wwPDB X-ray Structure Validation Summary Report (i)

Aug 31, 2022 – 01:22 pm BST

PDB ID	:	7OJ4
Title	:	Crystal structure of apo NS3 helicase from tick-borne encephalitis virus
Authors	:	Anindita, P.D.; Havlickova, P.; Kascakova, B.; Grinkevich, P.; Brynda, J.; Franta, Z.
Deposited on	:	2021-05-13
Resolution	:	1.83 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*
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:

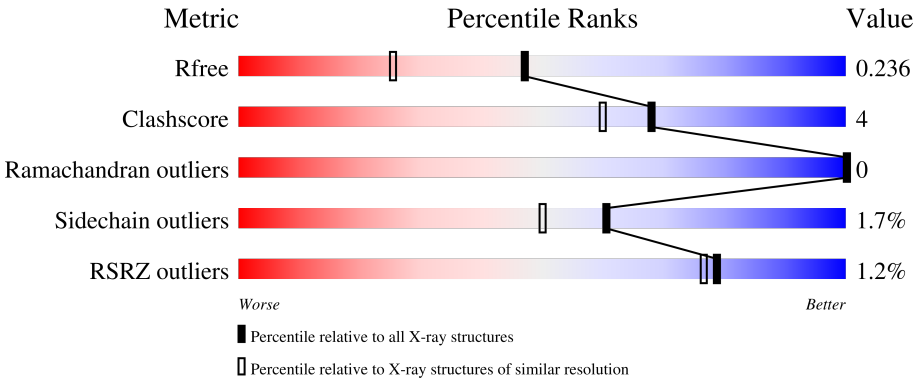
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.30
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.30

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.83 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	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RSRZ outliers	127900	7371 (1.84-1.80)

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
1	A	473	% 81% 8% · 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3582 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 NS3 helicase domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	426	3375	2106	620	635	14	0	1	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	149	MET	-	initiating methionine	UNP A0A2S1PWV0
A	150	GLY	-	expression tag	UNP A0A2S1PWV0
A	151	HIS	-	expression tag	UNP A0A2S1PWV0
A	152	HIS	-	expression tag	UNP A0A2S1PWV0
A	153	HIS	-	expression tag	UNP A0A2S1PWV0
A	154	HIS	-	expression tag	UNP A0A2S1PWV0
A	155	HIS	-	expression tag	UNP A0A2S1PWV0
A	156	HIS	-	expression tag	UNP A0A2S1PWV0
A	157	HIS	-	expression tag	UNP A0A2S1PWV0
A	158	HIS	-	expression tag	UNP A0A2S1PWV0
A	159	HIS	-	expression tag	UNP A0A2S1PWV0
A	160	HIS	-	expression tag	UNP A0A2S1PWV0
A	161	SER	-	expression tag	UNP A0A2S1PWV0
A	162	SER	-	expression tag	UNP A0A2S1PWV0
A	163	GLY	-	expression tag	UNP A0A2S1PWV0
A	164	HIS	-	expression tag	UNP A0A2S1PWV0
A	165	ILE	-	expression tag	UNP A0A2S1PWV0
A	166	ASP	-	expression tag	UNP A0A2S1PWV0
A	167	ASP	-	expression tag	UNP A0A2S1PWV0
A	168	ASP	-	expression tag	UNP A0A2S1PWV0
A	169	ASP	-	expression tag	UNP A0A2S1PWV0
A	170	LYS	-	expression tag	UNP A0A2S1PWV0
A	171	HIS	-	expression tag	UNP A0A2S1PWV0
A	172	MET	-	expression tag	UNP A0A2S1PWV0

- Molecule 2 is water.

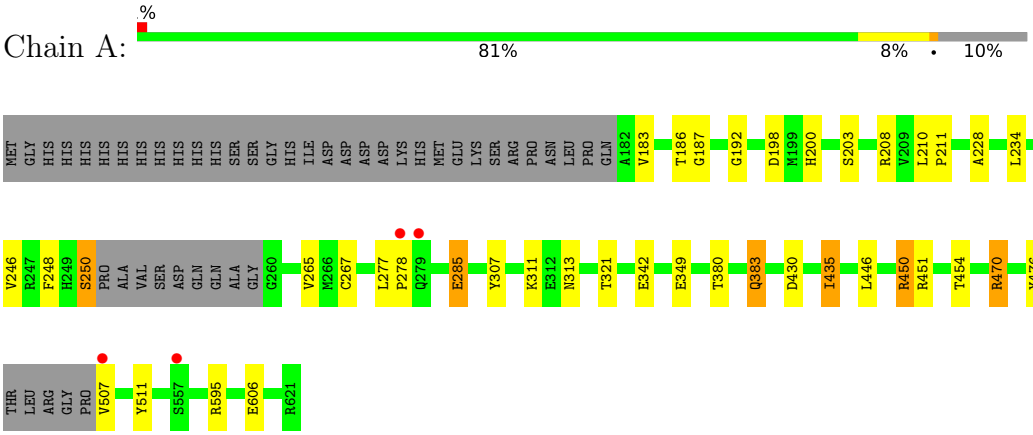
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	206	Total 207	O 207	0	1

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: NS3 helicase domain

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	73.30Å 73.30Å 196.05Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.06 – 1.83 49.01 – 1.83	Depositor EDS
% Data completeness (in resolution range)	99.8 (49.06-1.83) 99.9 (49.01-1.83)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.68 (at 1.83Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.181 , 0.229 0.185 , 0.236	Depositor DCC
R_{free} test set	2366 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	34.3	Xtriage
Anisotropy	0.356	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3582	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.13% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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	0.94	5/3448 (0.1%)	0.99	6/4670 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	250	SER	C-O	18.60	1.58	1.23
1	A	349	GLU	CD-OE1	8.44	1.34	1.25
1	A	383[A]	GLN	CD-OE1	6.48	1.38	1.24
1	A	383[B]	GLN	CD-OE1	6.48	1.38	1.24
1	A	606	GLU	CD-OE1	6.33	1.32	1.25

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	450	ARG	CB-CA-C	-6.81	96.79	110.40
1	A	450	ARG	CB-CG-CD	-6.55	94.56	111.60
1	A	250	SER	CA-C-O	-5.50	108.54	120.10
1	A	595	ARG	NE-CZ-NH1	5.49	123.04	120.30
1	A	285	GLU	CB-CG-CD	-5.48	99.40	114.20

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	511	TYR	Peptide

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	3375	0	3294	25	0
2	A	207	0	0	1	0
All	All	3582	0	3294	25	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.

The worst 5 of 25 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:186:THR:HG22	1:A:187:GLY:H	1.06	1.06
1:A:186:THR:HG22	1:A:187:GLY:N	1.85	0.91
1:A:186:THR:CG2	1:A:187:GLY:H	1.84	0.90
1:A:234:LEU:CD1	1:A:265:VAL:HG12	2.26	0.66
1:A:454:THR:HG23	1:A:482:ASP:OD1	1.97	0.65

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	421/473 (89%)	408 (97%)	13 (3%)	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	356/400 (89%)	350 (98%)	6 (2%)	60 50

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	435	ILE
1	A	470	ARG
1	A	507	VAL
1	A	200	HIS
1	A	183	VAL

Sometimes 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 monosaccharides in this entry.

5.6 Ligand geometry

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^{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			OWAB(Å^2)	Q<0.9
1	A	426/473 (90%)	-0.05	5 (1%)	79	76	26, 42, 80, 109	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	507	VAL	5.1
1	A	279	GLN	4.4
1	A	278	PRO	2.4
1	A	557	SER	2.3
1	A	243	GLY	2.1

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

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.