



# Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 12:07 am BST

PDB ID : 2WV9  
Title : Crystal Structure of the NS3 protease-helicase from Murray Valley encephalitis virus  
Authors : Assenberg, R.; Mastrangelo, E.; Walter, T.S.; Verma, A.; Milani, M.; Owens, R.J.; Stuart, D.I.; Grimes, J.M.; Mancini, E.J.  
Deposited on : 2009-10-15  
Resolution : 2.75 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

<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  
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

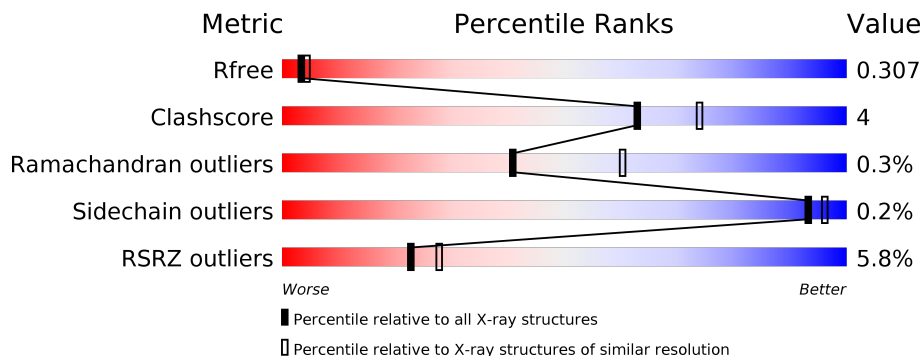
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.75 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	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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 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	673	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4766 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 FLAVIVIRIN PROTEASE NS2B REGULATORY SUBUNIT, FLAVIVIRIN PROTEASE NS3 CATALYTIC SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	601	4677	2952	830	875	20	0	0	0

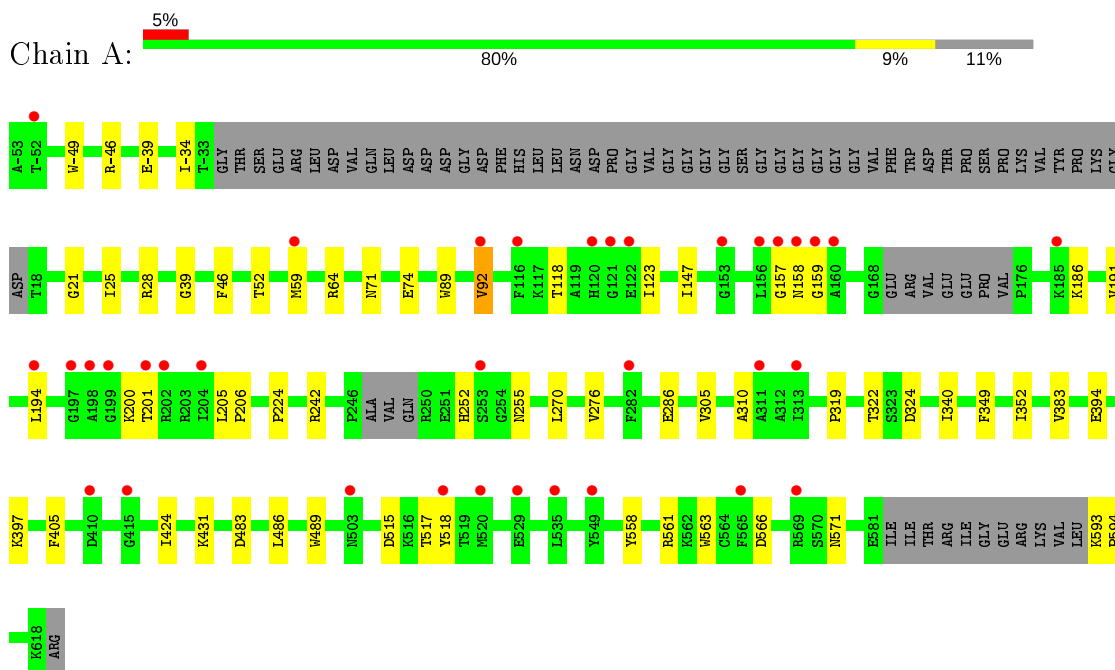
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	89	Total	O	0	0
			89	89		

### 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: FLAVIVIRIN PROTEASE NS2B REGULATORY SUBUNIT, FLAVIVIRIN PROTEASE NS3 CATALYTIC SUBUNIT



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	41.91Å 105.46Å 80.07Å 90.00° 97.42° 90.00°	Depositor
Resolution (Å)	43.94 – 2.75 43.93 – 2.75	Depositor EDS
% Data completeness (in resolution range)	98.6 (43.94-2.75) 98.6 (43.93-2.75)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.64 (at 2.77Å)	Xtrriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.267 , 0.300 0.273 , 0.307	Depositor DCC
$R_{free}$ test set	884 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	50.5	Xtrriage
Anisotropy	0.095	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 51.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.38$ , $\langle L^2 \rangle = 0.21$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	4766	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	6.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.12% 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	0.32	0/4785	0.48	0/6496

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 [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	4677	0	4610	40	0
2	A	89	0	0	1	0
All	All	4766	0	4610	40	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 (40) 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:A:489:TRP:HB3	1:A:517:THR:CG2	2.15	0.76
1:A:489:TRP:HB3	1:A:517:THR:HG21	1.72	0.71
1:A:558:TYR:O	1:A:561:ARG:NH1	2.25	0.69
1:A:89:TRP:HB2	1:A:147:ILE:HD13	1.76	0.67

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:186:LYS:O	1:A:310:ALA:O	2.14	0.66
1:A:194:LEU:HB3	1:A:200:LYS:HG3	1.75	0.66
1:A:431:LYS:HD3	1:A:486:LEU:HD21	1.78	0.65
1:A:515:ASP:HA	1:A:518:TYR:CE2	2.33	0.62
1:A:593:LYS:CB	1:A:594:PRO:HD3	2.30	0.62
1:A:593:LYS:HB2	1:A:594:PRO:HD3	1.82	0.61
1:A:92:VAL:HG12	1:A:92:VAL:O	2.02	0.60
1:A:157:GLY:O	1:A:159:GLY:N	2.35	0.59
1:A:340:ILE:HD12	1:A:424:ILE:HG23	1.85	0.57
1:A:59:MET:HG3	1:A:64:ARG:HG3	1.88	0.55
1:A:-46:ARG:HH12	1:A:21:GLY:HA3	1.71	0.55
1:A:571:ASN:O	1:A:594:PRO:HD2	2.07	0.53
1:A:593:LYS:CB	1:A:594:PRO:CD	2.86	0.53
1:A:224:PRO:HG3	1:A:286:GLU:HG3	1.91	0.52
1:A:349:PHE:HB3	1:A:352:ILE:HD12	1.92	0.51
1:A:563:TRP:HA	1:A:566:ASP:OD1	2.12	0.50
1:A:322:THR:HG22	1:A:324:ASP:H	1.77	0.50
1:A:92:VAL:HG13	2:A:2026:HOH:O	2.11	0.50
1:A:242:ARG:HD3	1:A:252:HIS:HB3	1.94	0.49
1:A:-46:ARG:NH1	1:A:21:GLY:HA3	2.26	0.49
1:A:593:LYS:HB2	1:A:594:PRO:CD	2.42	0.49
1:A:394:GLU:HA	1:A:397:LYS:HE3	1.95	0.48
1:A:71:ASN:HB3	1:A:74:GLU:HB2	1.95	0.48
1:A:383:VAL:HG22	1:A:405:PHE:HB2	1.95	0.47
1:A:270:LEU:HD23	1:A:276:VAL:HG21	1.96	0.47
1:A:-39:GLU:HG2	1:A:-34:ILE:HG21	1.96	0.46
1:A:25:ILE:HD12	1:A:52:THR:HG21	1.98	0.45
1:A:305:VAL:HA	1:A:310:ALA:O	2.16	0.44
1:A:191:VAL:HG21	1:A:319:PRO:HB3	1.99	0.44
1:A:39:GLY:HA3	1:A:46:PHE:CZ	2.53	0.43
1:A:118:THR:HG21	1:A:123:ILE:HD12	1.99	0.43
1:A:483:ASP:HB2	1:A:486:LEU:HD22	2.02	0.42
1:A:92:VAL:CG1	1:A:92:VAL:O	2.67	0.41
1:A:201:THR:HG22	1:A:201:THR:O	2.20	0.41
1:A:205:LEU:HB3	1:A:206:PRO:HD3	2.02	0.41
1:A:-49:TRP:CZ3	1:A:28:ARG:HG3	2.56	0.40

There are no symmetry-related clashes.

## 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	591/673 (88%)	572 (97%)	17 (3%)	2 (0%)	41 60

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	158	ASN
1	A	92	VAL

### 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	491/547 (90%)	490 (100%)	1 (0%)	93 96

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	255	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	108	ASN
1	A	377	GLN
1	A	474	HIS

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	508	GLN

### 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 [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

### 6.1 Protein, DNA and RNA chains

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	601/673 (89%)	0.60	35 (5%) <span style="border: 1px solid red; padding: 2px;">23</span> <span style="border: 1px solid red; padding: 2px;">28</span>	2, 5, 11, 33	0

All (35) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	198	ALA	7.6
1	A	197	GLY	5.9
1	A	92	VAL	4.9
1	A	116	PHE	4.8
1	A	158	ASN	4.3
1	A	121	GLY	4.0
1	A	160	ALA	3.7
1	A	194	LEU	3.6
1	A	120	HIS	3.3
1	A	157	GLY	3.0
1	A	410	ASP	3.0
1	A	122	GLU	2.9
1	A	59	MET	2.8
1	A	503	ASN	2.7
1	A	253	SER	2.7
1	A	529	GLU	2.7
1	A	415	GLY	2.7
1	A	518	TYR	2.7
1	A	156	LEU	2.6
1	A	185	LYS	2.4
1	A	565	PHE	2.4
1	A	202	ARG	2.4
1	A	520	MET	2.3
1	A	159	GLY	2.3
1	A	204	ILE	2.3
1	A	311	ALA	2.2
1	A	199	GLY	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	153	GLY	2.2
1	A	549	TYR	2.2
1	A	201	THR	2.2
1	A	569	ARG	2.2
1	A	313	ILE	2.1
1	A	535	LEU	2.1
1	A	-52	THR	2.1
1	A	282	PHE	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.