



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 08:46 am GMT

PDB ID : 5GPI  
Title : Crystal Structures of Unlinked NS2B-NS3 Protease from Zika Virus and Its Complex with a Reverse Peptide Inhibitor  
Authors : Phoo, W.W.; Zhang, Z.Z.  
Deposited on : 2016-08-02  
Resolution : 1.58 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
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)	:	recalc28949

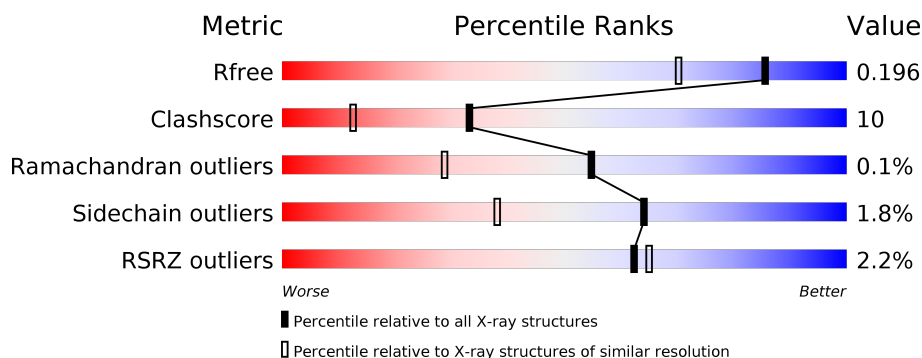
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.58 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	4211 (1.60-1.56)
Clashscore	112137	4539 (1.60-1.56)
Ramachandran outliers	110173	4423 (1.60-1.56)
Sidechain outliers	110143	4420 (1.60-1.56)
RSRZ outliers	101464	4232 (1.60-1.56)

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	B	178	<div> <div>2%</div> <div> <div></div> <div>70%</div> <div>16%</div> <div>13%</div> </div> </div>
1	D	178	<div> <div>%</div> <div> <div></div> <div>72%</div> <div>10%</div> <div>17%</div> </div> </div>
1	F	178	<div> <div>4%</div> <div> <div></div> <div>66%</div> <div>19%</div> <div>15%</div> </div> </div>
1	H	178	<div> <div>2%</div> <div> <div></div> <div>70%</div> <div>16%</div> <div>14%</div> </div> </div>
2	A	53	<div> <div></div> <div> <div></div> <div>60%</div> <div>9%</div> <div>28%</div> </div> </div>
2	C	53	<div> <div></div> <div> <div></div> <div>58%</div> <div>9%</div> <div>30%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	53	<div><div></div><div>4%</div><div>60%</div><div>11%</div><div>28%</div></div>
2	G	53	<div><div></div><div>68%</div><div>8%</div><div>25%</div></div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 6293 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 NS3 protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	153	Total	C	N	O	S	0	1	0
			1147	728	200	214	5			
1	D	148	Total	C	N	O	S	0	1	0
			1098	696	189	208	5			
1	B	154	Total	C	N	O	S	0	4	0
			1160	738	200	217	5			
1	F	152	Total	C	N	O	S	0	0	0
			1113	709	195	204	5			

- Molecule 2 is a protein called NS2B cofactor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	37	Total	C	N	O	S	0	0	0
			280	175	46	58	1			
2	A	38	Total	C	N	O	S	0	0	0
			280	174	46	59	1			
2	E	38	Total	C	N	O	S	0	0	0
			282	177	47	57	1			
2	G	40	Total	C	N	O	S	0	0	0
			299	185	48	65	1			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	H	156	Total	O	0	0
			156	156		
3	C	27	Total	O	0	0
			27	27		
3	A	41	Total	O	0	0
			41	41		
3	E	33	Total	O	0	0
			33	33		

*Continued on next page...*

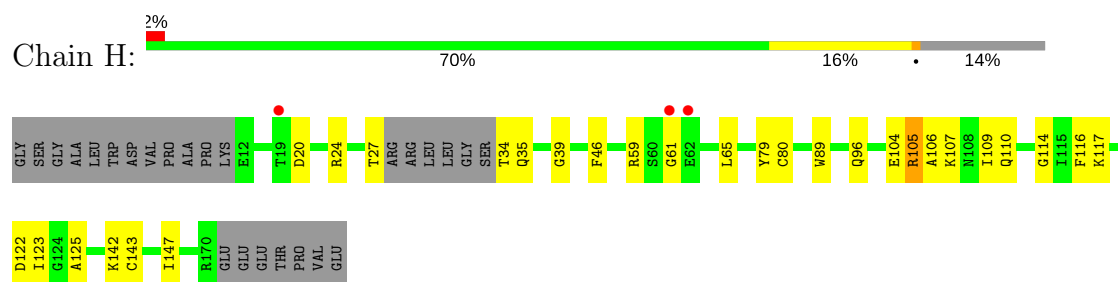
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	G	52	Total 52	O 52	0	0
3	D	102	Total 102	O 102	0	0
3	B	127	Total 127	O 127	0	0
3	F	96	Total 96	O 96	0	0

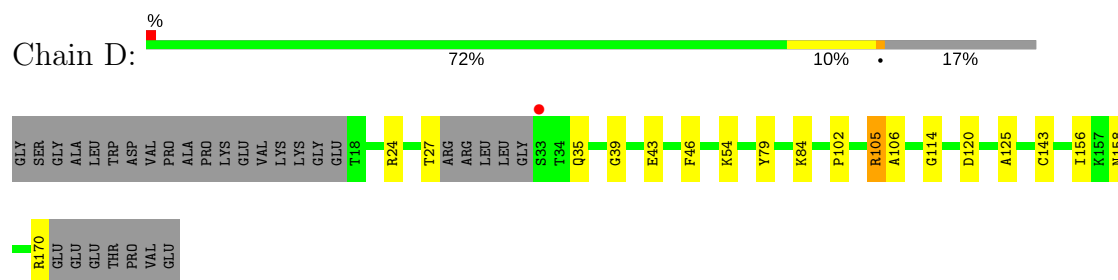
### 3 Residue-property plots [i](#)

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.

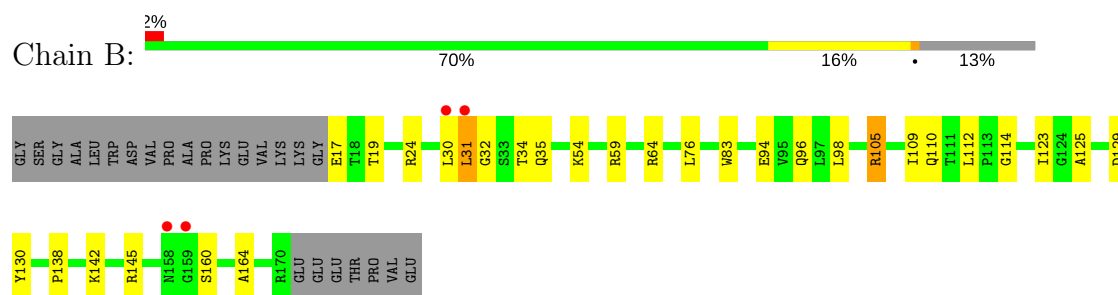
#### • Molecule 1: NS3 protease



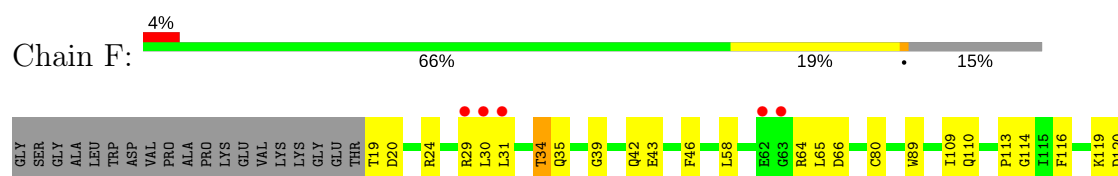
#### • Molecule 1: NS3 protease

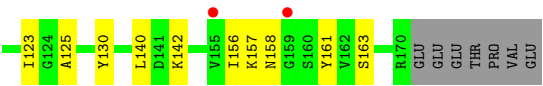


#### • Molecule 1: NS3 protease

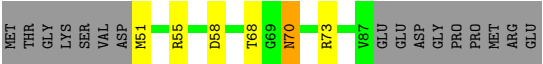


#### • Molecule 1: NS3 protease

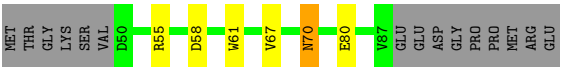




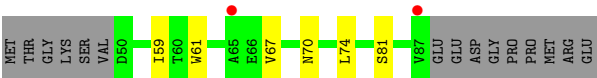
• Molecule 2: NS2B cofactor



• Molecule 2: NS2B cofactor



• Molecule 2: NS2B cofactor



• Molecule 2: NS2B cofactor



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	60.27Å 60.38Å 215.46Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.17 – 1.58 46.17 – 1.58	Depositor EDS
% Data completeness (in resolution range)	99.5 (46.17-1.58) 97.9 (46.17-1.58)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.90 (at 1.58Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, $R_{free}$	0.166 , 0.196 0.166 , 0.196	Depositor DCC
$R_{free}$ test set	5614 reflections (5.17%)	DCC
Wilson B-factor (Å <sup>2</sup> )	18.3	Xtriage
Anisotropy	0.228	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 35.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.38$ , $\langle L^2 \rangle = 0.20$	Xtriage
Estimated twinning fraction	0.326 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	6293	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.88% 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

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	B	0.28	0/1190	0.54	0/1619
1	D	0.27	0/1121	0.48	0/1523
1	F	0.27	0/1137	0.50	0/1546
1	H	0.27	0/1170	0.54	1/1585 (0.1%)
2	A	0.25	0/284	0.47	0/387
2	C	0.34	0/284	0.52	0/385
2	E	0.24	0/286	0.49	0/388
2	G	0.24	0/303	0.50	0/412
All	All	0.27	0/5775	0.51	1/7845 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	105	ARG	NE-CZ-NH1	-5.33	117.64	120.30

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	B	1160	0	1149	31	0
1	D	1098	0	1076	16	0
1	F	1113	0	1103	27	0
1	H	1147	0	1143	24	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	280	0	250	8	0
2	C	280	0	261	5	0
2	E	282	0	261	6	0
2	G	299	0	264	5	0
3	A	41	0	0	2	2
3	B	127	0	0	14	1
3	C	27	0	0	1	0
3	D	102	0	0	6	1
3	E	33	0	0	2	0
3	F	96	0	0	5	1
3	G	52	0	0	3	2
3	H	156	0	0	6	4
All	All	6293	0	5507	109	7

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 10.

The worst 5 of 109 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:D:106:ALA:O	3:D:201:HOH:O	1.88	0.90
1:B:129:ASP:OD2	3:B:201:HOH:O	1.89	0.89
1:B:17:GLU:OE1	3:B:202:HOH:O	1.98	0.81
1:H:104:GLU:OE2	3:H:201:HOH:O	1.99	0.79
1:B:31:LEU:CD2	1:B:32:GLY:H	1.97	0.78

The worst 5 of 7 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)
3:A:107:HOH:O	3:A:111:HOH:O[3_444]	2.01	0.19
3:H:330:HOH:O	3:G:114:HOH:O[3_554]	2.01	0.19
3:H:307:HOH:O	3:A:139:HOH:O[1_645]	2.09	0.11
3:D:285:HOH:O	3:F:283:HOH:O[4_555]	2.13	0.07
3:G:139:HOH:O	3:G:143:HOH:O[3_554]	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	B	156/178 (88%)	151 (97%)	5 (3%)	0	100	100
1	D	145/178 (82%)	142 (98%)	3 (2%)	0	100	100
1	F	150/178 (84%)	139 (93%)	11 (7%)	0	100	100
1	H	150/178 (84%)	145 (97%)	4 (3%)	1 (1%)	25	7
2	A	36/53 (68%)	35 (97%)	1 (3%)	0	100	100
2	C	35/53 (66%)	34 (97%)	1 (3%)	0	100	100
2	E	36/53 (68%)	35 (97%)	1 (3%)	0	100	100
2	G	38/53 (72%)	38 (100%)	0	0	100	100
All	All	746/924 (81%)	719 (96%)	26 (4%)	1 (0%)	55	29

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	61	GLY

### 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	B	119/140 (85%)	115 (97%)	4 (3%)	42	14
1	D	113/140 (81%)	111 (98%)	2 (2%)	64	38
1	F	112/140 (80%)	111 (99%)	1 (1%)	82	68
1	H	118/140 (84%)	118 (100%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	A	28/45 (62%)	27 (96%)	1 (4%)	40	12
2	C	29/45 (64%)	27 (93%)	2 (7%)	18	3
2	E	28/45 (62%)	28 (100%)	0	100	100
2	G	30/45 (67%)	30 (100%)	0	100	100
All	All	577/740 (78%)	567 (98%)	10 (2%)	64	40

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	105	ARG
1	B	19	THR
1	B	54	LYS
1	D	54	LYS
1	B	31	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	35	GLN
1	D	108	ASN
1	B	110	GLN
2	A	70	ASN
1	B	108	ASN

### 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 [i](#)

There are no ligands in this entry.

## 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 [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	B	154/178 (86%)	-0.02	4 (2%) 56 58	9, 15, 39, 58	0
1	D	148/178 (83%)	-0.15	1 (0%) 87 89	13, 24, 45, 60	0
1	F	152/178 (85%)	0.18	7 (4%) 33 33	13, 28, 56, 78	0
1	H	153/178 (85%)	-0.03	3 (1%) 65 68	8, 15, 38, 67	0
2	A	38/53 (71%)	-0.10	0 100 100	12, 20, 32, 45	0
2	C	37/53 (69%)	0.03	0 100 100	19, 29, 48, 52	0
2	E	38/53 (71%)	0.34	2 (5%) 27 28	21, 34, 62, 66	0
2	G	40/53 (75%)	-0.11	0 100 100	13, 22, 43, 45	0
All	All	760/924 (82%)	0.01	17 (2%) 62 65	8, 22, 48, 78	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	30	LEU	8.6
1	F	31	LEU	4.2
2	E	87	VAL	3.7
1	F	63	GLY	3.6
1	F	29	ARG	3.5

### 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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.