

Full wwPDB X-ray Structure Validation Report (i)

Jun 24, 2024 – 04:29 PM EDT

PDB ID	:	5019
Title	:	Japanese encephalitis virus non-structural protein 1 C-terminal domain
Authors	:	Poonsiri, T.; Wright, G.S.A.; Antonyuk, S.V.
Deposited on		
Resolution	:	2.10 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 (i) 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 (1)) were used in the production of this report:

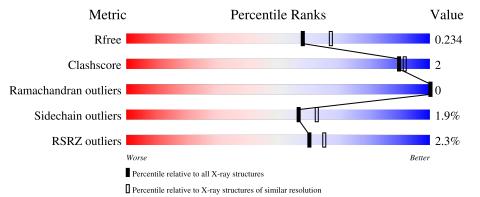
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: $X\text{-}RAY \, DIFFRACTION$

The reported resolution of this entry is 2.10 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	$\begin{array}{c} \textbf{Whole archive} \\ \textbf{(\#Entries)} \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
R_{free}	130704	5197(2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 >=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 <=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	
			2%	
1	А	182	90%	6% • •



2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2924 atoms, of which 1350 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.

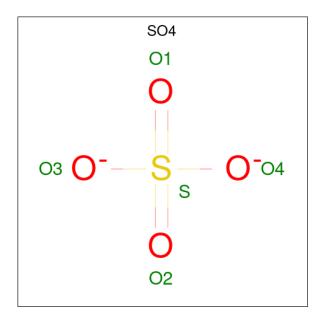
• Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
1	А	176	Total 2722	C 869	Н 1324	N 245	О 274	S 10	33	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	171	MET	-	initiating methionine	UNP P27395

• Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



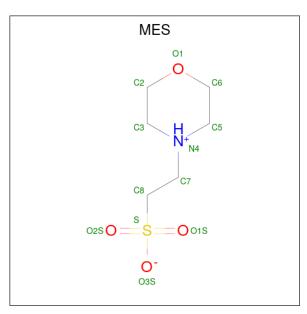
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0



Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
2	А	1	TotalOS541	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0

• Molecule 3 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf		
2	Δ	1	Total	С	Η	Ν	0	S	0	0
5	D A	1	25	6	13	1	4	1	0	0
9	٨	1	Total	С	Η	Ν	Ο	S	0	0
J	A	1	25	6	13	1	4	1	0	0



• Molecule 4 is water.

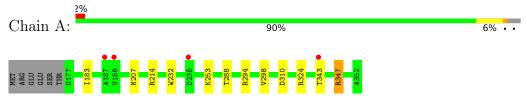
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	92	Total O 92 92	0	0



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: Genome polyprotein





4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 21 21 21	Depositor
Cell constants	49.42Å 78.24Å 163.18Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.30 - 2.10	Depositor
Resolution (A)	47.30 - 2.10	EDS
% Data completeness	99.7 (47.30-2.10)	Depositor
(in resolution range)	99.7 (47.30-2.10)	EDS
R _{merge}	0.10	Depositor
R _{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$2.22 (at 2.10 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
D D.	0.189 , 0.228	Depositor
R, R_{free}	0.198 , 0.234	DCC
R_{free} test set	949 reflections (5.02%)	wwPDB-VP
Wilson B-factor $(Å^2)$	36.1	Xtriage
Anisotropy	0.763	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.43 , 48.8	EDS
L-test for twinning ²	$ \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	0.96	EDS
Total number of atoms	2924	wwPDB-VP
Average B, all atoms $(Å^2)$	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 7.99% of the height of the origin peak. No significant pseudotranslation is detected.

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.



¹Intensities estimated from amplitudes.

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: MES, $\mathrm{SO4}$

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		
IVIOI	Unain	RMSZ	# Z > 5	RMSZ	# Z > 5	
1	А	0.81	0/1434	0.95	7/1951~(0.4%)	

There are no bond length outliers.

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$Observed(^{o})$	$Ideal(^{o})$
1	А	347	ARG	NE-CZ-NH1	6.95	123.78	120.30
1	А	294	ARG	NE-CZ-NH1	6.46	123.53	120.30
1	А	294	ARG	NE-CZ-NH2	-6.19	117.20	120.30
1	А	347	ARG	NE-CZ-NH2	-6.14	117.23	120.30
1	А	214	ARG	NE-CZ-NH1	5.90	123.25	120.30
1	А	324	ARG	NE-CZ-NH1	5.75	123.17	120.30
1	А	214	ARG	NE-CZ-NH2	-5.33	117.64	120.30

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	А	1398	1324	1318	5	0
2	А	60	0	0	2	0
3	А	24	26	26	0	0
4	А	92	0	0	1	1



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	1574	1350	1344	5	1

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 2.

All (5) 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:347:ARG:HD3	2:A:403:SO4:O4	1.86	0.75
1:A:232:TRP:CD2	1:A:253:LYS:HB2	2.45	0.51
1:A:298:VAL:HG22	1:A:347:ARG:HG2	1.94	0.50
1:A:183:ILE:HD13	4:A:563:HOH:O	2.17	0.45
1:A:207:ASN:HB3	2:A:410:SO4:O2	2.19	0.42

All (1) 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)
4:A:525:HOH:O	4:A:525:HOH:O[6_555]	1.30	0.90

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	А	174/182~(96%)	167~(96%)	7~(4%)	0	100 100	

There are no Ramachandran outliers to report.



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 side chain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	А	156/162~(96%)	153~(98%)	3~(2%)	57 63	

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	288	THR
1	А	310	ASP
1	А	343	THR

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)

14 ligands are 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



Mal	T a	Chain	Dec	Link	Bo	ond leng	ths	В	ond ang	les
Mol	Type	Chain	Res	LIIIK	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2
2	SO4	А	402	-	4,4,4	0.27	0	6,6,6	0.72	0
2	SO4	А	412	-	4,4,4	0.38	0	6,6,6	0.20	0
2	SO4	А	403	-	4,4,4	0.14	0	6,6,6	0.37	0
2	SO4	А	401	-	4,4,4	0.36	0	6,6,6	0.26	0
2	SO4	А	407	-	4,4,4	0.41	0	6,6,6	0.33	0
2	SO4	А	406	-	4,4,4	0.42	0	$6,\!6,\!6$	0.45	0
3	MES	А	414	-	12,12,12	2.07	1 (8%)	14,16,16	1.87	<mark>5 (35%)</mark>
2	SO4	А	410	-	4,4,4	0.37	0	6,6,6	0.32	0
2	SO4	А	411	-	4,4,4	0.32	0	6,6,6	0.42	0
2	SO4	А	405	-	4,4,4	0.42	0	6,6,6	0.29	0
2	SO4	А	408	-	4,4,4	0.38	0	$6,\!6,\!6$	0.28	0
2	SO4	А	409	-	4,4,4	0.46	0	6,6,6	0.30	0
3	MES	А	413	-	12,12,12	1.98	1 (8%)	14,16,16	1.75	5 (35%)
2	SO4	А	404	-	4,4,4	0.35	0	6,6,6	1.27	0

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).

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	MES	А	413	-	-	3/6/14/14	0/1/1/1
3	MES	А	414	-	-	5/6/14/14	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	А	414	MES	C8-S	-6.54	1.68	1.77
3	А	413	MES	C8-S	-6.24	1.68	1.77

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$Observed(^{o})$	$Ideal(^{o})$
3	А	413	MES	O2S-S-C8	3.78	111.46	106.92
3	А	414	MES	O1S-S-C8	3.31	110.90	106.92
3	А	414	MES	C6-C5-N4	3.29	115.09	110.10
3	А	414	MES	C2-C3-N4	2.57	114.00	110.10



501	19
-----	----

Mol	Chain	Res	Type	Atoms	Ζ	$Observed(^{o})$	$Ideal(^{o})$
3	А	414	MES	O2S-S-C8	2.54	109.98	106.92
3	А	414	MES	O3S-S-C8	2.51	109.83	105.77
3	А	413	MES	O1S-S-C8	2.50	109.93	106.92
3	А	413	MES	O3S-S-C8	2.29	109.47	105.77
3	А	413	MES	C6-C5-N4	2.07	113.24	110.10
3	А	413	MES	O3S-S-O1S	-2.05	106.28	111.27

Continued from previous page...

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	А	413	MES	C7-C8-S-O1S
3	А	414	MES	C7-C8-S-O2S
3	А	414	MES	C8-C7-N4-C3
3	А	414	MES	C8-C7-N4-C5
3	А	413	MES	C7-C8-S-O3S
3	А	413	MES	C7-C8-S-O2S
3	А	414	MES	C7-C8-S-O1S
3	A	414	MES	C7-C8-S-O3S

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	А	403	SO4	1	0
2	А	410	SO4	1	0

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^{th} 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		$\mathbf{OWAB}(\mathbf{A}^2)$	Q < 0.9	
1	А	176/182~(96%)	0.35	4 (2%)	60	65	27, 39, 61, 70	6 (3%)

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	343	THR	2.6
1	А	187	ALA	2.5
1	А	235	ASP	2.2
1	А	188	VAL	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 monosaccharides 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^{th} 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	$B-factors(Å^2)$	Q < 0.9
3	MES	А	414	12/12	0.76	0.29	98,115,125,128	0
2	SO4	А	406	5/5	0.79	0.26	79,96,97,102	0
2	SO4	А	407	5/5	0.84	0.17	97,99,107,114	0



Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
	÷ –	Chan					. ,	•
3	MES	A	413	12/12	0.85	0.21	$49,\!55,\!60,\!65$	25
2	SO4	А	405	5/5	0.86	0.17	85,88,94,109	0
2	SO4	А	412	5/5	0.89	0.32	$115,\!115,\!123,\!123$	0
2	SO4	А	411	5/5	0.90	0.22	$82,\!92,\!108,\!108$	0
2	SO4	А	408	5/5	0.90	0.18	88,90,94,98	0
2	SO4	А	410	5/5	0.91	0.13	89,94,103,113	0
2	SO4	А	409	5/5	0.94	0.37	93,94,105,106	0
2	SO4	А	402	5/5	0.96	0.13	59,65,72,78	0
2	SO4	А	404	5/5	0.96	0.13	$68,\!68,\!72,\!75$	0
2	SO4	А	401	5/5	0.96	0.23	67,70,82,90	0
2	SO4	А	403	5/5	0.98	0.09	$56,\!57,\!66,\!72$	0

Continued from previous page...

6.5 Other polymers (i)

There are no such residues in this entry.

