



# Full wwPDB X-ray Structure Validation Report ⓘ

May 22, 2020 – 05:54 pm BST

PDB ID : 2JLQ  
Title : Dengue virus 4 NS3 helicase structure, apo enzyme.  
Authors : Luo, D.H.; Xu, T.; Watson, R.P.; Becker, D.S.; Sampath, A.; Jahnke, W.;  
Yeong, S.S.; Wang, C.H.; Lim, S.P.; Vasudevan, S.G.; Lescar, J.  
Deposited on : 2008-09-15  
Resolution : 1.67 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  
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

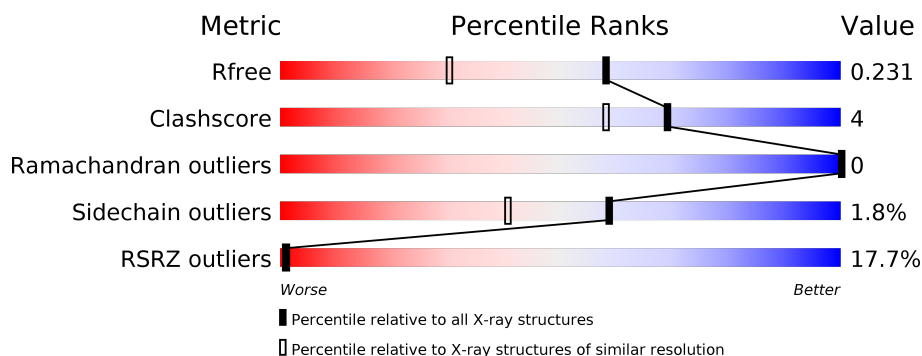
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.67 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	6780 (1.70-1.66)
Clashscore	141614	7310 (1.70-1.66)
Ramachandran outliers	138981	7173 (1.70-1.66)
Sidechain outliers	138945	7172 (1.70-1.66)
RSRZ outliers	127900	6661 (1.70-1.66)

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	451	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4030 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 NS3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	451	Total	C	N	O	S	0	17	0
			3675	2323	650	683	19			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	213	LEU	LYS	conflict	UNP Q2YHF0
A	250	ASP	GLU	conflict	UNP Q2YHF0
A	292	CYS	SER	conflict	UNP Q2YHF0
A	321	SER	THR	conflict	UNP Q2YHF0
A	381	ARG	LYS	conflict	UNP Q2YHF0
A	480	LYS	ARG	conflict	UNP Q2YHF0

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Cl	0	0
			1	1		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		

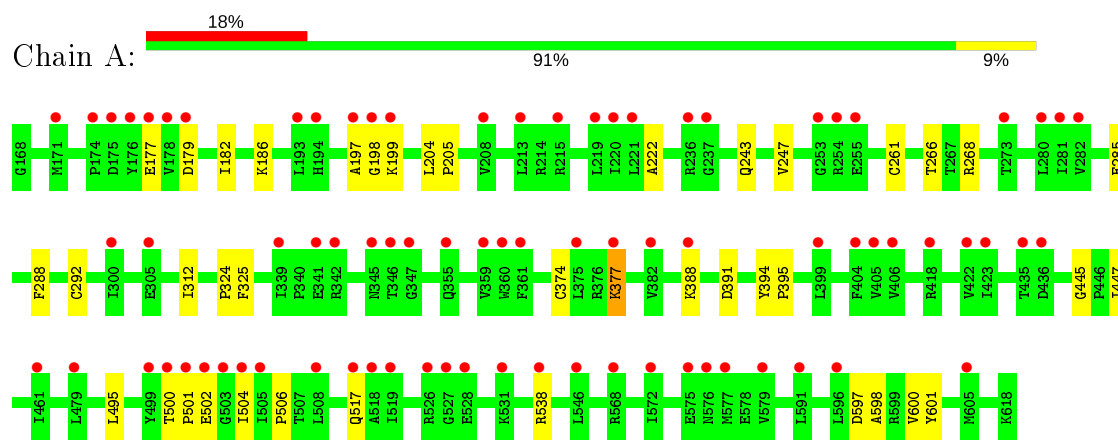
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	348	Total	O	0	0
			348	348		

### 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: SERINE PROTEASE SUBUNIT NS3



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.82Å 92.61Å 102.96Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.96 – 1.67 20.96 – 1.67	Depositor EDS
% Data completeness (in resolution range)	99.6 (20.96-1.67) 99.6 (20.96-1.67)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.40 (at 1.67Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.211 , 0.234 0.207 , 0.231	Depositor DCC
$R_{free}$ test set	2989 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	27.2	Xtriage
Anisotropy	0.063	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 46.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4030	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.02% 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: GOL, 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	0.21	0/3806	0.35	0/5151

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	3675	0	3712	27	0
2	A	1	0	0	0	0
3	A	6	0	8	0	0
4	A	348	0	0	3	0
All	All	4030	0	3720	27	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 (27) 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:292[B]:CYS:SG	4:A:2105:HOH:O	2.30	0.87
1:A:285[A]:GLU:OE2	4:A:2098:HOH:O	2.02	0.76
1:A:495:LEU:HD22	1:A:506:PRO:HB2	1.68	0.76
1:A:247[B]:VAL:HG13	1:A:268:ARG:HH21	1.54	0.73
1:A:197:ALA:N	1:A:198:GLY:HA2	2.03	0.72
1:A:597:ASP:O	1:A:600:VAL:HG22	1.95	0.67
1:A:247[B]:VAL:HG13	1:A:268:ARG:NH2	2.13	0.63
1:A:197:ALA:H	1:A:198:GLY:HA2	1.67	0.60
1:A:197:ALA:N	1:A:198:GLY:CA	2.69	0.55
1:A:177:GLU:HG3	1:A:177:GLU:O	2.08	0.53
1:A:179:ASP:O	1:A:182:ILE:HG12	2.10	0.51
1:A:374:CYS:O	1:A:377:LYS:HG3	2.09	0.51
1:A:182:ILE:HD12	1:A:312:ILE:HD11	1.93	0.50
1:A:222:ALA:O	1:A:261:CYS:HA	2.11	0.50
1:A:598:ALA:HA	1:A:601:TYR:CZ	2.48	0.49
1:A:204:LEU:HB3	1:A:205:PRO:HD3	1.96	0.48
1:A:500:THR:HG23	1:A:502:GLU:H	1.79	0.47
1:A:500:THR:OG1	1:A:501:PRO:HD2	2.15	0.46
1:A:394:TYR:N	1:A:395:PRO:CD	2.80	0.45
1:A:243:GLN:HB3	1:A:268:ARG:NH2	2.31	0.45
1:A:197:ALA:HB3	4:A:2027:HOH:O	2.17	0.44
1:A:500:THR:HG22	1:A:504:ILE:H	1.83	0.43
1:A:324:PRO:HG2	1:A:325:PHE:CD2	2.55	0.42
1:A:204:LEU:N	1:A:205:PRO:CD	2.83	0.41
1:A:445:GLY:HA2	1:A:447:ILE:HG23	2.04	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	466/451 (103%)	458 (98%)	8 (2%)	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	402/388 (104%)	395 (98%)	7 (2%)	60 43

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	199	LYS
1	A	288	PHE
1	A	377	LYS
1	A	388	LYS
1	A	391	ASP
1	A	517	GLN
1	A	538	ARG

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

Mol	Chain	Res	Type
1	A	194	HIS
1	A	517	GLN
1	A	529	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, 1 is monoatomic - leaving 1 for Mogul analysis.

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	$\# Z  > 2$
3	GOL	A	1620	-	5,5,5	0.35	0	5,5,5	0.24	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
3	GOL	A	1620	-	-	0/4/4/4	-

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 ⓘ

### 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	451/451 (100%)	0.97	80 (17%) <b>1</b> <b>1</b>	19, 27, 36, 45	1 (0%)

All (80) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	500	THR	9.9
1	A	501	PRO	9.8
1	A	197	ALA	8.4
1	A	499	TYR	5.5
1	A	517	GLN	5.4
1	A	175	ASP	5.3
1	A	177	GLU	5.0
1	A	347	GLY	5.0
1	A	236	ARG	4.8
1	A	345	ASN	4.8
1	A	503	GLY	4.6
1	A	253	GLY	4.6
1	A	575	GLU	4.5
1	A	360	TRP	4.3
1	A	237	GLY	4.3
1	A	502	GLU	3.9
1	A	519	ILE	3.9
1	A	193	LEU	3.8
1	A	174	PRO	3.6
1	A	505	ILE	3.6
1	A	219	LEU	3.5
1	A	436	ASP	3.4
1	A	577	MET	3.4
1	A	346	THR	3.4
1	A	375	LEU	3.3
1	A	605	MET	3.2
1	A	377	LYS	3.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	342	ARG	3.1
1	A	221	LEU	3.0
1	A	176	TYR	2.9
1	A	255	GLU	2.9
1	A	199	LYS	2.8
1	A	220	ILE	2.8
1	A	418	ARG	2.7
1	A	406	VAL	2.7
1	A	215	ARG	2.7
1	A	538	ARG	2.7
1	A	423	ILE	2.7
1	A	382	VAL	2.7
1	A	179	ASP	2.6
1	A	359	VAL	2.6
1	A	576	ASN	2.6
1	A	405	VAL	2.6
1	A	504	ILE	2.6
1	A	479[A]	LEU	2.6
1	A	254	ARG	2.5
1	A	572	ILE	2.5
1	A	526	ARG	2.5
1	A	531	LYS	2.5
1	A	355	GLN	2.5
1	A	213	LEU	2.5
1	A	281	ILE	2.4
1	A	404	PHE	2.4
1	A	422	VAL	2.4
1	A	305	GLU	2.4
1	A	591	LEU	2.4
1	A	198	GLY	2.4
1	A	596	LEU	2.3
1	A	178	VAL	2.3
1	A	518	ALA	2.3
1	A	194	HIS	2.3
1	A	527	GLY	2.3
1	A	273	THR	2.2
1	A	300	ILE	2.2
1	A	339	ILE	2.2
1	A	388	LYS	2.2
1	A	171	MET	2.2
1	A	280	LEU	2.2
1	A	528	GLU	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	435	THR	2.1
1	A	508	LEU	2.1
1	A	461	ILE	2.1
1	A	361	PHE	2.1
1	A	399	LEU	2.0
1	A	546	LEU	2.0
1	A	568	ARG	2.0
1	A	341	GLU	2.0
1	A	208	VAL	2.0
1	A	282	VAL	2.0
1	A	579	VAL	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](#)

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, 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	B-factors( $\text{\AA}^2$ )	Q<0.9
3	GOL	A	1620	6/6	0.96	0.10	19,21,21,22	0
2	CL	A	1619	1/1	0.98	0.13	37,37,37,37	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.