



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 17, 2022 – 10:13 am BST

PDB ID : 7PFY  
Title : Crystal Structure of Unlinked NS2B-NS3 Protease from Zika Virus in Complex with Inhibitor MI-2241  
Authors : Huber, S.; Heine, A.; Steinmetzer, T.  
Deposited on : 2021-08-12  
Resolution : 1.38 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 types of validation reports are described at <http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : **FAILED**  
Mogul : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.29  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0267  
CCP4 : 7.1.010 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.29

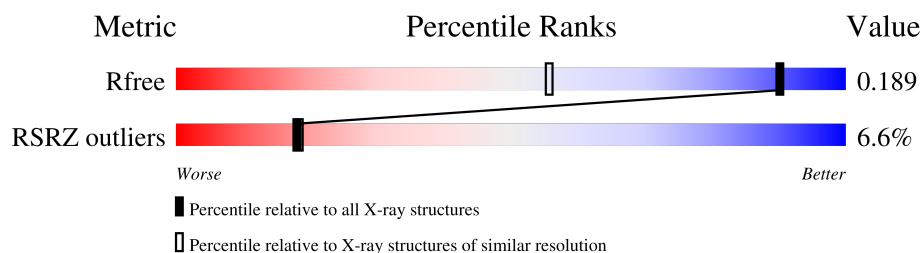
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.38 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	2907 (1.40-1.36)
RSRZ outliers	127900	2846 (1.40-1.36)

MolProbity failed to run properly - the sequence quality summary graphics cannot be shown.

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3138 atoms, of which 1409 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 Serine protease subunit NS2B.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	40	Total	C	H	N	O	S	0	2	0
			567	186	266	49	65	1			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	44	MET	-	initiating methionine	UNP Q32ZE1
A	45	THR	-	expression tag	UNP Q32ZE1

- Molecule 2 is a protein called Serine protease NS3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	153	Total	C	H	N	O	S	0	16	0
			2317	742	1143	202	224	6			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	GLY	-	expression tag	UNP Q32ZE1
B	107	LYS	ARG	conflict	UNP Q32ZE1

- Molecule 3 is a protein called Inhibitor MI-2241.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	D	6	Total	C	N	O	0	0	0
			46	29	11	6			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	42	Total	O	0	1
			43	43		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	148	Total 148	O 148	0	0
4	D	17	Total 17	O 17	0	0

MolProbity failed to run properly - this section is therefore empty.

### 3 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	47.27Å 60.92Å 83.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.34 – 1.38 37.34 – 1.38	Depositor EDS
% Data completeness (in resolution range)	99.0 (37.34-1.38) 99.0 (37.34-1.38)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.40 (at 1.38Å)	Xtrriage
Refinement program	PHENIX 1.16_3549	Depositor
R, $R_{free}$	0.167 , 0.189 0.167 , 0.189	Depositor DCC
$R_{free}$ test set	2482 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.5	Xtrriage
Anisotropy	0.632	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3138	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.94% 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.

## 4 Model quality [i](#)

### 4.1 Standard geometry [i](#)

MolProbity failed to run properly - this section is therefore empty.

### 4.2 Too-close contacts [i](#)

MolProbity failed to run properly - this section is therefore empty.

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

#### 4.3.1 Protein backbone [i](#)

MolProbity failed to run properly - this section is therefore empty.

#### 4.3.2 Protein sidechains [i](#)

MolProbity failed to run properly - this section is therefore empty.

#### 4.3.3 RNA [i](#)

MolProbity failed to run properly - this section is therefore empty.

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

2 non-standard protein/DNA/RNA residues are modelled in this entry.

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	BAL	D	3	3	4,4,5	0.38	0	3,3,5	0.59	0
3	V7T	D	1	3	10,11,12	1.20	1 (10%)	9,12,14	1.52	2 (22%)

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	BAL	D	3	3	-	0/1/2/3	-
3	V7T	D	1	3	-	0/10/11/13	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	1	V7T	C2-N	3.30	1.38	1.33

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	1	V7T	N2-C2-N	-2.98	112.88	119.55
3	D	1	V7T	C-CA-N	2.15	113.62	109.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 4.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 4.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 4.7 Other polymers [i](#)

There are no such residues in this entry.

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

There are no chain breaks in this entry.

## 5 Fit of model and data [i](#)

### 5.1 Protein, DNA and RNA chains [i](#)

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	40/53 (75%)	0.15	5 (12%) <b>3</b> <b>4</b>	20, 25, 45, 55	0
2	B	153/178 (85%)	0.45	8 (5%) <b>27</b> <b>28</b>	14, 21, 37, 62	0
3	D	4/6 (66%)	-0.08	0 <b>100</b> <b>100</b>	20, 22, 23, 26	0
All	All	197/237 (83%)	0.38	13 (6%) <b>18</b> <b>18</b>	14, 22, 37, 62	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	30	LEU	11.6
2	B	31	LEU	7.2
2	B	29	ARG	5.3
2	B	18[A]	THR	3.8
2	B	63[A]	GLY	3.4
1	A	49	VAL	2.7
1	A	88	GLU	2.4
2	B	170	ARG	2.4
2	B	19	THR	2.4
2	B	32	GLY	2.3
1	A	65	ALA	2.3
1	A	50	ASP	2.1
1	A	67	VAL	2.1

### 5.2 Non-standard residues in protein, DNA, RNA chains [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	BAL	D	3	5/6	0.93	0.09	21,22,24,25	0
3	V7T	D	1	12/13	0.98	0.09	18,19,21,21	0

### 5.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

### 5.4 Ligands [i](#)

There are no ligands in this entry.

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

There are no such residues in this entry.