Full wwPDB X-ray Structure Validation Report ⓘ

Nov 6, 2023 – 11:46 pm GMT

PDB ID	:	7PGC
Title	:	Crystal Structure of Unlinked NS2B-NS3 Protease from Zika Virus in Complex with Inhibitor MI-2191
Authors	:	Huber, S.; Heine, A.; Steinmetzer, T.
Deposited on	:	2021-08-13
Resolution	:	1.55 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

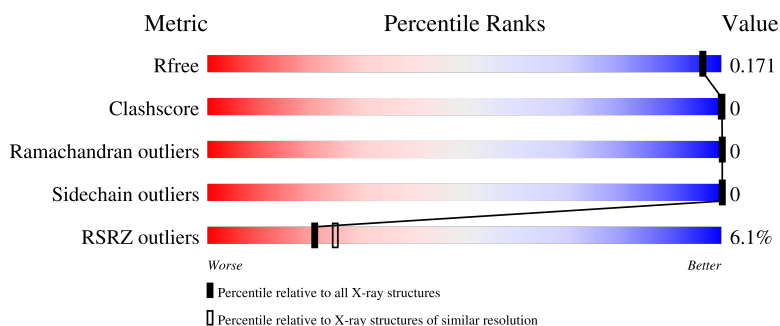|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.55 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	Percentile Ranks	Value
Rfree		0.171
Clashscore		0
Ramachandran outliers		0
Sidechain outliers		0
RSRZ outliers		6.1%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1483 (1.56-1.56)
Clashscore	141614	1529 (1.56-1.56)
Ramachandran outliers	138981	1498 (1.56-1.56)
Sidechain outliers	138945	1495 (1.56-1.56)
RSRZ outliers	127900	1465 (1.56-1.56)

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	53	6% / 74% / 26%
2	B	178	5% / 87% / 13%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	HMF	C	2	-	X	-	-

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 3001 atoms, of which 1333 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 NS2B.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	39	Total	C	H	N	O	S	0	2	0
			554	183	259	48	63	1			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	44	MET	-	initiating methionine	UNP Q32ZE1
A	45	THR	-	expression tag	UNP Q32ZE1

- Molecule 2 is a protein called Serine protease NS3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	155	Total	C	H	N	O	S	0	2	0
			2202	713	1074	192	218	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	GLY	-	expression tag	UNP Q32ZE1
B	107	LYS	ARG	conflict	UNP Q32ZE1

- Molecule 3 is a protein called Inhibitor MI-2191.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	5	Total	C	N	O	0	0	0
			53	38	10	5			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	47	Total	O	0	1
			48	48		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	134	Total 135	O 135	0	1
4	C	9	Total 9	O 9	0	0

SEQUENCE-PLOTS INFOmissingINFO

3   Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	48.94Å 60.15Å 82.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.96 – 1.55 37.96 – 1.55	Depositor EDS
% Data completeness (in resolution range)	99.4 (37.96-1.55) 99.4 (37.96-1.55)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.04	Depositor
$< I/\sigma(I) >$ [1]	1.21 (at 1.55Å)	Xtriage
Refinement program	PHENIX 1.16_3549	Depositor
R, R_{free}	0.145 , 0.171 0.145 , 0.171	Depositor DCC
R_{free} test set	1790 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	22.7	Xtriage
Anisotropy	0.334	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 59.0	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	3001	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.26% 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.

4 Model quality ⓘ

4.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: HMF, V7T, V8N

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	0.38	0/311	0.66	0/422
2	B	0.39	0/1159	0.64	0/1580
3	C	0.35	0/17	0.68	0/19
All	All	0.39	0/1487	0.65	0/2021

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

4.2 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	295	259	249	0	0
2	B	1128	1074	1066	0	0
3	C	53	0	36	0	0
4	A	48	0	0	0	0
4	B	135	0	0	0	0
4	C	9	0	0	0	0
All	All	1668	1333	1351	0	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 0.

There are no clashes within the asymmetric unit.

There are no symmetry-related clashes.

4.3 Torsion angles (i)

4.3.1 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	39/53 (74%)	37 (95%)	2 (5%)	0	100	100
2	B	155/178 (87%)	149 (96%)	6 (4%)	0	100	100
3	C	1/5 (20%)	1 (100%)	0	0	100	100
All	All	195/236 (83%)	187 (96%)	8 (4%)	0	100	100

There are no Ramachandran outliers to report.

4.3.2 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	30/45 (67%)	30 (100%)	0	100	100
2	B	112/140 (80%)	112 (100%)	0	100	100
3	C	2/2 (100%)	2 (100%)	0	100	100
All	All	144/187 (77%)	144 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

4.3.3 RNA ⓘ

There are no RNA molecules in this entry.

4.4 Non-standard residues in protein, DNA, RNA chains ⓘ

3 non-standard protein/DNA/RNA residues 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 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	V8N	C	3	3	11,11,12	0.39	0	13,13,15	0.53	0
3	HMF	C	2	3	11,12,13	4.59	6 (54%)	9,14,16	4.89	8 (88%)
3	V7T	C	1	3	10,11,12	0.65	0	9,12,14	2.03	2 (22%)

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	V8N	C	3	3	-	0/5/5/6	0/1/1/1
3	HMF	C	2	3	-	0/6/7/9	0/1/1/1
3	V7T	C	1	3	-	1/10/11/13	-

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	2	HMF	C3-C2	6.82	1.53	1.38
3	C	2	HMF	C5-C6	6.68	1.53	1.38
3	C	2	HMF	C6-C1	6.52	1.52	1.38
3	C	2	HMF	C2-C1	6.41	1.52	1.38
3	C	2	HMF	C4-C3	5.30	1.52	1.38
3	C	2	HMF	C4-C5	5.15	1.51	1.38

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	2	HMF	C4-C3-C2	-5.93	111.16	120.19
3	C	2	HMF	C4-C5-C6	-5.86	111.26	120.19
3	C	2	HMF	C6-C1-C2	-5.65	109.29	118.17
3	C	2	HMF	C3-C2-C1	-5.62	112.01	120.63
3	C	2	HMF	C5-C6-C1	-5.36	112.40	120.63
3	C	2	HMF	C5-C4-C3	-5.02	110.61	119.93
3	C	1	V7T	N-C2-N1	4.72	128.82	120.59
3	C	2	HMF	CG-C1-C6	-3.89	111.40	121.23
3	C	2	HMF	CG-C1-C2	-3.67	111.95	121.23
3	C	1	V7T	N2-C2-N	-3.13	112.54	119.55

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	1	V7T	O-C-CA-CB

There are no ring outliers.

No monomer is involved in short contacts.

4.5 Carbohydrates (i)

There are no monosaccharides in this entry.

4.6 Ligand geometry (i)

There are no ligands in this entry.

4.7 Other polymers (i)

There are no such residues in this entry.

4.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

5 Fit of model and data ⓘ

5.1 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, 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	39/53 (73%)	0.25	3 (7%)	13	16	22, 28, 52, 56	0
2	B	155/178 (87%)	0.18	9 (5%)	23	26	18, 24, 41, 68	0
3	C	2/5 (40%)	0.15	0	100	100	21, 21, 21, 25	0
All	All	196/236 (83%)	0.20	12 (6%)	21	25	18, 26, 47, 68	0

All (12) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	30	LEU	10.1
1	A	66	GLU	6.5
2	B	31	LEU	6.1
1	A	49	VAL	5.0
2	B	18	THR	4.8
2	B	29	ARG	3.9
2	B	32	GLY	3.7
2	B	17	GLU	3.2
2	B	16	GLY	3.2
2	B	170	ARG	2.9
1	A	65	ALA	2.8
2	B	61	GLY	2.7

5.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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(Å^2)	Q<0.9
3	HMF	C	2	12/13	0.95	0.12	18,25,34,35	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
3	V7T	C	1	12/13	0.97	0.16	17,18,20,21	0
3	V8N	C	3	11/12	0.97	0.10	19,19,19,19	0

5.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.4 Ligands ⓘ

There are no ligands in this entry.

5.5 Other polymers ⓘ

There are no such residues in this entry.