



# Full wwPDB X-ray Structure Validation Report ⓘ

Oct 16, 2017 – 10:35 PM EDT

PDB ID : 5TFN  
Title : CRYSTAL STRUCTURE OF THE ZIKA VIRUS NS2B-NS3 PROTEASE in super-open conformation  
Authors : Aleshin, A.E.; Bankston, L.; Liddington, R.C.  
Deposited on : unknown  
Resolution : 3.00 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

<http://wwpdb.org/validation/2016/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)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

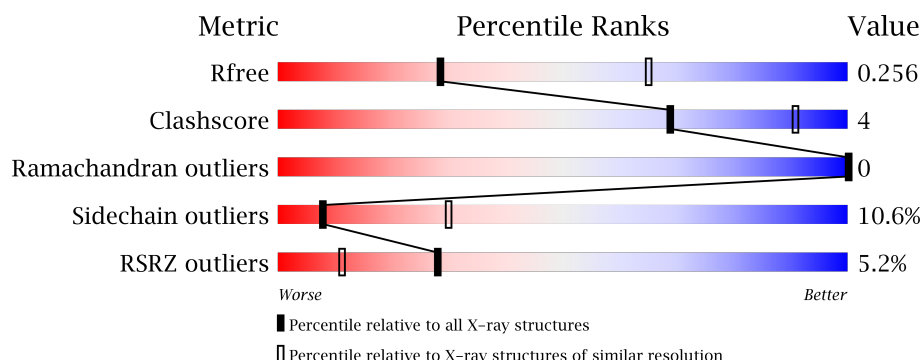
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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))
$R_{free}$	100719	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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. 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	241	<div> <div>2%</div> <div>54%</div> <div>14%</div> <div>32%</div> </div>
1	B	241	<div> <div>5%</div> <div>54%</div> <div>11%</div> <div>34%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2475 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 NS2B-NS3 Protease CHIMERA,Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	165	Total	C	N	O	S	0	0	0
			1249	785	219	239	6			
1	B	160	Total	C	N	O	S	0	0	0
			1220	770	213	231	6			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	46	GLY	-	expression tag	UNP A0A160JCU6
A	47	SER	-	expression tag	UNP A0A160JCU6
A	95	ALA	-	linker	UNP A0A160JCU6
A	96	GLY	-	linker	UNP A0A160JCU6
A	97	GLY	-	linker	UNP A0A160JCU6
A	98	GLY	-	linker	UNP A0A160JCU6
A	99	GLY	-	linker	UNP A0A160JCU6
A	100	SER	-	linker	UNP A0A160JCU6
A	101	GLY	-	linker	UNP A0A160JCU6
A	102	GLY	-	linker	UNP A0A160JCU6
A	103	GLY	-	linker	UNP A0A160JCU6
A	104	GLY	-	linker	UNP A0A160JCU6
A	1056	SER	ALA	engineered mutation	UNP Q32ZE1
A	1092	HIS	LEU	conflict	UNP Q32ZE1
A	1135	ALA	SER	conflict	UNP Q32ZE1
A	1169	ARG	LYS	conflict	UNP Q32ZE1
B	46	GLY	-	expression tag	UNP A0A160JCU6
B	47	SER	-	expression tag	UNP A0A160JCU6
B	95	ALA	-	linker	UNP A0A160JCU6
B	96	GLY	-	linker	UNP A0A160JCU6
B	97	GLY	-	linker	UNP A0A160JCU6
B	98	GLY	-	linker	UNP A0A160JCU6
B	99	GLY	-	linker	UNP A0A160JCU6
B	100	SER	-	linker	UNP A0A160JCU6
B	101	GLY	-	linker	UNP A0A160JCU6

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	102	GLY	-	linker	UNP A0A160JCU6
B	103	GLY	-	linker	UNP A0A160JCU6
B	104	GLY	-	linker	UNP A0A160JCU6
B	1056	SER	ALA	engineered mutation	UNP Q32ZE1
B	1092	HIS	LEU	conflict	UNP Q32ZE1
B	1135	ALA	SER	conflict	UNP Q32ZE1
B	1169	ARG	LYS	conflict	UNP Q32ZE1

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	4	Total O 4 4	0	0
2	B	2	Total O 2 2	0	0

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.

- [illegible]

- Chain B:
- 
- | Position | Amino Acid | Information Content (bits) |
|----------|------------|----------------------------|
| 1        | GLY        | 1.8                        |
| 2        | SER        | 1.7                        |
| 3        | SER        | 1.6                        |
| 4        | SER        | 1.5                        |
| 5        | VAL        | 1.4                        |
| 6        | GLY        | 1.3                        |
| 7        | GLY        | 1.2                        |
| 8        | GLY        | 1.1                        |
| 9        | GLY        | 1.0                        |
| 10       | GLY        | 0.9                        |
| 11       | GLY        | 0.8                        |
| 12       | GLY        | 0.7                        |
| 13       | GLY        | 0.6                        |
| 14       | GLY        | 0.5                        |
| 15       | GLY        | 0.4                        |
| 16       | GLY        | 0.3                        |
| 17       | GLY        | 0.2                        |
| 18       | GLY        | 0.1                        |
| 19       | GLY        | 0.0                        |
| 20       | GLY        | 0.0                        |
| 21       | GLY        | 0.0                        |
| 22       | GLY        | 0.0                        |
| 23       | GLY        | 0.0                        |
| 24       | GLY        | 0.0                        |
| 25       | GLY        | 0.0                        |
| 26       | GLY        | 0.0                        |
| 27       | GLY        | 0.0                        |
| 28       | GLY        | 0.0                        |
| 29       | GLY        | 0.0                        |
| 30       | GLY        | 0.0                        |
| 31       | GLY        | 0.0                        |
| 32       | GLY        | 0.0                        |
| 33       | GLY        | 0.0                        |
| 34       | GLY        | 0.0                        |
| 35       | GLY        | 0.0                        |
| 36       | GLY        | 0.0                        |
| 37       | GLY        | 0.0                        |
| 38       | GLY        | 0.0                        |
| 39       | GLY        | 0.0                        |
| 40       | GLY        | 0.0                        |

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	54.96 Å 54.96 Å 250.11 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.40 – 3.00 38.40 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.7 (38.40-3.00) 97.4 (38.40-2.70)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.00 (at 2.69 Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, $R_{free}$	0.210 , 0.259 0.215 , 0.256	Depositor DCC
$R_{free}$ test set	436 reflections (5.51%)	DCC
Wilson B-factor (Å <sup>2</sup> )	76.2	Xtriage
Anisotropy	0.446	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 80.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2475	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	112.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 13.54% 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	0.81	0/1275	1.01	3/1728 (0.2%)
1	B	0.77	0/1246	0.99	6/1689 (0.4%)
All	All	0.79	0/2521	1.00	9/3417 (0.3%)

There are no bond length outliers.

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1049	MET	CG-SD-CE	7.80	112.68	100.20
1	B	1049	MET	CG-SD-CE	6.92	111.27	100.20
1	B	1031	LEU	CB-CG-CD1	6.76	122.49	111.00
1	B	1030	LEU	CB-CG-CD1	6.21	121.56	111.00
1	B	1064	ARG	NE-CZ-NH2	5.70	123.15	120.30
1	B	1019	THR	N-CA-CB	5.46	120.67	110.30
1	A	1064	ARG	NE-CZ-NH2	-5.17	117.72	120.30
1	B	1024	ARG	NE-CZ-NH2	5.04	122.82	120.30
1	A	68	THR	N-CA-C	5.01	124.53	111.00

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	1249	0	1221	12	2
1	B	1220	0	1197	10	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	4	0	0	1	0
2	B	2	0	0	0	0
All	All	2475	0	2418	21	2

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 4.

All (21) 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:1024:ARG:NH1	1:B:1104:GLU:O	2.33	0.60
1:A:1024:ARG:NH1	1:A:1104:GLU:O	2.36	0.55
1:B:1065:LEU:HD22	1:B:1080:CYS:SG	2.49	0.52
1:B:1029:ARG:HG2	1:B:1030:LEU:N	2.26	0.51
1:A:1045:VAL:HG22	1:A:1081:GLY:HA3	1.93	0.49
1:A:1065:LEU:HD22	1:A:1080:CYS:SG	2.53	0.49
1:B:1045:VAL:HG22	1:B:1081:GLY:HA3	1.96	0.48
1:A:1031:LEU:H	1:A:1031:LEU:HD12	1.79	0.47
1:A:1024:ARG:NH2	1:A:1035:GLN:OE1	2.44	0.47
1:A:1039:GLY:HA3	1:A:1046:PHE:CZ	2.50	0.46
1:B:1039:GLY:HA3	1:B:1046:PHE:CZ	2.53	0.44
1:B:1024:ARG:NH2	1:B:1035:GLN:OE1	2.47	0.44
1:A:1090:ASP:HB2	1:A:1092:HIS:CE1	2.53	0.43
1:B:1029:ARG:CG	1:B:1030:LEU:N	2.82	0.43
1:B:1045:VAL:HG21	1:B:1084:LYS:HE2	2.00	0.43
1:A:1157:LYS:HG2	1:A:1160:SER:HA	2.02	0.42
1:B:61:TRP:CE3	1:B:1141:ASP:O	2.72	0.42
1:A:1045:VAL:HG21	1:A:1084:LYS:HE2	2.00	0.42
1:A:1133:GLY:N	2:A:301:HOH:O	2.48	0.41
1:A:1031:LEU:HD13	1:B:1151:GLY:O	2.22	0.41
1:A:61:TRP:CE3	1:A:1142:LYS:HA	2.56	0.40

All (2) 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:A:69:GLY:O	1:B:49:VAL:O[6_554]	2.12	0.08
1:A:69:GLY:C	1:B:49:VAL:O[6_554]	2.13	0.07



## 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	161/241 (67%)	153 (95%)	8 (5%)	0	100	100
1	B	156/241 (65%)	151 (97%)	5 (3%)	0	100	100
All	All	317/482 (66%)	304 (96%)	13 (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	129/187 (69%)	114 (88%)	15 (12%)	6	26
1	B	126/187 (67%)	114 (90%)	12 (10%)	10	36
All	All	255/374 (68%)	228 (89%)	27 (11%)	8	30

All (27) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	63	LYS
1	A	66	GLU
1	A	67	VAL
1	A	68	THR
1	A	1019	THR
1	A	1030	LEU
1	A	1042	GLN
1	A	1051	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	1098	LEU
1	A	1109	ILE
1	A	1111	THR
1	A	1115	ILE
1	A	1116	PHE
1	A	1120	ASP
1	A	1160	SER
1	B	63	LYS
1	B	1019	THR
1	B	1030	LEU
1	B	1042	GLN
1	B	1059	ARG
1	B	1098	LEU
1	B	1109	ILE
1	B	1111	THR
1	B	1115	ILE
1	B	1116	PHE
1	B	1120	ASP
1	B	1142	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	1051	HIS

### 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

There are no ligands in this entry.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

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

### 6.1 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	165/241 (68%)	-0.19	4 (2%) 59 30	69, 91, 151, 175	0
1	B	160/241 (66%)	0.31	13 (8%) 13 5	82, 118, 177, 197	0
All	All	325/482 (67%)	0.06	17 (5%) 28 11	69, 104, 167, 197	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1016	GLY	5.7
1	B	1122	ASP	5.0
1	A	1160	SER	3.6
1	B	1115	ILE	3.4
1	B	1119	LYS	3.2
1	B	1017	GLU	3.1
1	B	66	GLU	3.1
1	B	1117	LYS	3.0
1	B	1062	GLU	3.0
1	B	1116	PHE	2.9
1	B	1156	ILE	2.4
1	B	1127	ALA	2.4
1	A	1158	ASN	2.4
1	B	1018	THR	2.3
1	B	1128	LEU	2.2
1	B	1016	GLY	2.1
1	A	64	ASP	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.