



# Full wwPDB NMR Structure Validation Report ⓘ

May 28, 2020 – 10:45 pm BST

PDB ID : 2KU4  
Title : Horse prion protein  
Authors : Perez, D.R.; Damberger, F.F.; Wuthrich, K.  
Deposited on : 2010-02-12

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 ⓘ symbol.

---

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

Cyrange : Kirchner and Güntert (2011)  
NmrClust : Kelley et al. (1996)  
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 : 2.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

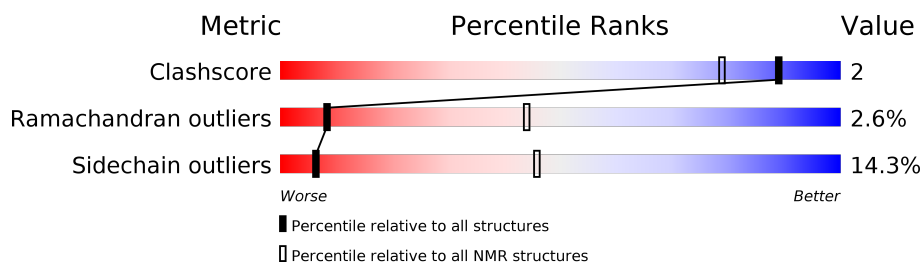
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*


The overall completeness of chemical shifts assignment is 64%.

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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	113	 81% 15% .

## 2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:121-A:229 (109)	0.31	6

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 3 clusters and 2 single-model clusters were found.

Cluster number	Models
1	3, 6, 9, 10, 14, 16, 18, 19
2	2, 7, 12, 13, 17, 20
3	4, 8, 11, 15
Single-model clusters	1; 5

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1798 atoms, of which 873 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Major prion protein.

Mol	Chain	Residues	Atoms						Trace
1	A	113	Total	C	H	N	O	S	0
			1798	578	873	160	180	7	

There are 2 discrepancies between the modelled and reference sequences:

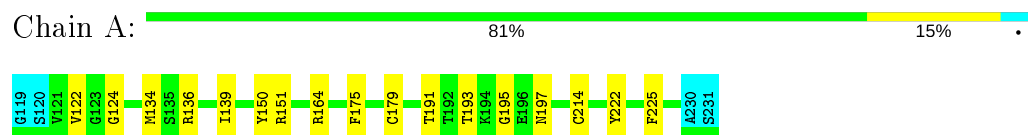
Chain	Residue	Modelled	Actual	Comment	Reference
A	119	GLY	-	expression tag	UNP O97964
A	120	SER	-	expression tag	UNP O97964

## 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: Major prion protein

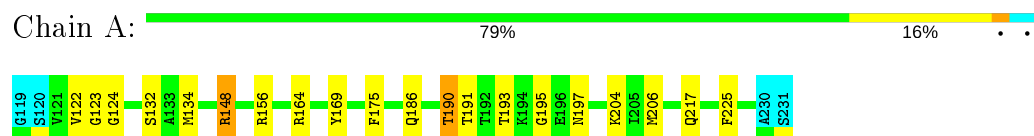


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

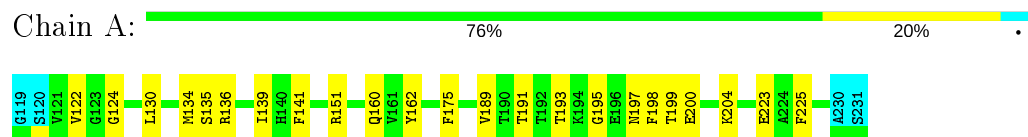
#### 4.2.1 Score per residue for model 1

- Molecule 1: Major prion protein



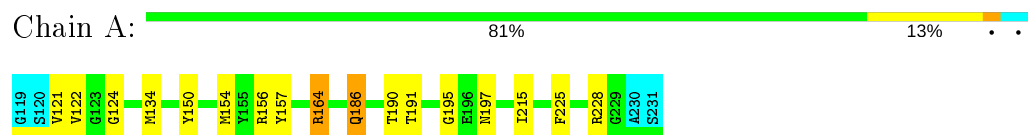
#### 4.2.2 Score per residue for model 2

- Molecule 1: Major prion protein



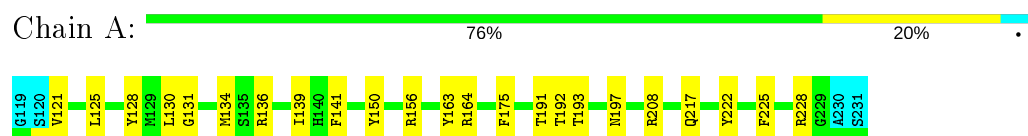
### 4.2.3 Score per residue for model 3

- Molecule 1: Major prion protein



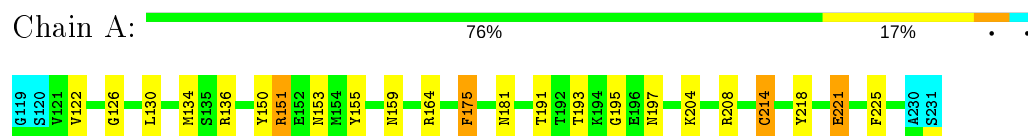
### 4.2.4 Score per residue for model 4

- Molecule 1: Major prion protein



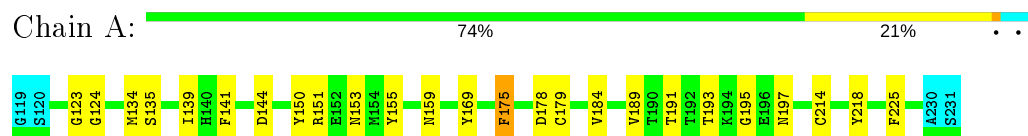
### 4.2.5 Score per residue for model 5

- Molecule 1: Major prion protein



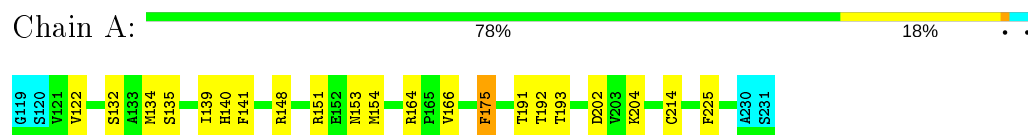
### 4.2.6 Score per residue for model 6 (medoid)

- Molecule 1: Major prion protein



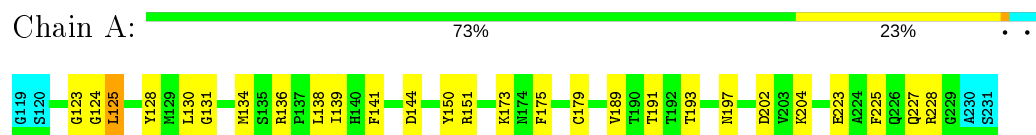
### 4.2.7 Score per residue for model 7

- Molecule 1: Major prion protein



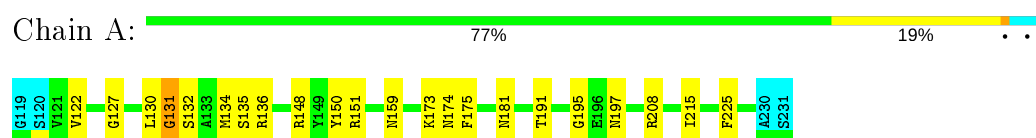
### 4.2.8 Score per residue for model 8

- Molecule 1: Major prion protein



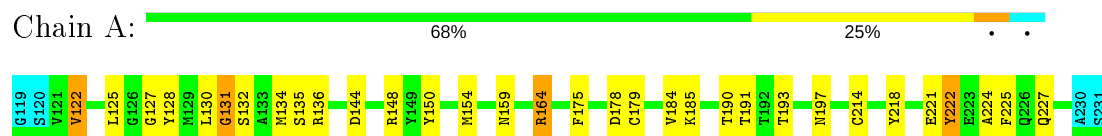
### 4.2.9 Score per residue for model 9

- Molecule 1: Major prion protein



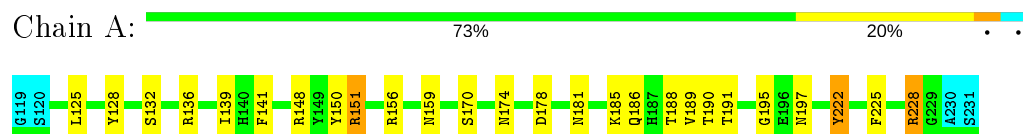
### 4.2.10 Score per residue for model 10

- Molecule 1: Major prion protein



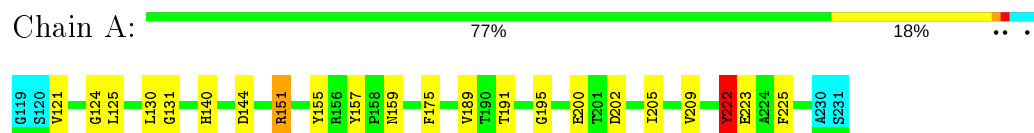
### 4.2.11 Score per residue for model 11

- Molecule 1: Major prion protein



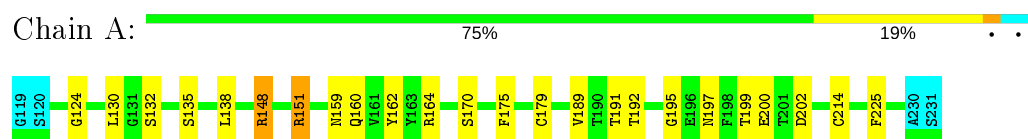
### 4.2.12 Score per residue for model 12

- Molecule 1: Major prion protein



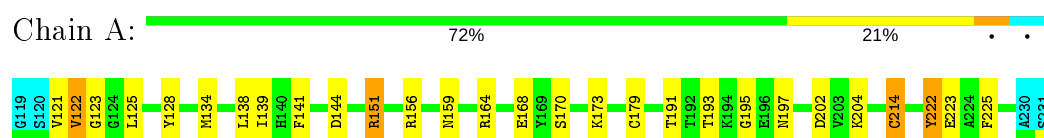
### 4.2.13 Score per residue for model 13

- Molecule 1: Major prion protein



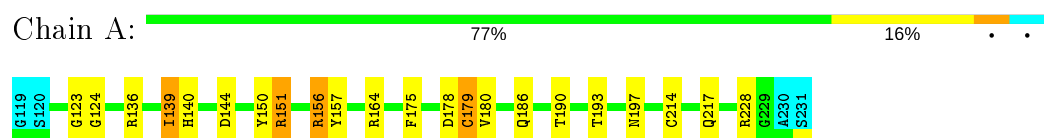
### 4.2.14 Score per residue for model 14

- Molecule 1: Major prion protein



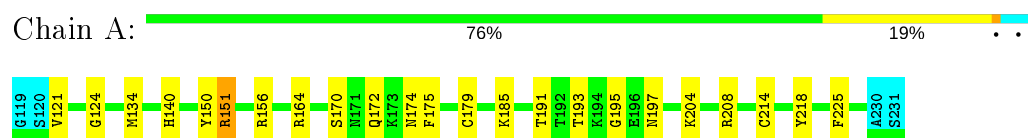
### 4.2.15 Score per residue for model 15

- Molecule 1: Major prion protein



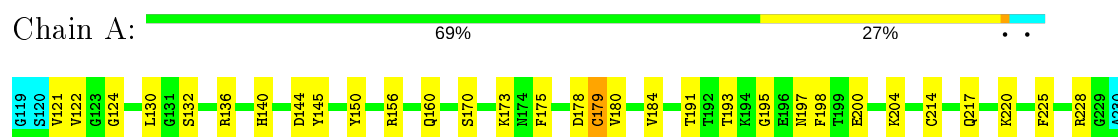
### 4.2.16 Score per residue for model 16

- Molecule 1: Major prion protein



### 4.2.17 Score per residue for model 17

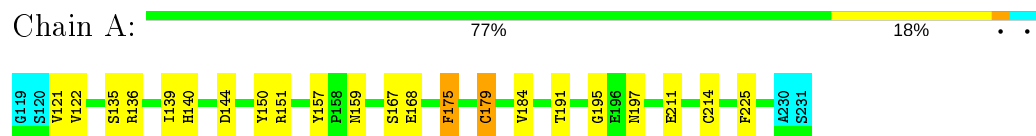
- Molecule 1: Major prion protein



S231

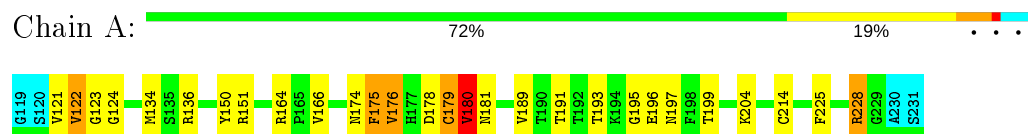
#### 4.2.18 Score per residue for model 18

- Molecule 1: Major prion protein



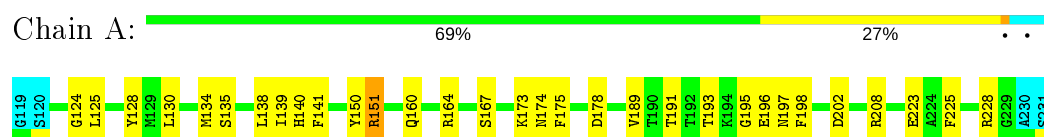
#### 4.2.19 Score per residue for model 19

- Molecule 1: Major prion protein



#### 4.2.20 Score per residue for model 20

- Molecule 1: Major prion protein



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
OPALp	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	921
Number of shifts mapped to atoms	921
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	64%

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

COVALENT-GEOMETRY INFOmissingINFO

### 5.1 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 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	904	855	855	3±1
All	All	18080	17100	17100	64

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:122:VAL:HG13	1:A:123:GLY:H	0.69	1.48	19	2
1:A:130:LEU:HD11	1:A:160:GLN:HB3	0.64	1.69	13	4
1:A:138:LEU:HD11	1:A:151:ARG:HB3	0.63	1.69	20	2
1:A:139:ILE:HG22	1:A:140:HIS:H	0.61	1.54	15	2
1:A:179:CYS:SG	1:A:214:CYS:CB	0.60	2.89	17	2
1:A:186:GLN:O	1:A:190:THR:HG22	0.59	1.98	15	2
1:A:125:LEU:HD23	1:A:128:TYR:HB2	0.56	1.77	10	2
1:A:130:LEU:HD23	1:A:131:GLY:N	0.56	2.16	4	5
1:A:175:PHE:CE1	1:A:214:CYS:SG	0.55	2.99	6	3
1:A:176:VAL:O	1:A:180:VAL:HG23	0.55	2.02	19	1
1:A:205:ILE:O	1:A:209:VAL:HG23	0.54	2.03	12	1
1:A:185:LYS:O	1:A:189:VAL:HG12	0.51	2.05	11	1
1:A:125:LEU:HD22	1:A:128:TYR:CG	0.51	2.41	11	2
1:A:179:CYS:CB	1:A:214:CYS:SG	0.49	3.00	19	5
1:A:175:PHE:CE2	1:A:214:CYS:SG	0.49	3.06	10	1
1:A:186:GLN:O	1:A:190:THR:HG23	0.49	2.07	3	2
1:A:179:CYS:SG	1:A:180:VAL:N	0.47	2.87	15	2
1:A:130:LEU:HD13	1:A:162:TYR:CZ	0.47	2.44	13	2
1:A:130:LEU:HD13	1:A:162:TYR:CE1	0.47	2.44	13	1
1:A:139:ILE:HG21	1:A:141:PHE:CE1	0.47	2.44	6	8
1:A:217:GLN:HE22	1:A:220:LYS:HE3	0.46	1.69	17	1
1:A:221:GLU:HA	1:A:224:ALA:HB3	0.45	1.88	10	1
1:A:125:LEU:HD22	1:A:128:TYR:CD1	0.44	2.47	11	2
1:A:166:VAL:CG1	1:A:175:PHE:CD2	0.44	3.01	7	1
1:A:175:PHE:CE1	1:A:218:TYR:HB2	0.42	2.50	5	2
1:A:125:LEU:HD13	1:A:128:TYR:HB2	0.41	1.92	8	2
1:A:139:ILE:HG22	1:A:140:HIS:N	0.41	2.30	7	1
1:A:153:ASN:HA	1:A:155:TYR:CE1	0.41	2.50	6	2
1:A:122:VAL:HG21	1:A:190:THR:HG21	0.40	1.91	3	1
1:A:172:GLN:HA	1:A:218:TYR:CE1	0.40	2.52	16	1

## 5.2 Torsion angles ⓘ

### 5.2.1 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 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	109/113 (96%)	90±3 (83±3%)	16±4 (15±3%)	3±1 (3±1%)	8	44

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	2180/2260 (96%)	1801 (83%)	322 (15%)	57 (3%)	8 44

All 14 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	195	GLY	15
1	A	124	GLY	12
1	A	121	VAL	8
1	A	122	VAL	6
1	A	123	GLY	4
1	A	131	GLY	2
1	A	170	SER	2
1	A	189	VAL	2
1	A	191	THR	1
1	A	176	VAL	1
1	A	186	GLN	1
1	A	126	GLY	1
1	A	197	ASN	1
1	A	188	THR	1

### 5.2.2 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 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	99/101 (98%)	85±3 (86±3%)	14±3 (14±3%)	6 46
All	All	1980/2020 (98%)	1696 (86%)	284 (14%)	6 46

All 53 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	225	PHE	19
1	A	191	THR	18
1	A	197	ASN	17
1	A	175	PHE	15
1	A	193	THR	14

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	134	MET	14
1	A	164	ARG	11
1	A	151	ARG	10
1	A	204	LYS	9
1	A	159	ASN	9
1	A	144	ASP	8
1	A	135	SER	8
1	A	136	ARG	8
1	A	179	CYS	7
1	A	178	ASP	7
1	A	132	SER	7
1	A	228	ARG	6
1	A	202	ASP	6
1	A	173	LYS	5
1	A	122	VAL	5
1	A	174	ASN	5
1	A	189	VAL	5
1	A	223	GLU	5
1	A	222	TYR	4
1	A	156	ARG	4
1	A	181	ASN	4
1	A	200	GLU	4
1	A	140	HIS	4
1	A	170	SER	3
1	A	199	THR	3
1	A	154	MET	3
1	A	148	ARG	3
1	A	217	GLN	3
1	A	192	THR	3
1	A	138	LEU	2
1	A	227	GLN	2
1	A	168	GLU	2
1	A	215	ILE	2
1	A	185	LYS	2
1	A	125	LEU	2
1	A	190	THR	2
1	A	196	GLU	2
1	A	167	SER	2
1	A	206	MET	1
1	A	180	VAL	1
1	A	166	VAL	1
1	A	153	ASN	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	221	GLU	1
1	A	130	LEU	1
1	A	211	GLU	1
1	A	139	ILE	1
1	A	214	CYS	1
1	A	184	VAL	1

### 5.2.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.3 Non-standard residues in protein, DNA, RNA chains [i](#)

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

## 5.4 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.5 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.6 Other polymers [i](#)

There are no such molecules in this entry.

## 5.7 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 64% for the well-defined parts and 63% for the entire structure.

### 6.1 Chemical shift list 1

File name: input\_cs.cif

Chemical shift list name: *assigned\_chem\_shift\_list\_1*

#### 6.1.1 Bookkeeping

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	921
Number of shifts mapped to atoms	921
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

#### 6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	113	$0.02 \pm 0.25$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	104	$0.37 \pm 0.14$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	0	—	None (insufficient data)
$^{15}\text{N}$	109	$0.17 \pm 0.22$	None needed ( $< 0.5$ ppm)

#### 6.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 64%, i.e. 888 atoms were assigned a chemical shift out of a possible 1391. 14 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	316/539 (59%)	101/215 (47%)	109/218 (50%)	106/106 (100%)
Sidechain	532/707 (75%)	304/416 (73%)	206/250 (82%)	22/41 (54%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	40/145 (28%)	7/76 (9%)	33/66 (50%)	0/3 (0%)
Overall	888/1391 (64%)	412/707 (58%)	348/534 (65%)	128/150 (85%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 63%, i.e. 901 atoms were assigned a chemical shift out of a possible 1419. 14 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	326/559 (58%)	104/223 (47%)	113/226 (50%)	109/110 (99%)
Sidechain	535/715 (75%)	304/421 (72%)	209/253 (83%)	22/41 (54%)
Aromatic	40/145 (28%)	7/76 (9%)	33/66 (50%)	0/3 (0%)
Overall	901/1419 (63%)	415/720 (58%)	355/545 (65%)	131/154 (85%)

#### 6.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

#### 6.1.5 Random Coil Index (RCI) plots ⓘ

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:

