



Full wwPDB X-ray Structure Validation Report ⓘ

Jul 20, 2017 – 06:16 AM EDT

PDB ID : 5VP9
Title : Crystal structure of HCV NS3/4A protease in complex with AM-07, an analogue of 5172-mcP1P3
Authors : Matthew, A.N.; Schiffer, C.A.
Deposited on : unknown
Resolution : 1.86 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

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20029824
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

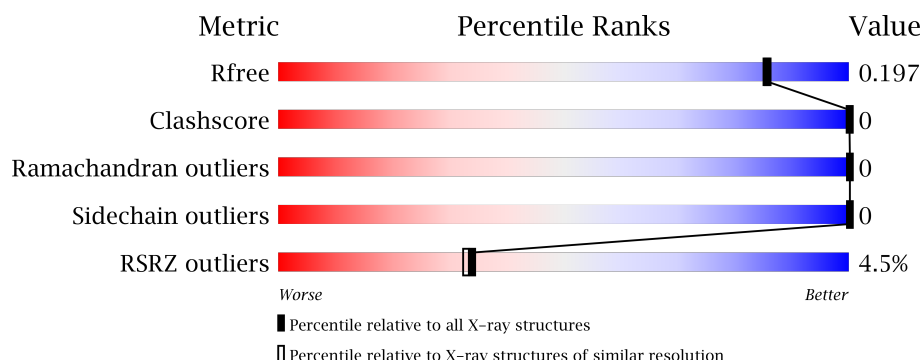
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.86 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}	100719	1923 (1.86-1.86)
Clashscore	112137	2083 (1.86-1.86)
Ramachandran outliers	110173	2060 (1.86-1.86)
Sidechain outliers	110143	2060 (1.86-1.86)
RSRZ outliers	101464	1932 (1.86-1.86)

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 ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	203	<div> <div>4%</div> <div>97%</div> <div>..</div> </div>

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
2	9H7	A	1201	-	-	-	X

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3192 atoms, of which 1511 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.

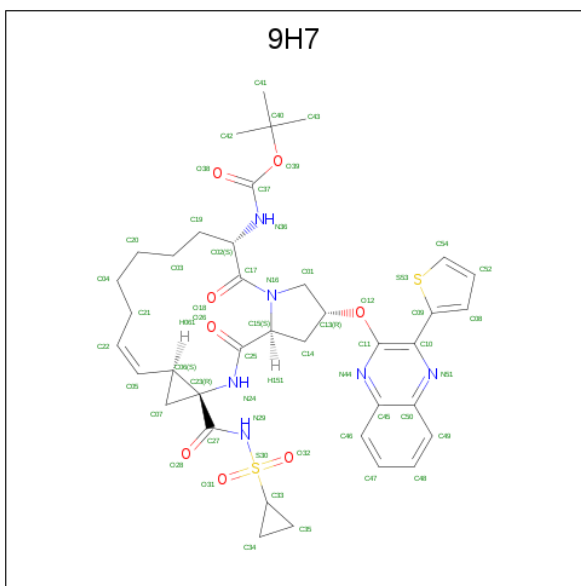
- Molecule 1 is a protein called NS4A cofactor – NS3 protein chimera.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	198	Total	C	H	N	O	S	0	1	0
			2930	905	1466	267	284	8			

There are 27 discrepancies between the modelled and reference sequences:

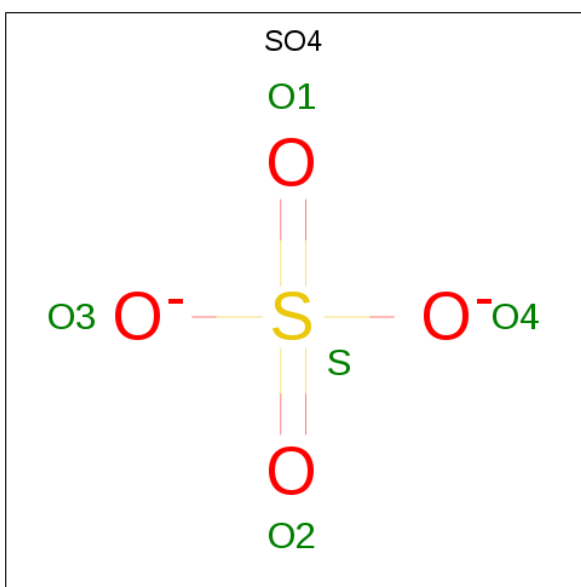
Chain	Residue	Modelled	Actual	Comment	Reference
A	980	GLY	-	expression tag	UNP A8DG50
A	981	SER	-	expression tag	UNP A8DG50
A	982	HIS	-	expression tag	UNP A8DG50
A	983	MET	-	expression tag	UNP A8DG50
A	984	ALA	-	expression tag	UNP A8DG50
A	985	SER	-	expression tag	UNP A8DG50
A	986	MET	-	expression tag	UNP A8DG50
A	987	LYS	-	expression tag	UNP A8DG50
A	988	LYS	-	expression tag	UNP A8DG50
A	989	LYS	-	expression tag	UNP A8DG50
A	991	SER	CYS	engineered mutation	UNP A8DG50
A	998	ILE	VAL	engineered mutation	UNP A8DG50
A	999	ASN	ILE	engineered mutation	UNP A8DG50
A	1001	SER	-	linker	UNP A8DG50
A	1002	GLY	-	linker	UNP A8DG50
A	1003	ASP	ILE	engineered mutation	UNP A8DG50
A	1013	GLU	LEU	engineered mutation	UNP A8DG50
A	1014	GLU	LEU	engineered mutation	UNP A8DG50
A	1017	GLN	ILE	engineered mutation	UNP A8DG50
A	1018	GLU	ILE	engineered mutation	UNP A8DG50
A	1021	GLN	LEU	engineered mutation	UNP A8DG50
A	1040	THR	ALA	engineered mutation	UNP A8DG50
A	1047	SER	CYS	engineered mutation	UNP A8DG50
A	1052	LEU	CYS	engineered mutation	UNP A8DG50
A	1072	THR	ILE	engineered mutation	UNP A8DG50
A	1086	GLN	PRO	engineered mutation	UNP A8DG50
A	1159	SER	CYS	engineered mutation	UNP A8DG50

- Molecule 2 is tert-butyl [(2R,6S,12Z,13aS,14aR,16aS)-14a-[(cyclopropylsulfonyl)carbamoyl]-5,16-dioxo-2-{[3-(thiophen-2-yl)quinoxalin-2-yl]oxy}-1,2,3,5,6,7,8,9,10,11,13a,14,14a,15,16,16a-hexadecahydrocyclopropa[e]pyrrolo[1,2-a][1,4]diazacyclopentadecin-6-yl]carbamate (three-letter code: 9H7) (formula: C₃₈H₄₆N₆O₈S₂).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	S	0	0
			99	38	45	6	8	2		

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Zn	0	0
			1	1		

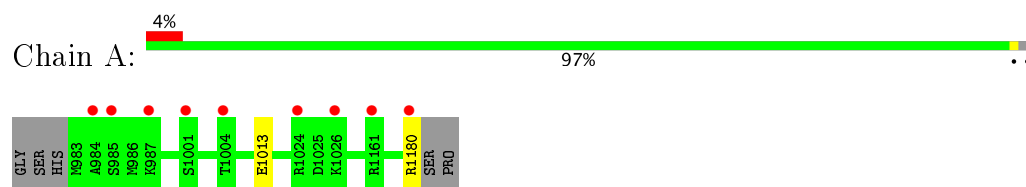
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	157	Total	O	0	0
			157	157		

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: NS4A cofactor – NS3 protein chimera



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	54.79Å 58.35Å 60.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.99 – 1.86 20.99 – 1.86	Depositor EDS
% Data completeness (in resolution range)	99.8 (20.99-1.86) 94.7 (20.99-1.86)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.37 (at 1.86Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.168 , 0.208 0.161 , 0.197	Depositor DCC
R_{free} test set	795 reflections (4.99%)	DCC
Wilson B-factor (Å ²)	19.5	Xtriage
Anisotropy	0.551	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.47 , 56.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.014 for -h,l,k	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	3192	wwPDB-VP
Average B, all atoms (Å ²)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.22% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, SO4, 9H7

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	0.30	0/1488	0.55	0/2019

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

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	1464	1466	1466	0	1
2	A	54	45	0	0	0
3	A	5	0	0	0	0
4	A	1	0	0	0	0
5	A	157	0	0	0	1
All	All	1681	1511	1466	0	2

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.

All (2) 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)
1:A:1013:GLU:OE2	1:A:1180:ARG:NH2[3_545]	2.18	0.02
5:A:1301:HOH:O	5:A:1419:HOH:O[4_445]	2.18	0.02

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	197/203 (97%)	195 (99%)	2 (1%)	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 sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	159/165 (96%)	159 (100%)	0	100	100

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

Some 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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 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	$\# Z > 2$
2	9H7	A	1201	-	58,60,60	5.25	35 (60%)	70,90,90	2.38	12 (17%)
3	SO4	A	1202	-	4,4,4	0.21	0	6,6,6	0.12	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
2	9H7	A	1201	-	-	0/55/83/83	0/4/7/7
3	SO4	A	1202	-	-	0/0/0/0	0/0/0/0

All (35) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1201	9H7	C09-S53	-16.87	1.56	1.72
2	A	1201	9H7	C14-C15	-10.06	1.33	1.53
2	A	1201	9H7	C01-C13	-9.56	1.36	1.52
2	A	1201	9H7	C54-S53	-2.84	1.56	1.70
2	A	1201	9H7	O39-C40	-2.63	1.43	1.48
2	A	1201	9H7	O26-C25	-2.09	1.19	1.23

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1201	9H7	C34-C33	2.05	1.53	1.50
2	A	1201	9H7	C35-C33	2.21	1.53	1.50
2	A	1201	9H7	C10-C09	2.35	1.54	1.49
2	A	1201	9H7	C23-N24	2.75	1.50	1.45
2	A	1201	9H7	O31-S30	3.00	1.47	1.43
2	A	1201	9H7	S30-N29	3.25	1.70	1.62
2	A	1201	9H7	O32-S30	3.29	1.47	1.43
2	A	1201	9H7	O39-C37	3.76	1.42	1.34
2	A	1201	9H7	C50-C45	4.65	1.52	1.42
2	A	1201	9H7	C37-N36	4.91	1.47	1.34
2	A	1201	9H7	C06-C05	5.80	1.56	1.49
2	A	1201	9H7	C25-N24	5.84	1.46	1.34
2	A	1201	9H7	C45-N44	6.36	1.48	1.37
2	A	1201	9H7	C50-N51	6.39	1.48	1.37
2	A	1201	9H7	O12-C11	6.41	1.42	1.35
2	A	1201	9H7	C11-N44	6.41	1.46	1.30
2	A	1201	9H7	C01-N16	6.55	1.58	1.47
2	A	1201	9H7	C52-C08	6.61	1.61	1.39
2	A	1201	9H7	C48-C47	6.86	1.54	1.38
2	A	1201	9H7	C17-N16	7.01	1.49	1.34
2	A	1201	9H7	C14-C13	7.03	1.68	1.52
2	A	1201	9H7	C10-C11	7.24	1.56	1.43
2	A	1201	9H7	C48-C49	7.39	1.53	1.36
2	A	1201	9H7	C47-C46	7.41	1.54	1.36
2	A	1201	9H7	C15-N16	7.71	1.62	1.47
2	A	1201	9H7	C49-C50	7.94	1.56	1.41
2	A	1201	9H7	C10-N51	8.18	1.46	1.32
2	A	1201	9H7	C46-C45	8.27	1.56	1.41
2	A	1201	9H7	O28-C27	8.76	1.38	1.22

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1201	9H7	C10-C11-N44	-5.15	119.12	124.03
2	A	1201	9H7	C06-C23-N24	-5.11	107.26	117.61
2	A	1201	9H7	C52-C54-S53	-4.98	107.82	113.18
2	A	1201	9H7	C06-C23-C27	-4.17	109.27	117.08
2	A	1201	9H7	C25-C15-N16	-4.05	102.26	112.59
2	A	1201	9H7	C01-N16-C15	-3.59	106.25	111.69
2	A	1201	9H7	O38-C37-N36	-2.94	119.80	124.87
2	A	1201	9H7	C11-N44-C45	2.07	120.42	116.30
2	A	1201	9H7	C14-C13-C01	2.16	105.53	103.63

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
2	A	1201	9H7	C23-C27-N29	3.53	121.83	115.64
2	A	1201	9H7	O39-C37-N36	4.37	117.89	110.06
2	A	1201	9H7	C07-C23-N24	13.43	135.78	117.85

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 [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, 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	198/203 (97%)	-0.02	9 (4%) 34 33	12, 20, 42, 53	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	984	ALA	3.6
1	A	1161	ARG	3.3
1	A	1180	ARG	2.8
1	A	1024	ARG	2.7
1	A	1004	THR	2.7
1	A	987	LYS	2.3
1	A	1001	SER	2.1
1	A	1026	LYS	2.1
1	A	985	SER	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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
2	9H7	A	1201	54/54	0.88	0.22	2.39	15,31,49,53	0
3	SO4	A	1202	5/5	0.98	0.09	-0.02	24,25,28,28	0
4	ZN	A	1203	1/1	0.99	0.04	-2.92	22,22,22,22	0

6.5 Other polymers [i](#)

There are no such residues in this entry.