



Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 06:41 pm BST

PDB ID : 6UM3
Title : Structure of Zika virus NS2b-NS3 protease mutant stabilizing the super-open conformation
Authors : Aleshin, A.E.; Shiryaev, S.A.; Liddington, R.C.
Deposited on : 2019-10-08
Resolution : 2.50 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/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	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

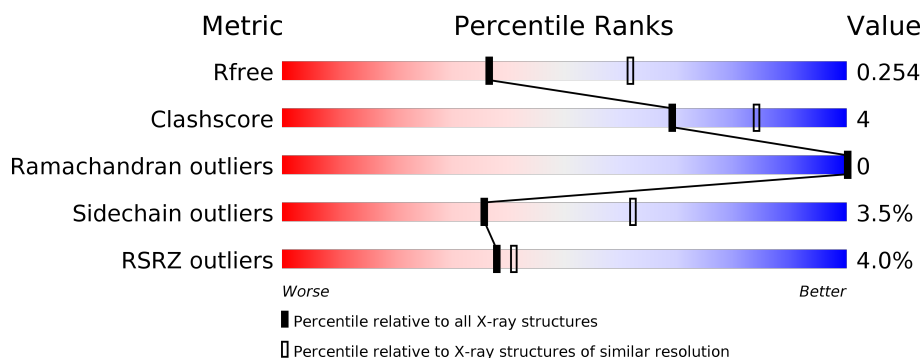
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 ≥ 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	241	<div> <div>2%</div> <div> <div></div> <div>58%</div> <div>8%</div> <div>33%</div> </div> </div>
1	B	241	<div> <div>3%</div> <div> <div></div> <div>61%</div> <div>6%</div> <div>33%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2531 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 fusion.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	162	Total	C	N	O	S	0	0	0
			1231	774	214	237	6			
1	B	161	Total	C	N	O	S	0	0	0
			1226	771	214	235	6			

There are 34 discrepancies between the modelled and reference sequences:

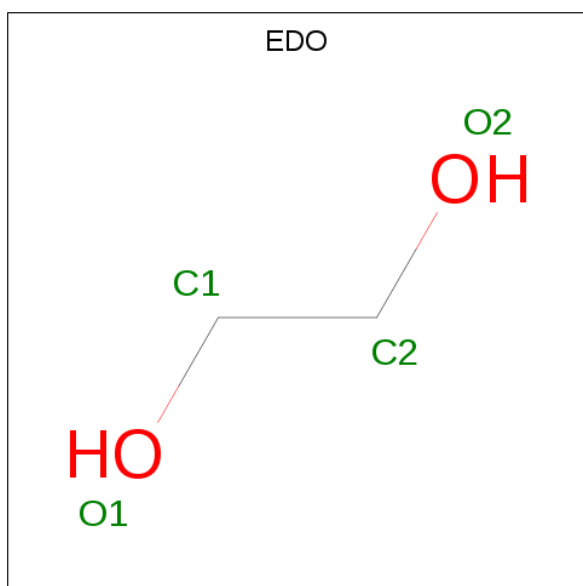
Chain	Residue	Modelled	Actual	Comment	Reference
A	46	GLY	-	expression tag	UNP A0A1V0E2E5
A	47	SER	-	expression tag	UNP A0A1V0E2E5
A	991	ALA	-	linker	UNP A0A1V0E2E5
A	992	GLY	-	linker	UNP A0A1V0E2E5
A	993	GLY	-	linker	UNP A0A1V0E2E5
A	994	GLY	-	linker	UNP A0A1V0E2E5
A	995	GLY	-	linker	UNP A0A1V0E2E5
A	996	SER	-	linker	UNP A0A1V0E2E5
A	997	GLY	-	linker	UNP A0A1V0E2E5
A	998	GLY	-	linker	UNP A0A1V0E2E5
A	999	GLY	-	linker	UNP A0A1V0E2E5
A	1000	GLY	-	linker	UNP A0A1V0E2E5
A	1080	SER	CYS	conflict	UNP A0A219YLK5
A	1088	CYS	ALA	conflict	UNP A0A219YLK5
A	1143	SER	CYS	conflict	UNP A0A219YLK5
A	1157	CYS	LYS	conflict	UNP A0A219YLK5
A	1178	SER	CYS	conflict	UNP A0A219YLK5
B	46	GLY	-	expression tag	UNP A0A1V0E2E5
B	47	SER	-	expression tag	UNP A0A1V0E2E5
B	991	ALA	-	linker	UNP A0A1V0E2E5
B	992	GLY	-	linker	UNP A0A1V0E2E5
B	993	GLY	-	linker	UNP A0A1V0E2E5
B	994	GLY	-	linker	UNP A0A1V0E2E5
B	995	GLY	-	linker	UNP A0A1V0E2E5
B	996	SER	-	linker	UNP A0A1V0E2E5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	997	GLY	-	linker	UNP A0A1V0E2E5
B	998	GLY	-	linker	UNP A0A1V0E2E5
B	999	GLY	-	linker	UNP A0A1V0E2E5
B	1000	GLY	-	linker	UNP A0A1V0E2E5
B	1080	SER	CYS	conflict	UNP A0A219YLK5
B	1088	CYS	ALA	conflict	UNP A0A219YLK5
B	1143	SER	CYS	conflict	UNP A0A219YLK5
B	1157	CYS	LYS	conflict	UNP A0A219YLK5
B	1178	SER	CYS	conflict	UNP A0A219YLK5

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			4	2	2		

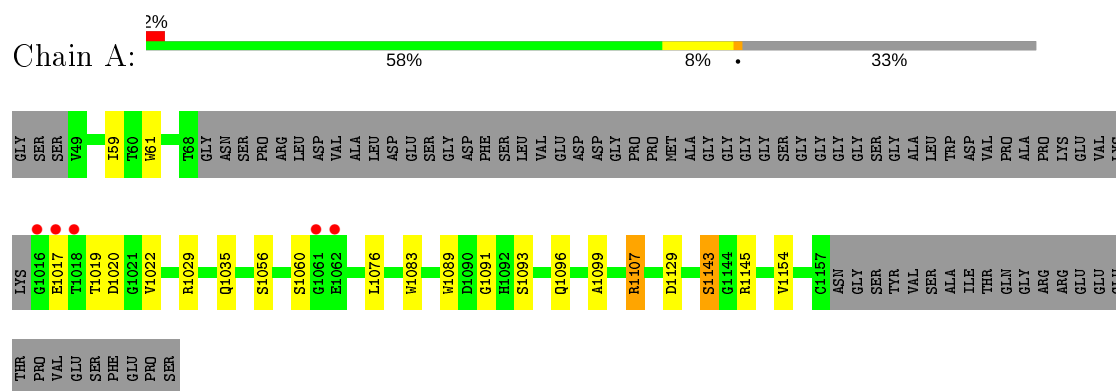
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	41	Total	O	0	0
			41	41		
3	B	29	Total	O	0	0
			29	29		

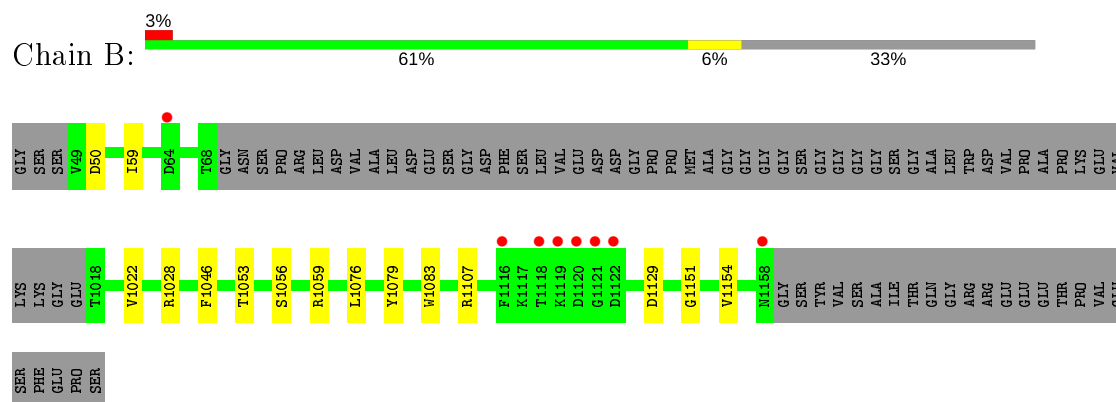
3 Residue-property plots [i](#)

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: NS2B-NS3 PROTEASE fusion



• Molecule 1: NS2B-NS3 PROTEASE fusion



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	51.74Å 55.90Å 61.67Å 90.00° 112.91° 90.00°	Depositor
Resolution (Å)	39.88 – 2.50 39.84 – 2.50	Depositor EDS
% Data completeness (in resolution range)	95.8 (39.88-2.50) 95.9 (39.84-2.50)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.04 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, R_{free}	0.188 , 0.249 0.195 , 0.254	Depositor DCC
R_{free} test set	525 reflections (4.81%)	wwPDB-VP
Wilson B-factor (Å ²)	41.9	Xtriage
Anisotropy	0.764	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 40.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.022 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	2531	wwPDB-VP
Average B, all atoms (Å ²)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.43% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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

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

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	0.69	0/1257	0.85	0/1705
1	B	0.67	0/1252	0.85	0/1699
All	All	0.68	0/2509	0.85	0/3404

There are no bond length outliers.

There are no bond angle outliers.

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	A	1231	0	1196	12	0
1	B	1226	0	1193	7	0
2	A	4	0	6	1	0
3	A	41	0	0	1	0
3	B	29	0	0	0	0
All	All	2531	0	2395	18	0

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 (18) 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:59:ILE:HD11	1:B:1022:VAL:HB	1.84	0.59
1:B:59:ILE:CD1	1:B:1022:VAL:HB	2.35	0.57
1:A:1060:SER:O	1:A:1060:SER:OG	2.20	0.55
1:A:59:ILE:CD1	1:A:1022:VAL:HB	2.39	0.53
1:A:1029:ARG:NH1	1:B:1151:GLY:O	2.44	0.51
1:A:1076:LEU:HD22	1:A:1083:TRP:CH2	2.47	0.50
1:A:1129:ASP:OD1	3:A:1301:HOH:O	2.20	0.48
1:A:1017:GLU:OE1	1:A:1017:GLU:N	2.47	0.48
1:A:1089:TRP:CZ3	1:A:1091:GLY:HA2	2.49	0.47
1:A:1143:SER:HB2	1:A:1145:ARG:HD3	1.96	0.47
1:B:1076:LEU:HD22	1:B:1083:TRP:CH2	2.50	0.46
1:B:1053:THR:O	1:B:1056:SER:HB3	2.16	0.46
1:A:1035:GLN:HB2	2:A:1201:EDO:H11	1.99	0.44
1:B:50:ASP:OD2	1:B:1059:ARG:NH2	2.51	0.44
1:A:61:TRP:CZ3	1:A:1096:GLN:HG3	2.53	0.44
1:A:1099:ALA:HB3	1:A:1107:ARG:HD2	2.00	0.44
1:A:1019:THR:HG22	1:A:1020:ASP:O	2.18	0.43
1:B:1046:PHE:HB3	1:B:1079:TYR:HB2	2.00	0.42

There are no symmetry-related clashes.

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	158/241 (66%)	156 (99%)	2 (1%)	0	100	100
1	B	157/241 (65%)	155 (99%)	2 (1%)	0	100	100
All	All	315/482 (65%)	311 (99%)	4 (1%)	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	129/189 (68%)	124 (96%)	5 (4%)	32	57
1	B	129/189 (68%)	125 (97%)	4 (3%)	40	67
All	All	258/378 (68%)	249 (96%)	9 (4%)	36	62

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1056	SER
1	A	1093	SER
1	A	1107	ARG
1	A	1143	SER
1	A	1154	VAL
1	B	1028	ARG
1	B	1107	ARG
1	B	1129	ASP
1	B	1154	VAL

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

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

1 ligand is 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	$\# Z > 2$
2	EDO	A	1201	-	3,3,3	0.27	0	2,2,2	0.68	0

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
2	EDO	A	1201	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1201	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1201	EDO	1	0

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, 95th 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 ²)	Q<0.9
1	A	162/241 (67%)	-0.10	5 (3%) 49 52	20, 36, 65, 102	0
1	B	161/241 (66%)	0.04	8 (4%) 28 30	21, 40, 75, 117	0
All	All	323/482 (67%)	-0.03	13 (4%) 38 41	20, 38, 75, 117	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1018	THR	5.3
1	A	1062	GLU	4.6
1	B	1118	THR	4.4
1	B	1116	PHE	3.9
1	B	1120	ASP	3.7
1	B	1119	LYS	3.2
1	B	64	ASP	3.1
1	A	1017	GLU	3.0
1	A	1061	GLY	2.9
1	A	1016	GLY	2.3
1	B	1122	ASP	2.3
1	B	1121	GLY	2.2
1	B	1158	ASN	2.1

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

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, 95th 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(\AA^2)	Q<0.9
2	EDO	A	1201	4/4	0.86	0.28	38,44,44,44	0

6.5 Other polymers

There are no such residues in this entry.