



# Full wwPDB NMR Structure Validation Report ⓘ

May 29, 2020 – 11:51 am BST

PDB ID : 5MQX  
Title : NMR solution structure of macro domain from Venezuelan equine encephalitis virus (VEEV) in complex with ADP-ribose  
Authors : Makrynitsa, G.I.; Ntonti, D.; Marousis, K.D.; Matsoukas, M.T.; Papageorgiou, N.; Coutard, B.; Bentrop, D.; Spyroulias, G.A.  
Deposited on : 2016-12-21

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
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
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

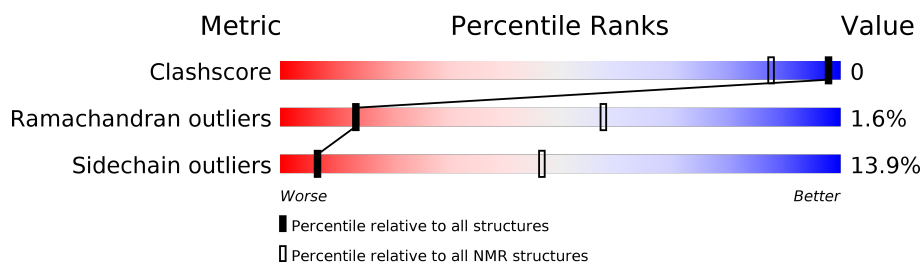
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 86%.

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	166	84% 11% . .

## 2 Ensemble composition and analysis ⓘ

This entry contains 21 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:3-A:160 (158)	0.52	1

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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 16, 19, 21
2	11, 12, 13, 14, 15, 17, 18, 20

### 3 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 2500 atoms, of which 1250 are hydrogens and 0 are deuteriums.

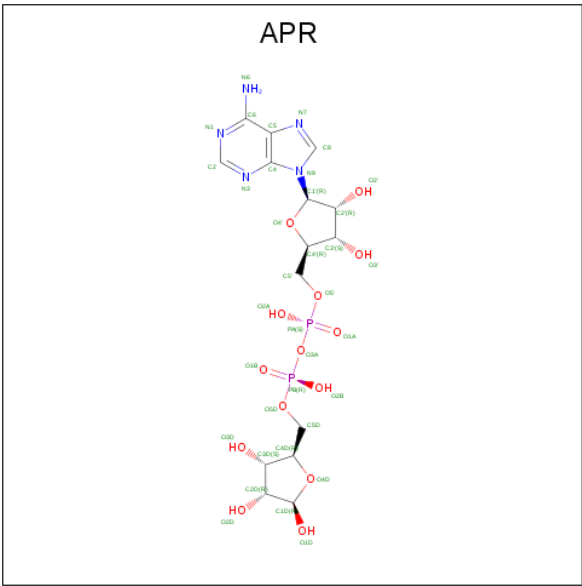
- Molecule 1 is a protein called Non-structural protein3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	160	2443	764	1229	214	233	3	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	161	HIS	-	expression tag	UNP P36328
A	162	HIS	-	expression tag	UNP P36328
A	163	HIS	-	expression tag	UNP P36328
A	164	HIS	-	expression tag	UNP P36328
A	165	HIS	-	expression tag	UNP P36328
A	166	HIS	-	expression tag	UNP P36328

- Molecule 2 is ADENOSINE-5-DIPHOSPHORIBOSE (three-letter code: APR) (formula: C<sub>15</sub>H<sub>23</sub>N<sub>5</sub>O<sub>14</sub>P<sub>2</sub>).



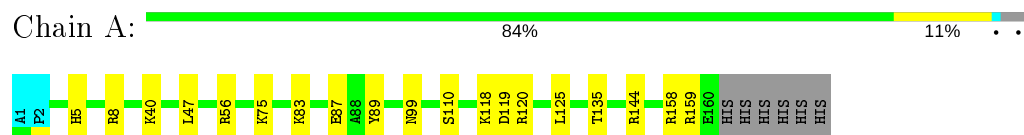
Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	P
2	A	1	57	15	21	5	14	2

## 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: Non-structural protein3

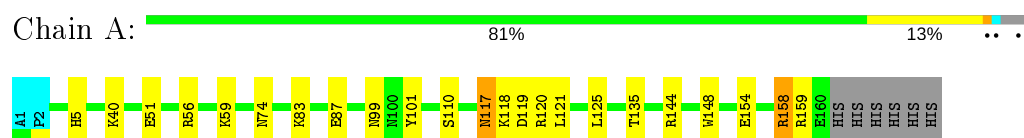


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

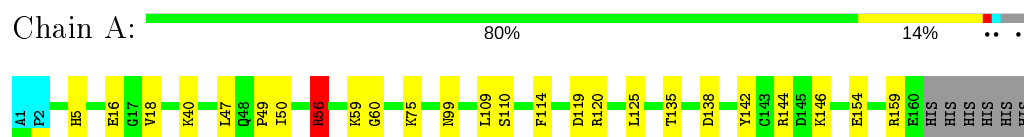
#### 4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Non-structural protein3



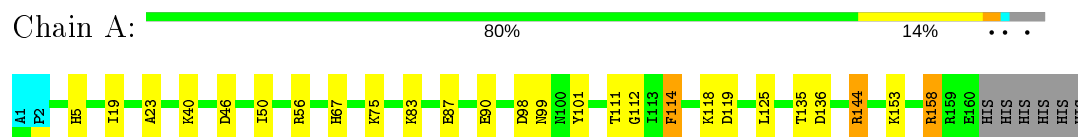
#### 4.2.2 Score per residue for model 2

- Molecule 1: Non-structural protein3



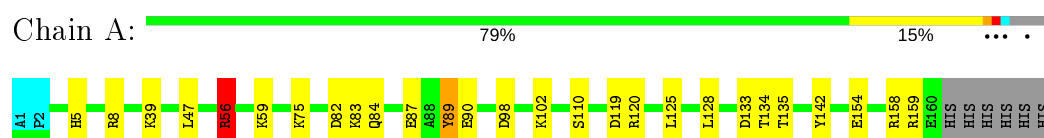
### 4.2.3 Score per residue for model 3

- Molecule 1: Non-structural protein3



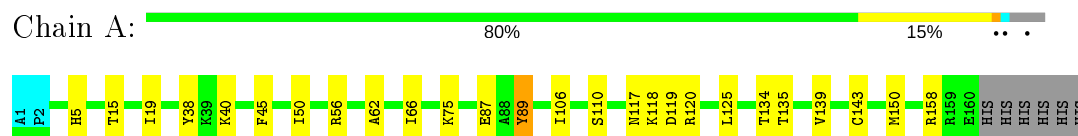
### 4.2.4 Score per residue for model 4

- Molecule 1: Non-structural protein3



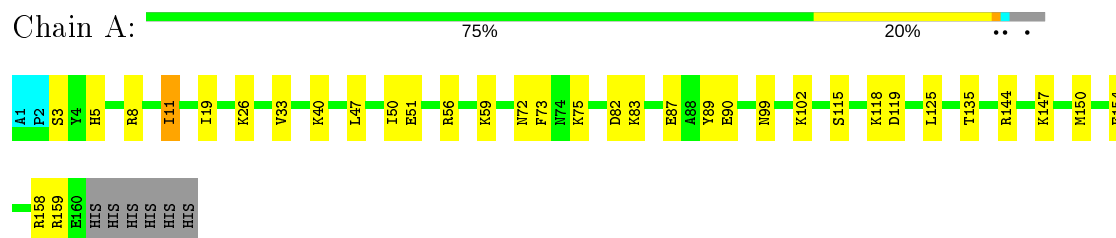
### 4.2.5 Score per residue for model 5

- Molecule 1: Non-structural protein3



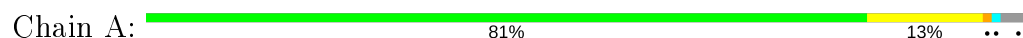
### 4.2.6 Score per residue for model 6

- Molecule 1: Non-structural protein3



### 4.2.7 Score per residue for model 7

- Molecule 1: Non-structural protein3

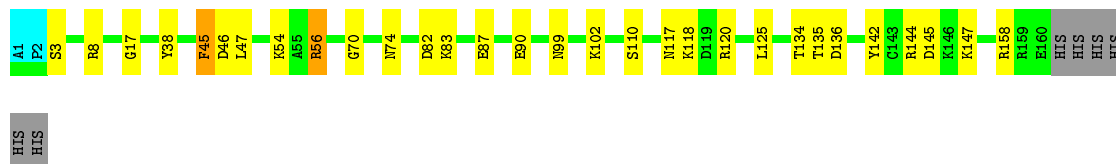




#### 4.2.8 Score per residue for model 8

- Molecule 1: Non-structural protein3

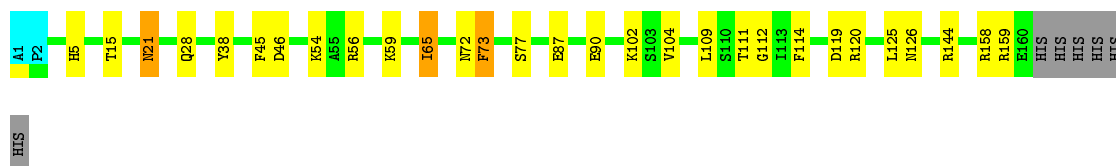
Chain A: 77% 17% ...



#### 4.2.9 Score per residue for model 9

- Molecule 1: Non-structural protein3

Chain A: 78% 16% ...



#### 4.2.10 Score per residue for model 10

- Molecule 1: Non-structural protein3

Chain A: 80% 14% ...



#### 4.2.11 Score per residue for model 11

- Molecule 1: Non-structural protein3

Chain A: 80% 14% ...








HIS  
HIS


## 4.2.17 Score per residue for model 17

- Molecule 1: Non-structural protein3

Chain A:  82% 12% ..


## 4.2.18 Score per residue for model 18

- Molecule 1: Non-structural protein3

Chain A:  84% 11% ..

## 4.2.19 Score per residue for model 19


- Molecule 1: Non-structural protein3

Chain A:  79% 14% ..

HIS

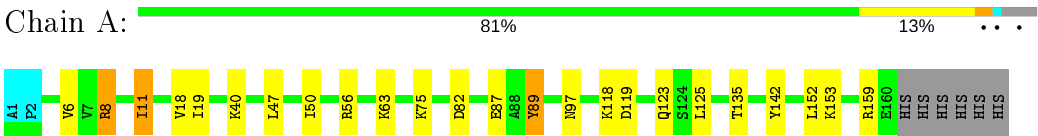
## 4.2.20 Score per residue for model 20

- Molecule 1: Non-structural protein3

Chain A:  80% 13% ..

4.2.21 Score per residue for model 21

- Molecule 1: Non-structural protein3



## 5 Refinement protocol and experimental data overview

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

Of the 112 calculated structures, 21 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
DYANA	structure calculation	
Amber	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	1860
Number of shifts mapped to atoms	1860
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	86%

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	1202	1215	1215	1±1
2	A	36	21	21	0±0
All	All	25998	25956	25956	20

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.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:106:ILE:HD12	1:A:139:VAL:HG13	0.51	1.82	5	1
1:A:33:VAL:HG11	2:A:201:APR:O2A	0.48	2.08	6	1
1:A:23:ALA:HB2	1:A:67:HIS:CE1	0.48	2.43	3	1
1:A:158:ARG:HA	1:A:158:ARG:CZ	0.46	2.40	14	1
1:A:57:LEU:CD2	1:A:59:LYS:HE2	0.46	2.41	10	1
1:A:51:GLU:H	1:A:51:GLU:CD	0.46	2.14	1	1
1:A:51:GLU:CD	1:A:51:GLU:H	0.45	2.15	12	1
1:A:21:ASN:HD21	1:A:65:ILE:HG22	0.44	1.70	9	1
1:A:56:ARG:HG2	1:A:56:ARG:HH21	0.44	1.72	2	1
1:A:38:TYR:CE1	1:A:45:PHE:CZ	0.44	3.06	8	1
1:A:89:TYR:CD1	1:A:128:LEU:HD13	0.42	2.49	19	1
1:A:6:VAL:HG13	1:A:152:LEU:HD12	0.41	1.90	21	1
1:A:49:PRO:C	1:A:50:ILE:HD12	0.41	2.36	2	1
1:A:75:LYS:HE2	1:A:75:LYS:HA	0.41	1.90	6	1
1:A:19:ILE:HD13	1:A:41:PHE:CE2	0.41	2.50	7	1
1:A:11:ILE:CD1	1:A:19:ILE:HG23	0.41	2.46	21	2
1:A:121:LEU:HD11	1:A:148:TRP:CD1	0.40	2.50	1	1
1:A:11:ILE:HG12	1:A:142:TYR:CD1	0.40	2.52	12	1
1:A:99:ASN:ND2	1:A:101:TYR:CZ	0.40	2.90	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	157/166 (95%)	137±2 (88±1%)	17±2 (11±2%)	2±1 (2±1%)	13	57
All	All	3297/3486 (95%)	2885 (88%)	360 (11%)	52 (2%)	13	57

All 16 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	47	LEU	15
1	A	110	SER	11
1	A	111	THR	5
1	A	72	ASN	3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	30	GLY	3
1	A	73	PHE	3
1	A	112	GLY	2
1	A	115	SER	2
1	A	114	PHE	1
1	A	45	PHE	1
1	A	107	PRO	1
1	A	117	ASN	1
1	A	62	ALA	1
1	A	46	ASP	1
1	A	146	LYS	1
1	A	29	PRO	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	127/134 (95%)	109±3 (86±2%)	18±3 (14±2%)	7	47
All	All	2667/2814 (95%)	2297 (86%)	370 (14%)	7	47

All 76 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	125	LEU	20
1	A	5	HIS	17
1	A	118	LYS	16
1	A	135	THR	15
1	A	87	GLU	14
1	A	40	LYS	13
1	A	119	ASP	12
1	A	99	ASN	12
1	A	83	LYS	12
1	A	56	ARG	10
1	A	75	LYS	10
1	A	109	LEU	9
1	A	59	LYS	8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	82	ASP	8
1	A	74	ASN	8
1	A	54	LYS	7
1	A	134	THR	7
1	A	90	GLU	7
1	A	50	ILE	7
1	A	8	ARG	7
1	A	153	LYS	6
1	A	117	ASN	6
1	A	147	LYS	6
1	A	154	GLU	6
1	A	26	LYS	5
1	A	150	MET	5
1	A	146	LYS	5
1	A	19	ILE	5
1	A	102	LYS	5
1	A	144	ARG	4
1	A	136	ASP	4
1	A	113	ILE	4
1	A	77	SER	4
1	A	18	VAL	4
1	A	98	ASP	3
1	A	45	PHE	3
1	A	72	ASN	3
1	A	73	PHE	3
1	A	158	ARG	3
1	A	67	HIS	3
1	A	38	TYR	3
1	A	89	TYR	3
1	A	143	CYS	3
1	A	114	PHE	3
1	A	133	ASP	3
1	A	84	GLN	3
1	A	15	THR	3
1	A	46	ASP	3
1	A	104	VAL	3
1	A	11	ILE	3
1	A	28	GLN	2
1	A	94	LYS	2
1	A	51	GLU	2
1	A	3	SER	2
1	A	63	LYS	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	123	GLN	2
1	A	76	VAL	2
1	A	97	ASN	2
1	A	149	GLU	1
1	A	58	VAL	1
1	A	142	TYR	1
1	A	138	ASP	1
1	A	21	ASN	1
1	A	4	TYR	1
1	A	65	ILE	1
1	A	126	ASN	1
1	A	16	GLU	1
1	A	25	SER	1
1	A	39	LYS	1
1	A	145	ASP	1
1	A	156	VAL	1
1	A	111	THR	1
1	A	100	ASN	1
1	A	47	LEU	1
1	A	66	ILE	1
1	A	128	LEU	1

### 5.2.3 RNA ⓘ

There are no RNA molecules in this entry.

### 5.3 Non-standard residues in protein, DNA, RNA chains ⓘ

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

### 5.4 Carbohydrates ⓘ

There are no carbohydrates in this entry.

LIGAND-GEOMETRY INFOmissingINFO

### 5.5 Other polymers ⓘ

There are no such molecules in this entry.

## 5.6 Polymer linkage issues ⓘ

There are no chain breaks in this entry.



## 6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 86% for the well-defined parts and 86% 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	1860
Number of shifts mapped to atoms	1860
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

#### 6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	158	$-0.06 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	145	$0.13 \pm 0.09$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	147	$0.15 \pm 0.08$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	149	$-0.17 \pm 0.39$	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 86%, i.e. 1614 atoms were assigned a chemical shift out of a possible 1876. 25 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	753/780 (97%)	304/311 (98%)	301/316 (95%)	148/153 (97%)
Sidechain	819/980 (84%)	501/569 (88%)	315/366 (86%)	3/45 (7%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	42/116 (36%)	22/62 (35%)	19/49 (39%)	1/5 (20%)
Overall	1614/1876 (86%)	827/942 (88%)	635/731 (87%)	152/203 (75%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 86%, i.e. 1632 atoms were assigned a chemical shift out of a possible 1895. 25 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	760/788 (96%)	306/314 (97%)	305/320 (95%)	149/154 (97%)
Sidechain	830/991 (84%)	508/576 (88%)	319/370 (86%)	3/45 (7%)
Aromatic	42/116 (36%)	22/62 (35%)	19/49 (39%)	1/5 (20%)
Overall	1632/1895 (86%)	836/952 (88%)	643/739 (87%)	153/204 (75%)

#### 6.1.4 Statistically unusual chemical shifts ⓘ

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	41	PHE	CE1	120.81	137.92 – 123.42	-6.8
1	A	72	ASN	HB2	0.89	4.36 – 1.26	-6.2
1	A	64	HIS	HE1	4.98	10.53 – 5.43	-5.9
1	A	72	ASN	HB3	0.89	4.41 – 1.11	-5.7
1	A	7	VAL	HB	0.37	3.59 – 0.39	-5.1

#### 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:

