD11	1.79	0.64
1:A:275:VAL:HG13	1:A:570:LEU:HD23	1.81	0.63
1:A:385:GLY:HA3	1:A:555:GLN:HE22	1.67	0.59
1:A:375:ILE:HD11	1:A:640:GLU:HG2	1.84	0.59
1:A:845:GLN:HG3	1:A:851:ILE:HG22	1.89	0.54
1:A:716:LEU:HD21	1:A:839:LEU:HD23	1.90	0.54
1:A:368:GLY:O	1:A:372:VAL:HG23	2.10	0.51
1:A:627:VAL:HG21	1:A:646:TRP:CD1	2.45	0.51
1:A:548:ASN:HD22	1:A:551:LYS:HE2	1.78	0.48
1:A:887:LYS:HE2	1:A:887:LYS:H	1.78	0.48
1:A:433:LEU:HD23	1:A:436:ARG:HH21	1.79	0.47
1:A:632:ASP:HB3	1:A:638:LEU:HD13	1.97	0.47
1:A:543:GLU:O	1:A:547[A]:HIS:HD2	1.99	0.46
1:A:638:LEU:HD11	1:A:643:ILE:HD11	1.96	0.46
1:A:372:VAL:HG21	1:A:628:LEU:HD21	1.98	0.46
1:A:646:TRP:CZ2	1:A:654:ARG:HG3	2.51	0.46
1:A:815:ARG:HA	1:A:819:GLU:HB2	1.97	0.45
1:A:320:ILE:HD11	1:A:341:ALA:HB1	1.99	0.44
1:A:718:MET:HG2	1:A:837:PRO:HD3	2.00	0.44
1:A:873:LEU:HA	1:A:873:LEU:HD23	1.85	0.44
1:A:520:ASP:HA	1:A:523:LYS:HE2	1.99	0.44
1:A:702[B]:ASP:HB2	1:A:705:GLN:HG2	2.00	0.43
1:A:849:SER:OG	1:A:851:ILE:HB	2.19	0.43
1:A:283:ILE:HD11	1:A:450:VAL:HG21	2.01	0.41
1:A:366:LEU:HB3	1:A:633:LEU:O	2.21	0.41
1:A:540:ARG:HA	1:A:540:ARG:HD2	1.97	0.41
1:A:702[A]:ASP:HB3	1:A:705:GLN:HG2	2.02	0.41
1:A:650:LYS:O	1:A:654:ARG:HG2	2.22	0.41

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	586/635 (92%)	565 (96%)	20 (3%)	1 (0%)	47 62

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	342	MET

### 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	517/552 (94%)	493 (95%)	24 (5%)	27 43

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

Mol	Chain	Res	Type
1	A	295	ASP
1	A	309	GLU
1	A	317	SER
1	A	319	MET
1	A	326	LEU
1	A	350	GLN
1	A	357	LYS
1	A	405	ASN
1	A	426	GLU
1	A	443	LEU
1	A	453	MET
1	A	481	ARG
1	A	589	MET
1	A	595	LYS
1	A	638	LEU
1	A	648	GLU
1	A	693	GLN
1	A	749	ARG
1	A	785	VAL
1	A	797	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	802	GLN
1	A	834	GLU
1	A	851	ILE
1	A	888	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (13) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	272	ASN
1	A	339	GLN
1	A	452	ASN
1	A	548	ASN
1	A	555	GLN
1	A	562	GLN
1	A	676	ASN
1	A	682	ASN
1	A	704	GLN
1	A	760	GLN
1	A	768	HIS
1	A	802	GLN
1	A	835	ASN

### 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 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	LNx	A	1004	-	13,20,20	0.45	0	14,27,27	0.33	0
3	LNx	A	1003	-	13,20,20	0.36	0	14,27,27	0.51	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	LNx	A	1004	-	-	0/8/12/12	0/2/2/2
3	LNx	A	1003	-	-	0/8/12/12	0/2/2/2

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, 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	590/635 (92%)	-0.00	40 (6%) 17 15	22, 41, 90, 115	0

All (40) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	407	ALA	8.6
1	A	353	VAL	6.9
1	A	312	ALA	5.9
1	A	310	VAL	4.8
1	A	744	ALA	4.5
1	A	743	GLY	4.5
1	A	406	ALA	4.5
1	A	348	PHE	4.3
1	A	315	SER	4.2
1	A	343	THR	4.1
1	A	355	LYS	4.1
1	A	344	ASP	3.8
1	A	318	SER	3.6
1	A	313	THR	3.6
1	A	290	THR	3.6
1	A	317	SER	3.3
1	A	745	GLY	3.3
1	A	304	TYR	3.2
1	A	471	ARG	3.2
1	A	293	TYR	3.2
1	A	316	ALA	3.0
1	A	886	MET	3.0
1	A	288	ASN	3.0
1	A	472	ALA	2.7
1	A	584	PRO	2.6
1	A	475	TYR	2.4
1	A	742	GLN	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	885	SER	2.3
1	A	477	TRP	2.2
1	A	887	LYS	2.2
1	A	342	MET	2.2
1	A	474	TRP	2.2
1	A	884	PRO	2.2
1	A	876	ASN	2.1
1	A	583	THR	2.1
1	A	311	LYS	2.1
1	A	419	ASP	2.1
1	A	272	ASN	2.1
1	A	426	GLU	2.1
1	A	280	ILE	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, 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
3	LNX	A	1004	19/19	0.83	0.24	80,82,87,88	0
3	LNX	A	1003	19/19	0.95	0.14	33,40,54,56	0
2	ZN	A	1001	1/1	1.00	0.09	45,45,45,45	0
2	ZN	A	1002	1/1	1.00	0.10	45,45,45,45	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.