Full wwPDB NMR Structure Validation Report (i)

Oct 31, 2021 – 12:32 AM EDT

PDB ID	:	2JY0
Title	:	Solution NMR structure of HCV NS2 protein, membrane segment (1-27)
Authors	:	Montserret, R.; Penin, F.
Deposited on	:	2007-12-04

This is a Full wwPDB NMR 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

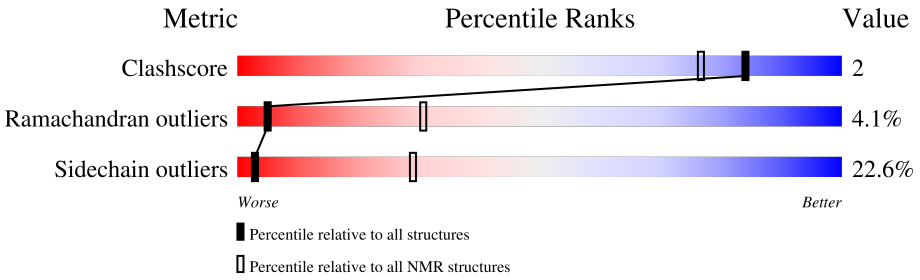
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 74%.

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)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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%

Mol	Chain	Length	Quality of chain
1	A	27	81% 15% •

2 Ensemble composition and analysis (i)

This entry contains 40 models.

Cyrange was unable to find well-defined residues.

Error message: Only domains with < 8 residues could be identified.

NmrClust was unable to cluster the ensemble.

Error message: Wrapper check: not enough residues in core to run NmrClust

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 403 atoms, of which 205 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protease NS2-3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	27	403	127	205	33	36	2	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	ALA	CYS	engineered mutation	UNP Q9WMX2
A	18	VAL	ILE	SEE REMARK 999	UNP Q9WMX2

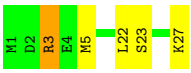
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: Protease NS2-3

Chain A:



4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

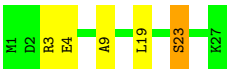
• Molecule 1: Protease NS2-3

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: Protease NS2-3

Chain A:



4.2.3 Score per residue for model 3

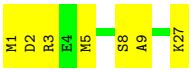
- Molecule 1: Protease NS2-3

Chain A: 74% 22% •

4.2.4 Score per residue for model 4

- Molecule 1: Protease NS2-3

Chain A: 74% 26%



4.2.5 Score per residue for model 5

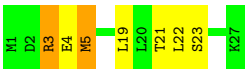
- Molecule 1: Protease NS2-3

Chain A: 67% 26% 7%

4.2.6 Score per residue for model 6

- Molecule 1: Protease NS2-3

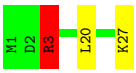
Chain A: 74% 19% 7%



4.2.7 Score per residue for model 7

- Molecule 1: Protease NS2-3

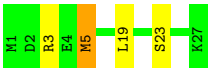
Chain A: 89% 7% •



4.2.8 Score per residue for model 8

- Molecule 1: Protease NS2-3

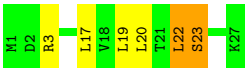
Chain A: 85% 11% •



4.2.9 Score per residue for model 9

- Molecule 1: Protease NS2-3

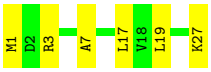
Chain A: 78% 15% 7%



4.2.10 Score per residue for model 10

- Molecule 1: Protease NS2-3

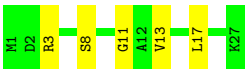
Chain A: 78% 22%



4.2.11 Score per residue for model 11

- Molecule 1: Protease NS2-3

Chain A: 81% 19%



4.2.12 Score per residue for model 12

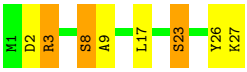
- Molecule 1: Protease NS2-3

Chain A: 81% 15% •

4.2.13 Score per residue for model 13

- Molecule 1: Protease NS2-3

Chain A: 70% 19% 11%



4.2.14 Score per residue for model 14

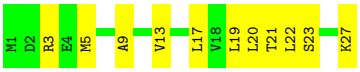
- Molecule 1: Protease NS2-3

Chain A: 85% 15%

4.2.15 Score per residue for model 15

- Molecule 1: Protease NS2-3

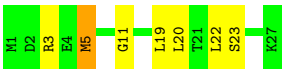
Chain A: 59% 41%



4.2.16 Score per residue for model 16

- Molecule 1: Protease NS2-3

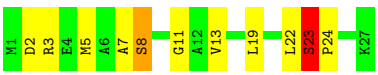
Chain A: 74% 22% •



4.2.17 Score per residue for model 17

- Molecule 1: Protease NS2-3

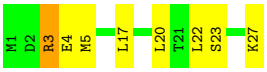
Chain A: 59% 33% • •



4.2.18 Score per residue for model 18

- Molecule 1: Protease NS2-3

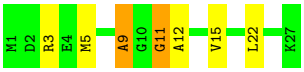
Chain A: 70% 26% •



4.2.19 Score per residue for model 19

- Molecule 1: Protease NS2-3

Chain A: 74% 19% 7%



4.2.20 Score per residue for model 20

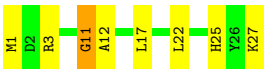
- Molecule 1: Protease NS2-3

Chain A: 67% 26% 7%

4.2.21 Score per residue for model 21

- Molecule 1: Protease NS2-3

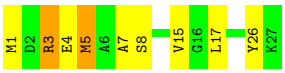
Chain A: 70% 26% •



4.2.22 Score per residue for model 22

- Molecule 1: Protease NS2-3

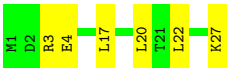
Chain A: 67% 26% 7%



4.2.23 Score per residue for model 23

- Molecule 1: Protease NS2-3

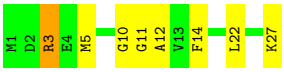
Chain A: 78% 22%



4.2.24 Score per residue for model 24

- Molecule 1: Protease NS2-3

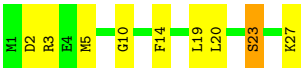
Chain A: 70% 26% •



4.2.25 Score per residue for model 25

- Molecule 1: Protease NS2-3

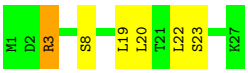
Chain A: 67% 30% •



4.2.26 Score per residue for model 26

- Molecule 1: Protease NS2-3

Chain A: 78% 19% •



4.2.27 Score per residue for model 27

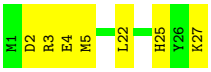
- Molecule 1: Protease NS2-3

Chain A: 81% 19%

4.2.28 Score per residue for model 28

- Molecule 1: Protease NS2-3

Chain A: 74% 26%



4.2.29 Score per residue for model 29

- Molecule 1: Protease NS2-3

Chain A: 67% 30% •

4.2.30 Score per residue for model 30

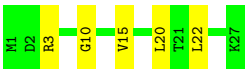
- Molecule 1: Protease NS2-3

Chain A: 78% 19% •

4.2.31 Score per residue for model 31

- Molecule 1: Protease NS2-3

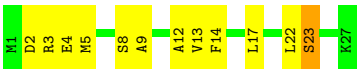
Chain A: 81% 19%



4.2.32 Score per residue for model 32

- Molecule 1: Protease NS2-3

Chain A: 56% 41% •



4.2.33 Score per residue for model 33

- Molecule 1: Protease NS2-3

Chain A: 81% 15% •

4.2.34 Score per residue for model 34

- Molecule 1: Protease NS2-3

Chain A: 74% 22% •

4.2.35 Score per residue for model 35

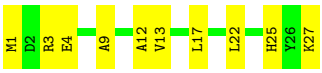
- Molecule 1: Protease NS2-3

Chain A: 67% 30% •

4.2.36 Score per residue for model 36

- Molecule 1: Protease NS2-3

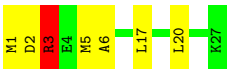
Chain A: 63% 37%



4.2.37 Score per residue for model 37

- Molecule 1: Protease NS2-3

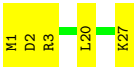
Chain A: 74% 22% •



4.2.38 Score per residue for model 38

- Molecule 1: Protease NS2-3

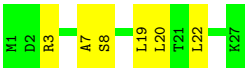
Chain A: 81% 19%



4.2.39 Score per residue for model 39

- Molecule 1: Protease NS2-3

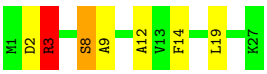
Chain A: 78% 22%



4.2.40 Score per residue for model 40

- Molecule 1: Protease NS2-3

Chain A: 74% 19%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing*.

Of the 50 calculated structures, 40 were deposited, based on the following criterion: *structures with the least restraint violations*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
X-PLOR NIH	structure solution	
X-PLOR NIH	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	270
Number of shifts mapped to atoms	270
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	74%

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

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	Planarity
1	A	0.0±0.0	1.0±0.0
All	All	0	40

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	3	ARG	Sidechain	40

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	198	205	205	1±1
All	All	7920	8200	8200	37

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 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:7:ALA:HB1	1:A:11:GLY:HA2	0.71	1.62	35	1
1:A:9:ALA:HB3	1:A:15:VAL:HG22	0.65	1.69	19	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:22:LEU:HD13	1:A:23:SER:N	0.64	2.08	9	1
1:A:19:LEU:HA	1:A:22:LEU:HD12	0.63	1.69	15	1
1:A:22:LEU:HD22	1:A:22:LEU:C	0.57	2.19	9	1
1:A:3:ARG:HG2	1:A:6:ALA:HB2	0.55	1.79	37	1
1:A:10:GLY:HA3	1:A:14:PHE:CB	0.51	2.35	24	1
1:A:8:SER:CB	1:A:13:VAL:HG23	0.51	2.36	11	1
1:A:8:SER:O	1:A:9:ALA:HB3	0.51	2.06	13	2
1:A:7:ALA:CA	1:A:15:VAL:HG21	0.50	2.36	22	1
1:A:10:GLY:HA3	1:A:15:VAL:HG23	0.50	1.84	31	1
1:A:11:GLY:O	1:A:12:ALA:HB3	0.48	2.07	19	3
1:A:9:ALA:HB3	1:A:15:VAL:CG2	0.48	2.38	19	1
1:A:19:LEU:O	1:A:22:LEU:HD12	0.47	2.10	9	1
1:A:8:SER:HB3	1:A:14:PHE:CB	0.47	2.40	40	1
1:A:7:ALA:HA	1:A:15:VAL:HG21	0.45	1.87	22	1
1:A:10:GLY:CA	1:A:14:PHE:HB3	0.45	2.42	25	1
1:A:21:THR:HA	1:A:26:TYR:CE1	0.45	2.47	29	1
1:A:8:SER:C	1:A:14:PHE:HB2	0.44	2.33	32	1
1:A:8:SER:HB2	1:A:13:VAL:HG23	0.44	1.89	17	1
1:A:10:GLY:HA2	1:A:14:PHE:HB2	0.44	1.90	5	1
1:A:19:LEU:HD23	1:A:19:LEU:O	0.44	2.13	27	1
1:A:9:ALA:O	1:A:13:VAL:HG23	0.43	2.14	15	1
1:A:9:ALA:CB	1:A:13:VAL:HB	0.43	2.44	36	1
1:A:7:ALA:HB1	1:A:13:VAL:HG21	0.42	1.90	17	1
1:A:11:GLY:O	1:A:15:VAL:HG23	0.42	2.14	29	1
1:A:9:ALA:HB1	1:A:13:VAL:HB	0.42	1.91	32	1
1:A:9:ALA:HB3	1:A:14:PHE:HB2	0.42	1.91	40	1
1:A:22:LEU:HD13	1:A:23:SER:H	0.42	1.75	9	1
1:A:7:ALA:HA	1:A:11:GLY:HA3	0.41	1.90	30	1
1:A:23:SER:CB	1:A:24:PRO:HD2	0.41	2.46	17	1
1:A:18:VAL:O	1:A:21:THR:HG22	0.41	2.16	20	1
1:A:10:GLY:HA3	1:A:14:PHE:HB3	0.41	1.91	25	1
1:A:23:SER:CB	1:A:24:PRO:CD	0.41	2.98	17	1

6.3 Torsion angles (i)

6.3.1 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 PDB entries followed by that with respect to all NMR entries. 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	25/27 (93%)	20±2 (80±6%)	4±1 (16±6%)	1±1 (4±3%)	5 31
All	All	1000/1080 (93%)	802 (80%)	157 (16%)	41 (4%)	5 31

All 10 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	23	SER	8
1	A	5	MET	7
1	A	11	GLY	6
1	A	2	ASP	4
1	A	3	ARG	4
1	A	12	ALA	4
1	A	9	ALA	2
1	A	10	GLY	2
1	A	8	SER	2
1	A	7	ALA	2

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	20/20 (100%)	15±2 (77±8%)	5±2 (23±8%)	3 29
All	All	800/800 (100%)	619 (77%)	181 (23%)	3 29

All 16 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	27	LYS	20
1	A	22	LEU	19
1	A	23	SER	17
1	A	3	ARG	17
1	A	17	LEU	17
1	A	20	LEU	17
1	A	5	MET	15
1	A	19	LEU	12

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	1	MET	11
1	A	4	GLU	11
1	A	2	ASP	9
1	A	8	SER	6
1	A	26	TYR	3
1	A	21	THR	3
1	A	25	HIS	3
1	A	14	PHE	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 74% for the well-defined parts and 74% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	270
Number of shifts mapped to atoms	270
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 74%, i.e. 225 atoms were assigned a chemical shift out of a possible 305. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	74/133 (56%)	52/53 (98%)	22/54 (41%)	0/26 (0%)
Sidechain	130/148 (88%)	80/86 (93%)	50/58 (86%)	0/4 (0%)
Aromatic	21/24 (88%)	11/13 (85%)	10/10 (100%)	0/1 (0%)
Overall	225/305 (74%)	143/152 (94%)	82/122 (67%)	0/31 (0%)

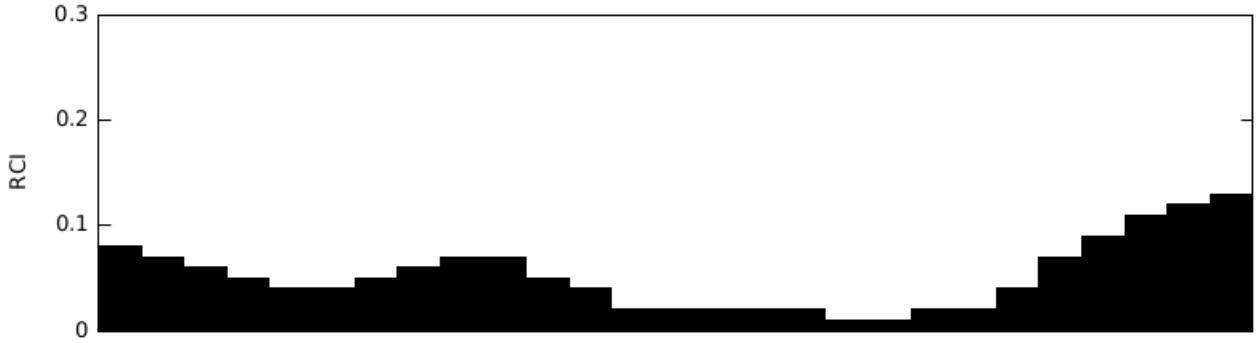
The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 74%, i.e. 225 atoms were assigned a chemical shift out of a possible 305. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	74/133 (56%)	52/53 (98%)	22/54 (41%)	0/26 (0%)
Sidechain	130/148 (88%)	80/86 (93%)	50/58 (86%)	0/4 (0%)
Aromatic	21/24 (88%)	11/13 (85%)	10/10 (100%)	0/1 (0%)
Overall	225/305 (74%)	143/152 (94%)	82/122 (67%)	0/31 (0%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A: