



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 28, 2016 – 04:28 PM EST

PDB ID : 5GJ4  
Title : Structure of NS2B-NS3 Protease from Zika Virus caught after self-cleavage  
Authors : Phoo, W.W.; Li, Y.  
Deposited on : 2016-06-27  
Resolution : 1.84 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20028320  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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-20028320

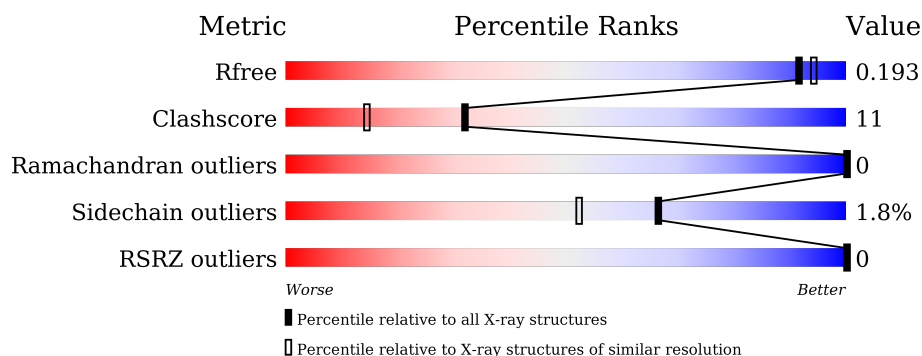
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.84 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))
$R_{free}$	91344	2634 (1.86-1.82)
Clashscore	102246	2862 (1.86-1.82)
Ramachandran outliers	100387	2831 (1.86-1.82)
Sidechain outliers	100360	2832 (1.86-1.82)
RSRZ outliers	91569	2639 (1.86-1.82)

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	61	<div> <div>61%</div> <div>10%</div> <div>30%</div> </div>
1	C	61	<div> <div>61%</div> <div>8%</div> <div>30%</div> </div>
1	E	61	<div> <div>51%</div> <div>18%</div> <div>30%</div> </div>
1	G	61	<div> <div>56%</div> <div>15%</div> <div>30%</div> </div>
2	B	177	<div> <div>73%</div> <div>14%</div> <div>14%</div> </div>
2	D	177	<div> <div>71%</div> <div>15%</div> <div>14%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	177	 72% 14% • 14%
2	H	177	 71% 15% 14%

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7043 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 Serine protease subunit NS2B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	43	Total	C	N	O	S	0	0	0
			334	205	56	72	1			
1	C	43	Total	C	N	O	S	0	0	0
			334	205	56	72	1			
1	E	43	Total	C	N	O	S	0	0	0
			334	205	56	72	1			
1	G	43	Total	C	N	O	S	0	0	0
			334	205	56	72	1			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	41	GLY	-	expression tag	UNP A0A142IX72
A	42	SER	-	expression tag	UNP A0A142IX72
A	43	HIS	-	expression tag	UNP A0A142IX72
A	44	MET	-	expression tag	UNP A0A142IX72
C	41	GLY	-	expression tag	UNP A0A142IX72
C	42	SER	-	expression tag	UNP A0A142IX72
C	43	HIS	-	expression tag	UNP A0A142IX72
C	44	MET	-	expression tag	UNP A0A142IX72
E	41	GLY	-	expression tag	UNP A0A142IX72
E	42	SER	-	expression tag	UNP A0A142IX72
E	43	HIS	-	expression tag	UNP A0A142IX72
E	44	MET	-	expression tag	UNP A0A142IX72
G	41	GLY	-	expression tag	UNP A0A142IX72
G	42	SER	-	expression tag	UNP A0A142IX72
G	43	HIS	-	expression tag	UNP A0A142IX72
G	44	MET	-	expression tag	UNP A0A142IX72

- Molecule 2 is a protein called Serine protease NS3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	153	Total	C	N	O	S	0	0	0
			1144	724	202	213	5			
2	D	152	Total	C	N	O	S	0	0	0
			1139	721	201	212	5			
2	F	152	Total	C	N	O	S	0	1	0
			1140	721	200	214	5			
2	H	153	Total	C	N	O	S	0	0	0
			1144	724	202	213	5			

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Cl	0	0
			1	1		
3	D	1	Total	Cl	0	0
			1	1		

- Molecule 4 is water.

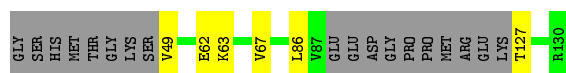
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	89	Total	O	0	0
			89	89		
4	B	199	Total	O	0	0
			199	199		
4	C	81	Total	O	0	0
			81	81		
4	D	199	Total	O	0	0
			199	199		
4	E	101	Total	O	0	0
			101	101		
4	F	210	Total	O	0	0
			210	210		
4	G	75	Total	O	0	0
			75	75		
4	H	184	Total	O	0	0
			184	184		

### 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 errors 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.

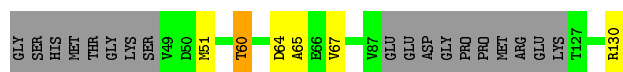
- Molecule 1: Serine protease subunit NS2B

Chain A: 



- Molecule 1: Serine protease subunit NS2B

Chain C: 



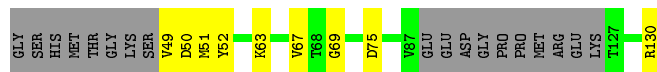
- Molecule 1: Serine protease subunit NS2B

Chain E: 



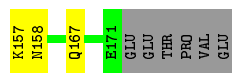
- Molecule 1: Serine protease subunit NS2B

Chain G: 



- Molecule 2: Serine protease NS3

Chain B: 



- Molecule 2: Serine protease NS3

SER	GLY	ALA	LEU	TRP	ASP	VAL	PRO	ALA	PRO	LYS	GLU	VAL	LYS	LYS	LYS	GLY	GLU	THR	T19	R28	V38	G39	V40	F46	M49	W50	E62	G63	R64	Y79	K84	L85	W89	D90	K107	N108	I109	Q110	T111	L112	T118	K119	D120	S135	G136	S137	P138	I139	L140	R145	R170	GLU	GLU	GLU	THR	PRO	VAL	GLU
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----

- Chain F:  72% 14% • 14%

[illegible]

- Chain H:  71% 15% 14%

SER	G144	GLY	ALA	LEU	TRP	ASP	VAL	PRO	ALA	PRO	LYS	GLU	VAL	LYS	LYS	LYS	GLY	GLU	THR	T19	D20	M26	T27	R28	R29	G32	V38	G39	E43	F46	R59	Q74	Y79	P82	R83	K84	R105	Q110	K119	D120	Y130	G133	G136	K142	C143
-----	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	88.79 Å 88.79 Å 138.10 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	31.50 – 1.84 31.50 – 1.84	Depositor EDS
% Data completeness (in resolution range)	97.8 (31.50-1.84) 94.6 (31.50-1.84)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.70 (at 1.84 Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.169 , 0.195 0.167 , 0.193	Depositor DCC
$R_{free}$ test set	4846 reflections (4.84%)	DCC
Wilson B-factor (Å <sup>2</sup> )	21.1	Xtriage
Anisotropy	0.012	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 32.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.39$ , $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	0.426 for -h,-k,l 0.487 for h,-h-k,-l 0.427 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	7043	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.10% 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](#)

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

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	0.30	0/337	0.53	0/452
1	C	0.30	0/337	0.51	0/452
1	E	0.30	0/337	0.52	0/452
1	G	0.30	0/337	0.52	0/452
2	B	0.31	0/1168	0.54	0/1585
2	D	0.31	0/1163	0.56	1/1578 (0.1%)
2	F	0.30	0/1167	0.54	0/1585
2	H	0.30	0/1168	0.54	0/1585
All	All	0.31	0/6014	0.54	1/8141 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	140	LEU	CA-CB-CG	5.33	127.55	115.30

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	334	0	315	9	0
1	C	334	0	315	6	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	334	0	315	15	0
1	G	334	0	315	16	0
2	B	1144	0	1139	21	0
2	D	1139	0	1137	27	0
2	F	1140	0	1130	23	0
2	H	1144	0	1139	30	0
3	B	1	0	0	0	0
3	D	1	0	0	0	0
4	A	89	0	0	6	0
4	B	199	0	0	12	3
4	C	81	0	0	4	0
4	D	199	0	0	9	3
4	E	101	0	0	8	1
4	F	210	0	0	8	2
4	G	75	0	0	9	0
4	H	184	0	0	10	3
All	All	7043	0	5805	124	6

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 11.

The worst 5 of 124 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)
2:F:86:ASP:OD1	4:F:201:HOH:O	1.85	0.94
1:A:49:VAL:O	4:A:201:HOH:O	1.96	0.82
2:B:107:LYS:NZ	4:B:2105:HOH:O	2.12	0.81
1:A:49:VAL:N	4:A:202:HOH:O	2.13	0.80
2:B:167:GLN:O	4:B:2101:HOH:O	1.98	0.80

The worst 5 of 6 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)
4:B:2204:HOH:O	4:F:366:HOH:O[3_554]	1.80	0.40
4:B:2204:HOH:O	4:F:344:HOH:O[3_554]	1.88	0.32
4:D:336:HOH:O	4:H:251:HOH:O[2_665]	2.09	0.11
4:B:2242:HOH:O	4:H:370:HOH:O[3_564]	2.12	0.08
4:D:485:HOH:O	4:H:384:HOH:O[2_665]	2.15	0.05

## 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	39/61 (64%)	39 (100%)	0	0	100	100
1	C	39/61 (64%)	38 (97%)	1 (3%)	0	100	100
1	E	39/61 (64%)	38 (97%)	1 (3%)	0	100	100
1	G	39/61 (64%)	39 (100%)	0	0	100	100
2	B	151/177 (85%)	147 (97%)	4 (3%)	0	100	100
2	D	150/177 (85%)	146 (97%)	4 (3%)	0	100	100
2	F	151/177 (85%)	145 (96%)	6 (4%)	0	100	100
2	H	151/177 (85%)	146 (97%)	5 (3%)	0	100	100
All	All	759/952 (80%)	738 (97%)	21 (3%)	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	36/51 (71%)	36 (100%)	0	100	100
1	C	36/51 (71%)	34 (94%)	2 (6%)	26	8
1	E	36/51 (71%)	34 (94%)	2 (6%)	26	8
1	G	36/51 (71%)	35 (97%)	1 (3%)	51	32
2	B	118/140 (84%)	116 (98%)	2 (2%)	68	54
2	D	118/140 (84%)	117 (99%)	1 (1%)	86	81

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	F	118/140 (84%)	117 (99%)	1 (1%)	86	81
2	H	118/140 (84%)	116 (98%)	2 (2%)	68	54
All	All	616/764 (81%)	605 (98%)	11 (2%)	66	52

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	140	LEU
1	E	50	ASP
1	G	51	MET
1	C	60	THR
2	F	170	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	158	ASN
2	D	108	ASN
2	F	35	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 2 are 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

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	43/61 (70%)	-0.51	0 100 100	17, 32, 40, 45	0
1	C	43/61 (70%)	-0.54	0 100 100	18, 33, 42, 47	0
1	E	43/61 (70%)	-0.42	0 100 100	17, 32, 45, 53	0
1	G	43/61 (70%)	-0.42	0 100 100	18, 33, 43, 52	0
2	B	153/177 (86%)	-0.61	0 100 100	16, 25, 40, 59	0
2	D	152/177 (85%)	-0.62	0 100 100	16, 25, 36, 42	0
2	F	152/177 (85%)	-0.63	0 100 100	15, 25, 38, 51	0
2	H	153/177 (86%)	-0.64	0 100 100	17, 24, 37, 58	0
All	All	782/952 (82%)	-0.59	0 100 100	15, 26, 41, 59	0

There are no RSRZ outliers to report.

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	CL	D	201	1/1	0.99	0.07	-0.14	19,19,19,19	0
3	CL	B	2001	1/1	0.99	0.07	-0.64	19,19,19,19	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.