Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 10:15 PM BST

PDB ID	:	2JXF
Title	:	The solution structure of HCV NS4B(40-69)
Authors	:	Montserret, R.; Penin, F.
Deposited on	:	2007-11-19

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
http://wwpdb.org/validation/2016/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:

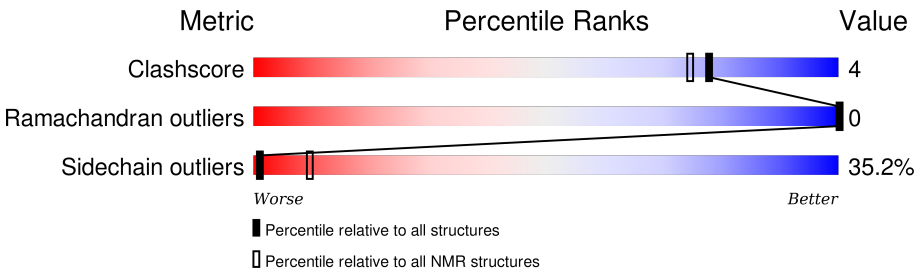
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	30	27% 17% 57%

2 Ensemble composition and analysis (i)

This entry contains 30 models. Model 24 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest rmsd deviation from the ensemble.*

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:11-A:23 (13)	0.15	24

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 4 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 3, 4, 11, 12, 13, 19, 20, 21, 22, 24, 29, 30
2	2, 6, 7, 9, 10, 14, 15, 17, 18, 23, 25, 26
3	5, 27, 28
4	8, 16

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 503 atoms, of which 247 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Genome polyprotein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	30	503	170	247	42	43	1	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: Genome polyprotein

Chain A:

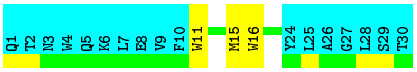
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Genome polyprotein

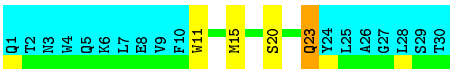
Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: Genome polyprotein

Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Genome polyprotein

Chain A: 27% 17% 57%

4.2.4 Score per residue for model 4

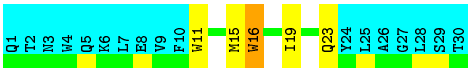
- Molecule 1: Genome polyprotein

Chain A: 23% 13% 7% 57%

4.2.5 Score per residue for model 5

- Molecule 1: Genome polyprotein

Chain A: 27% 13% • 57%



4.2.6 Score per residue for model 6

- Molecule 1: Genome polyprotein

Chain A: 30% 10% • 57%

4.2.7 Score per residue for model 7

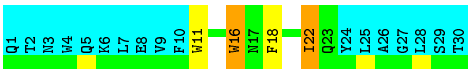
- Molecule 1: Genome polyprotein

Chain A: 33% 10% 57%

4.2.8 Score per residue for model 8

- Molecule 1: Genome polyprotein

Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Genome polyprotein

Chain A:

4.2.10 Score per residue for model 10

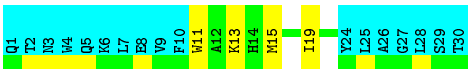
- Molecule 1: Genome polyprotein

Chain A:

4.2.11 Score per residue for model 11

- Molecule 1: Genome polyprotein

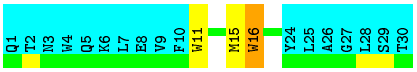
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Genome polyprotein

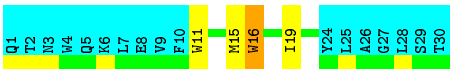
Chain A:



4.2.13 Score per residue for model 13

- Molecule 1: Genome polyprotein

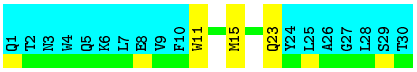
Chain A: 30% 10% • 57%



4.2.14 Score per residue for model 14

- Molecule 1: Genome polyprotein

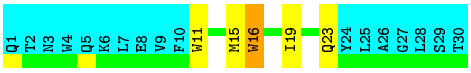
Chain A: 33% 10% 57%



4.2.15 Score per residue for model 15

- Molecule 1: Genome polyprotein

Chain A: 27% 13% • 57%



4.2.16 Score per residue for model 16

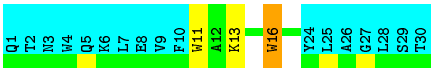
- Molecule 1: Genome polyprotein

Chain A: 30% 10% • 57%

4.2.17 Score per residue for model 17

- Molecule 1: Genome polyprotein

Chain A: 33% 7% • 57%



4.2.18 Score per residue for model 18

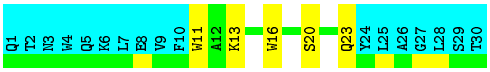
- Molecule 1: Genome polyprotein

Chain A: 33% 7% • 57%

4.2.19 Score per residue for model 19

- Molecule 1: Genome polyprotein

Chain A: 27% 17% 57%



4.2.20 Score per residue for model 20

- Molecule 1: Genome polyprotein

Chain A: 30% 13% 57%

4.2.21 Score per residue for model 21

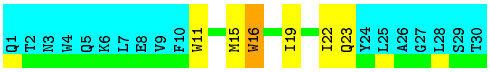
- Molecule 1: Genome polyprotein

Chain A: 33% 10% 57%

4.2.22 Score per residue for model 22

- Molecule 1: Genome polyprotein

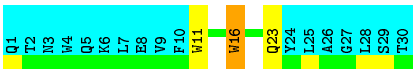
Chain A: 23% 17% • 57%



4.2.23 Score per residue for model 23

- Molecule 1: Genome polyprotein

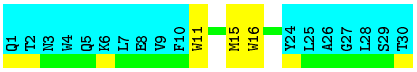
Chain A: 33% 7% • 57%



4.2.24 Score per residue for model 24 (medoid)

- Molecule 1: Genome polyprotein

Chain A: 33% 10% 57%



4.2.25 Score per residue for model 25

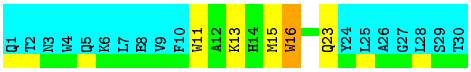
- Molecule 1: Genome polyprotein

Chain A: 23% 20% 57%

4.2.26 Score per residue for model 26

- Molecule 1: Genome polyprotein

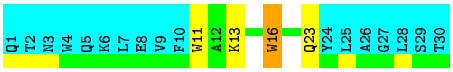
Chain A: 27% 13% • 57%



4.2.27 Score per residue for model 27

- Molecule 1: Genome polyprotein

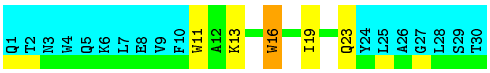
Chain A: 30% 10% • 57%



4.2.28 Score per residue for model 28

- Molecule 1: Genome polyprotein

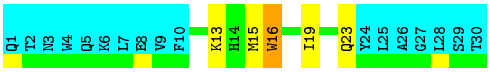
Chain A: 27% 13% • 57%



4.2.29 Score per residue for model 29

- Molecule 1: Genome polyprotein

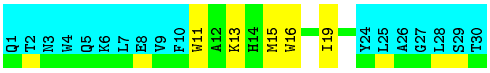
Chain A: 27% 13% • 57%



4.2.30 Score per residue for model 30

- Molecule 1: Genome polyprotein

Chain A: 27% 17% 57%



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, 30 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)	BMRB entry 15559
Number of chemical shift lists	1
Total number of shifts	332
Number of shifts mapped to atoms	332
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%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality (i)

6.1 Standard geometry (i)

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

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	114	107	107	1±1
All	All	3420	3210	3210	24

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.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:16:TRP:CZ3	1:A:19:ILE:HG21	0.50	2.41	18	5
1:A:16:TRP:CE3	1:A:19:ILE:HD13	0.50	2.41	29	3
1:A:16:TRP:CE3	1:A:16:TRP:HA	0.46	2.46	6	6
1:A:20:SER:HA	1:A:23:GLN:HG2	0.45	1.86	2	1
1:A:16:TRP:HA	1:A:16:TRP:CE3	0.45	2.47	13	8
1:A:18:PHE:O	1:A:22:ILE:HD13	0.42	2.14	8	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	13/30 (43%)	13±0 (100±1%)	0±0 (0±1%)	0±0 (0±0%)	100	100
All	All	390/900 (43%)	389 (100%)	1 (0%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	11/26 (42%)	7±1 (65±9%)	4±1 (35±9%)	1	9
All	All	330/780 (42%)	214 (65%)	116 (35%)	1	9

All 9 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	11	TRP	29
1	A	15	MET	19
1	A	16	TRP	19
1	A	23	GLN	18
1	A	19	ILE	13
1	A	13	LYS	13
1	A	22	ILE	2
1	A	20	SER	2
1	A	17	ASN	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 carbohydrates 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 72% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 15559

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	332
Number of shifts mapped to atoms	332
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)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	18	-0.48 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	26	0.54 ± 0.19	Should be applied
$^{13}C'$	0	—	—
^{15}N	0	—	—

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. 132 atoms were assigned a chemical shift out of a possible 178. 0 out of 0 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	35/65 (54%)	26/26 (100%)	9/26 (35%)	0/13 (0%)
Sidechain	64/73 (88%)	40/44 (91%)	24/26 (92%)	0/3 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	33/40 (82%)	17/21 (81%)	16/16 (100%)	0/3 (0%)
Overall	132/178 (74%)	83/91 (91%)	49/68 (72%)	0/19 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 72%, i.e. 284 atoms were assigned a chemical shift out of a possible 394. 0 out of 4 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	78/150 (52%)	60/60 (100%)	18/60 (30%)	0/30 (0%)
Sidechain	146/175 (83%)	91/103 (88%)	55/65 (85%)	0/7 (0%)
Aromatic	60/69 (87%)	31/36 (86%)	29/29 (100%)	0/4 (0%)
Overall	284/394 (72%)	182/199 (91%)	102/154 (66%)	0/41 (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: