



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 4, 2024 – 12:30 PM JST

PDB ID : 7DOC  
Title : Crystal structure of Zika NS2B-NS3 protease with compound 5  
Authors : Quek, J.P.  
Deposited on : 2020-12-14  
Resolution : 1.90 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

<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 : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.37.1  
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) : 2.37.1

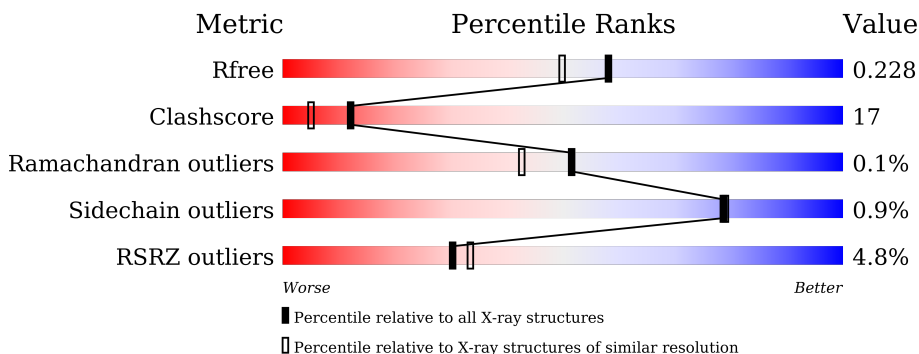
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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.



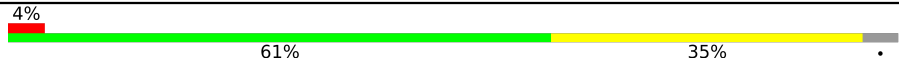


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 of 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	39	 8% 72% 28%
1	G	39	 3% 79% 21%
2	B	153	 5% 75% 24%
3	C	38	 11% 76% 24%
4	D	152	 3% 69% 31%
5	E	39	 74% 26%

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
6	F	158	
7	H	153	
8	I	5	

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 5976 atoms, of which 48 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 Core protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	39	289	179	47	62	1	0	0	0
1	G	39	287	178	44	64	1	0	0	0

- Molecule 2 is a protein called Core protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	153	1083	689	179	210	5	0	0	0

- Molecule 3 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	38	274	171	43	59	1	0	0	0

- Molecule 4 is a protein called Core protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	152	1066	678	175	208	5	0	0	0

- Molecule 5 is a protein called Serine protease subunit NS2B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	39	288	179	44	64	1	0	0	0

- Molecule 6 is a protein called Core protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	F	152	1104	700	187	212	5	0	0	0

- Molecule 7 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	H	153	1110	702	191	212	5	0	0	0

- Molecule 8 is a protein called 3-PYRIDIN-4-YL-2,4-DIHYDRO-INDENO[1,2-C.]PYRAZO  
LE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
8	I	5	104	33	48	14	8	1	0	0	0

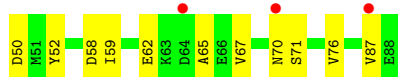
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	12	Total	O	0	0
			12	12		
9	B	50	Total	O	0	0
			50	50		
9	C	15	Total	O	0	0
			15	15		
9	D	60	Total	O	0	0
			60	60		
9	E	25	Total	O	0	0
			25	25		
9	F	89	Total	O	0	0
			89	89		
9	G	32	Total	O	0	0
			32	32		
9	H	88	Total	O	0	0
			88	88		

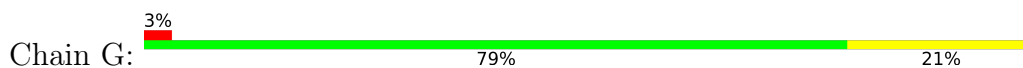
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

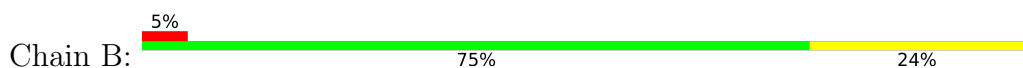
- Molecule 1: Core protein



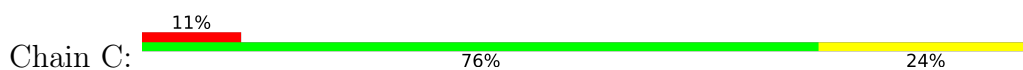
- Molecule 1: Core protein



- Molecule 2: Core protein

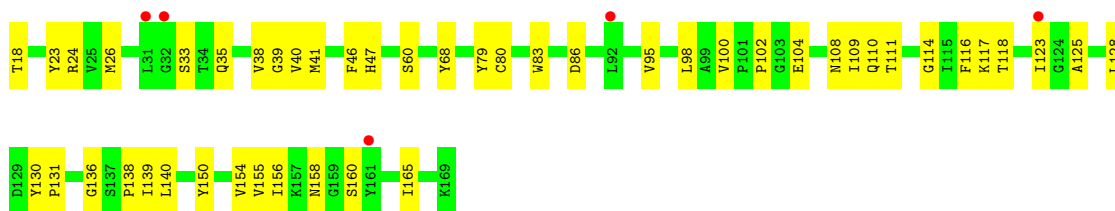


- Molecule 3: Genome polyprotein



- Molecule 4: Core protein





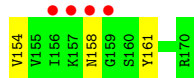
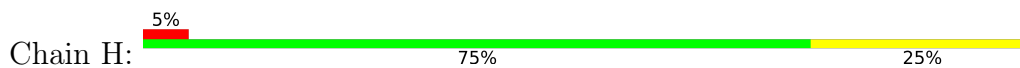
- Molecule 5: Serine protease subunit NS2B



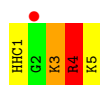
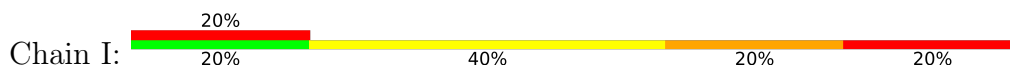
- Molecule 6: Core protein



- Molecule 7: Genome polyprotein



- Molecule 8: 3-PYRIDIN-4-YL-2,4-DIHYDRO-INDENO[1,2-C]PYRAZOLE



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	59.15Å 59.61Å 215.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.63 – 1.90 45.84 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.4 (45.63-1.90) 97.2 (45.84-1.90)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	34.93 (at 1.91Å)	Xtriage
Refinement program	PHENIX 1.16_3549	Depositor
R, $R_{free}$	0.209 , 0.229 0.210 , 0.228	Depositor DCC
$R_{free}$ test set	2981 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	35.4	Xtriage
Anisotropy	0.545	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 53.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	0.084 for k,h,-l	Xtriage
Reported twinning fraction	0.150 for k,h,-l	Depositor
Outliers	1 of 60244 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	5976	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 21.97 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.2714e-03.*

<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: HHC

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	0.31	0/293	0.62	0/399
1	G	0.24	0/291	0.45	0/397
2	B	0.26	0/1107	0.48	0/1517
3	C	0.26	0/278	0.47	0/380
4	D	0.25	0/1090	0.46	0/1494
5	E	0.26	0/292	0.49	0/398
6	F	0.29	0/1126	0.48	0/1534
7	H	0.26	0/1134	0.48	0/1546
8	I	3.13	3/33 (9.1%)	1.60	0/38
All	All	0.36	3/5644 (0.1%)	0.49	0/7703

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	I	4	ARG	CZ-NH2	9.59	1.45	1.33
8	I	4	ARG	NE-CZ	9.55	1.45	1.33
8	I	4	ARG	CZ-NH1	-5.95	1.25	1.33

There are no bond angle outliers.

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	289	0	256	8	2
1	G	287	0	249	8	1
2	B	1083	0	1017	33	0
3	C	274	0	239	13	0
4	D	1066	0	991	40	1
5	E	288	0	251	15	0
6	F	1104	0	1051	55	0
7	H	1110	0	1067	38	0
8	I	56	48	42	7	0
9	A	12	0	0	0	1
9	B	50	0	0	6	1
9	C	15	0	0	1	0
9	D	60	0	0	4	0
9	E	25	0	0	4	0
9	F	89	0	0	10	1
9	G	32	0	0	1	0
9	H	88	0	0	4	1
All	All	5928	48	5163	183	5

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 17.

The worst 5 of 183 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)
6:F:86:ASP:OD2	9:F:201:HOH:O	1.93	0.86
7:H:94:GLU:OE1	7:H:110:GLN:NE2	2.15	0.79
7:H:27:THR:CG2	7:H:34:THR:HB	2.13	0.78
4:D:116:PHE:HB2	4:D:123:ILE:HG23	1.66	0.77
7:H:27:THR:HG22	7:H:34:THR:HB	1.67	0.77

All (5) 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:58:ASP:OD1	1:A:70:ASN:OD1[3_555]	1.41	0.79
9:F:281:HOH:O	9:H:274:HOH:O[4_545]	1.97	0.23
9:A:109:HOH:O	9:B:243:HOH:O[3_555]	2.09	0.11
1:A:58:ASP:CG	1:A:70:ASN:OD1[3_555]	2.14	0.06
4:D:68:TYR:OH	1:G:80:GLU:O[3_455]	2.17	0.03

## 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	37/39 (95%)	35 (95%)	2 (5%)	0	100	100
1	G	37/39 (95%)	36 (97%)	1 (3%)	0	100	100
2	B	151/153 (99%)	145 (96%)	5 (3%)	1 (1%)	22	12
3	C	36/38 (95%)	35 (97%)	1 (3%)	0	100	100
4	D	150/152 (99%)	145 (97%)	5 (3%)	0	100	100
5	E	37/39 (95%)	36 (97%)	1 (3%)	0	100	100
6	F	146/158 (92%)	141 (97%)	5 (3%)	0	100	100
7	H	151/153 (99%)	143 (95%)	8 (5%)	0	100	100
8	I	3/5 (60%)	3 (100%)	0	0	100	100
All	All	748/776 (96%)	719 (96%)	28 (4%)	1 (0%)	51	42

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	30	LEU

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

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	29/33 (88%)	28 (97%)	1 (3%)	37	28
1	G	29/33 (88%)	29 (100%)	0	100	100
2	B	106/120 (88%)	105 (99%)	1 (1%)	78	79

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	C	27/32 (84%)	27 (100%)	0	100	100
4	D	103/119 (87%)	102 (99%)	1 (1%)	76	76
5	E	29/33 (88%)	29 (100%)	0	100	100
6	F	110/124 (89%)	110 (100%)	0	100	100
7	H	111/120 (92%)	111 (100%)	0	100	100
8	I	3/3 (100%)	1 (33%)	2 (67%)	0	0
All	All	547/617 (89%)	542 (99%)	5 (1%)	78	79

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	50	ASP
2	B	163	SER
4	D	33	SER
8	I	3	LYS
8	I	4	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 monosaccharides 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	A	39/39 (100%)	0.70	3 (7%) 13 15	43, 55, 73, 77	0
1	G	39/39 (100%)	0.48	1 (2%) 56 58	32, 40, 62, 71	0
2	B	153/153 (100%)	0.49	8 (5%) 27 30	33, 47, 63, 69	0
3	C	38/38 (100%)	0.75	4 (10%) 6 7	46, 55, 71, 91	0
4	D	152/152 (100%)	0.46	5 (3%) 46 49	32, 48, 72, 91	0
5	E	39/39 (100%)	0.34	0 100 100	36, 47, 61, 74	0
6	F	152/158 (96%)	0.37	7 (4%) 32 35	28, 36, 54, 76	0
7	H	153/153 (100%)	0.46	8 (5%) 27 30	25, 36, 62, 78	0
8	I	4/5 (80%)	1.80	1 (25%) 0 0	47, 51, 63, 76	0
All	All	769/776 (99%)	0.48	37 (4%) 30 33	25, 44, 67, 91	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
7	H	18	THR	4.8
2	B	133	GLY	3.8
8	I	2	GLY	3.5
2	B	31	LEU	3.4
4	D	92	LEU	3.3

### 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 monosaccharides 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.