



# Full wwPDB NMR Structure Validation Report ⓘ

Feb 13, 2017 – 12:38 am GMT

PDB ID : 2M26  
Title : NMR structure of the C-terminal domain of the protein HCFC1 from Mus musculus  
Authors : Serrano, P.; Geralt, M.; Dutta, S.K.; Wuthrich, K.; Joint Center for Structural Genomics (JCSG); Partnership for Stem Cell Biology (STEMCELL)  
Deposited on : 2012-12-14

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

<http://wwpdb.org/validation/2016/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 : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : trunk28760  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

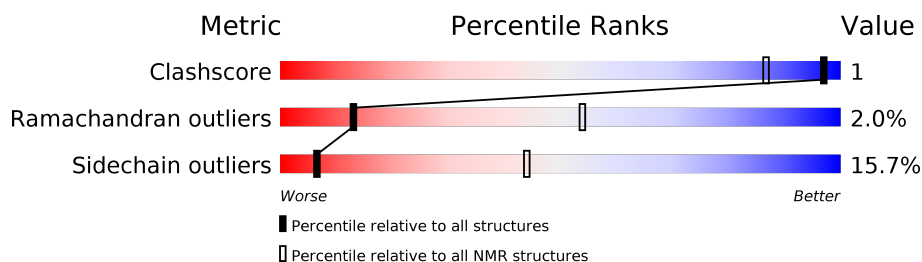
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 87%.

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	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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	126	

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 4 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 energy*.

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:6-A:46, A:62-A:118 (98)	0.82	4

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 3 single-model clusters were found.

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

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

There is only 1 type of molecule in this entry. The entry contains 1706 atoms, of which 770 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called HCF C-terminal chain 1.

Mol	Chain	Residues	Atoms						Trace
1	A	126	Total	C	H	N	O	S	0
			1706	590	770	159	182	5	

There is a discrepancy between the modelled and reference sequences:

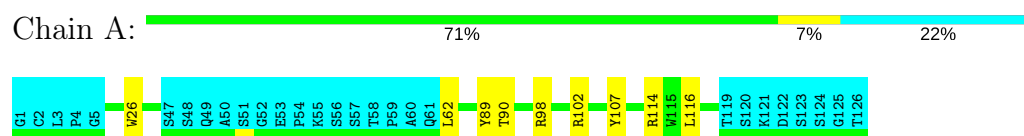
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	leader sequence	UNP Q61191

## 4 Residue-property plots

### 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: HCF C-terminal chain 1

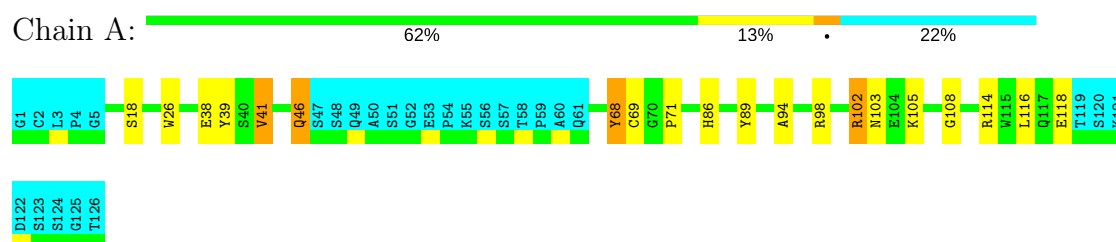


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

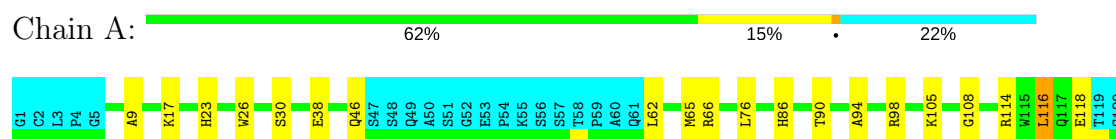
#### 4.2.1 Score per residue for model 1

- Molecule 1: HCF C-terminal chain 1



#### 4.2.2 Score per residue for model 2

- Molecule 1: HCF C-terminal chain 1





### 4.2.3 Score per residue for model 3

- Molecule 1: HCF C-terminal chain 1

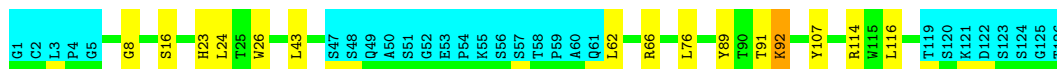
Chain A: 65% 12% 22%



### 4.2.4 Score per residue for model 4 (medoid)

- Molecule 1: HCF C-terminal chain 1

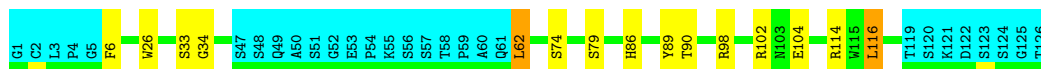
Chain A: 66% 11% 22%



### 4.2.5 Score per residue for model 5

- Molecule 1: HCF C-terminal chain 1

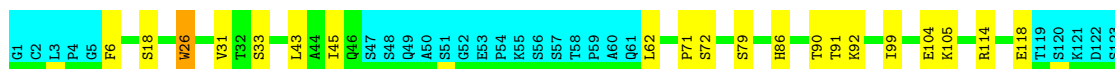
Chain A: 66% 10% 22%



### 4.2.6 Score per residue for model 6

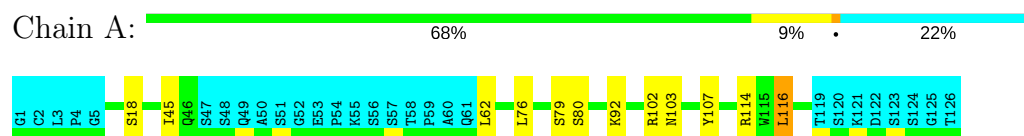
- Molecule 1: HCF C-terminal chain 1

Chain A: 62% 15% 22%



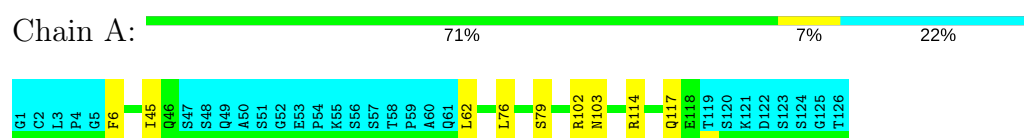
### 4.2.7 Score per residue for model 7

- Molecule 1: HCF C-terminal chain 1



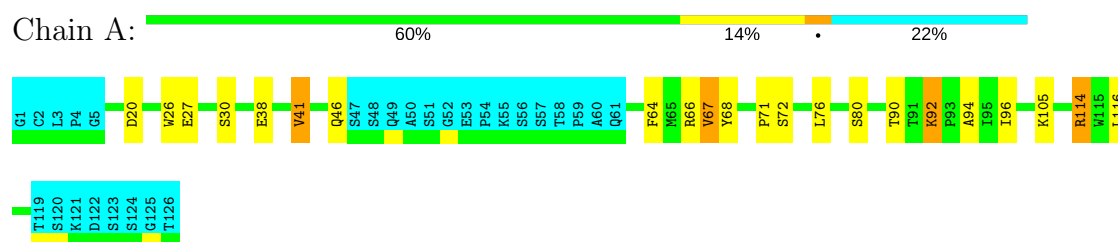
### 4.2.8 Score per residue for model 8

- Molecule 1: HCF C-terminal chain 1



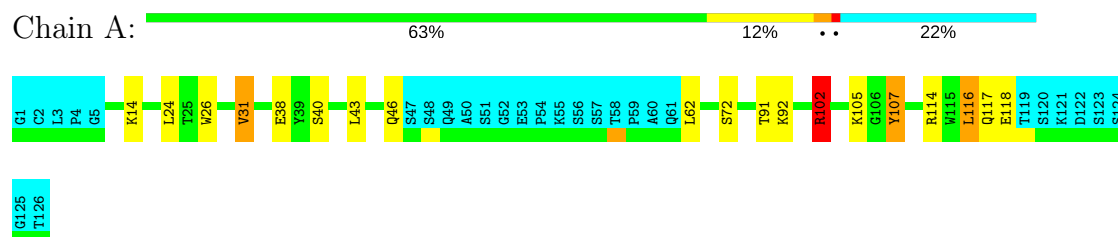
### 4.2.9 Score per residue for model 9

- Molecule 1: HCF C-terminal chain 1



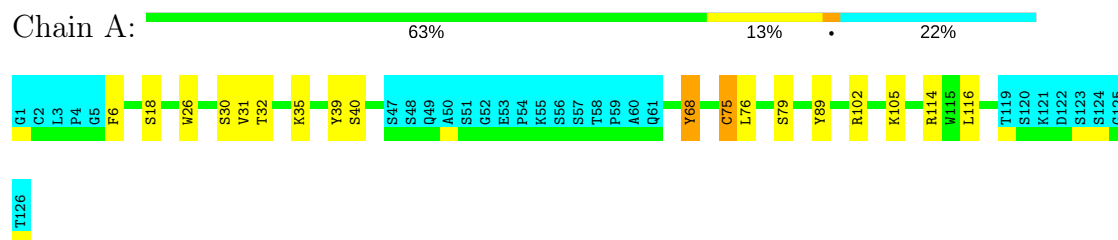
### 4.2.10 Score per residue for model 10

- Molecule 1: HCF C-terminal chain 1



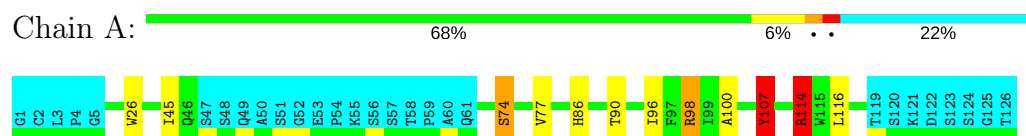
### 4.2.11 Score per residue for model 11

- Molecule 1: HCF C-terminal chain 1



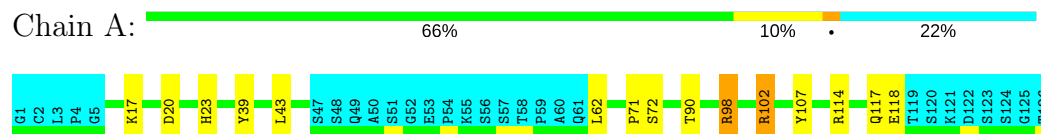
#### 4.2.12 Score per residue for model 12

- Molecule 1: HCF C-terminal chain 1



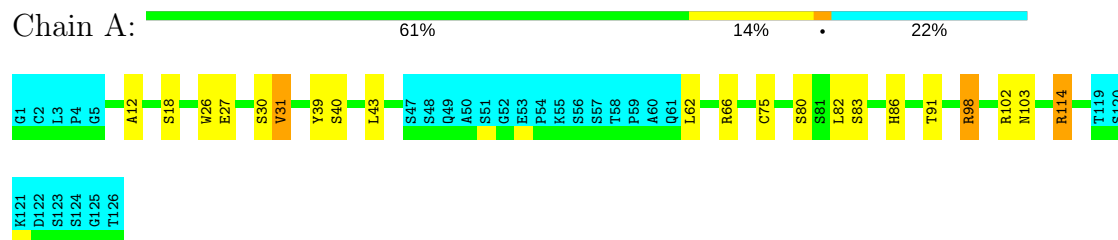
#### 4.2.13 Score per residue for model 13

- Molecule 1: HCF C-terminal chain 1



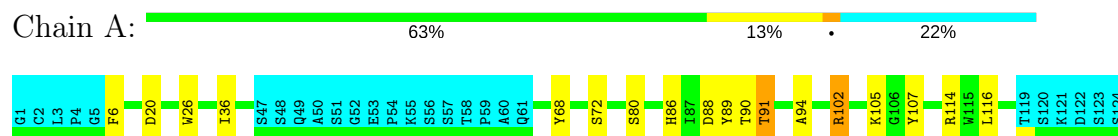
#### 4.2.14 Score per residue for model 14

- Molecule 1: HCF C-terminal chain 1



#### 4.2.15 Score per residue for model 15

- Molecule 1: HCF C-terminal chain 1





G125  
T126

#### 4.2.16 Score per residue for model 16

- Molecule 1: HCF C-terminal chain 1

Chain A:  64% 11% 22%

G1 C2 L3 P4 G5 D20 H23 W26 V31 Y39 I45 Q46 S47 S48 Q49 A50 S51 G52 E53 P54 K55 S56 S57 T58 P59 A60 Q61 L62 A63 Y68 S72 P73 S74 S81 T90 I99 K105 R114 W115 L116 T119 S120 K121 D122 S123 S124

G125  
T126

#### 4.2.17 Score per residue for model 17

- Molecule 1: HCF C-terminal chain 1

Chain A:  61% 15% 22%

G1 C2 L3 P4 G5 F6 P7 I15 S16 K17 S18 T25 G34 E38 Y42 Q46 S47 S48 Q49 A50 S51 G52 E53 P54 K55 S56 S57 T58 P59 A60 Q61 L62 A63 P64 S72 P73 S74 C75 Q78 L82 Y89 T90 G108 R114 E118 S120

K121 D122 S123 S124 G125 T126

#### 4.2.18 Score per residue for model 18

- Molecule 1: HCF C-terminal chain 1

Chain A:  68% 10% 22%

G1 C2 L3 P4 G5 W26 S33 I45 Q46 S47 S48 Q49 A50 S51 G52 E53 P54 K55 S56 S57 T58 P59 A60 Q61 L62 L76 V77 Q78 S79 H86 T90 T91 R114 Q117 E118 T119 S120 K121 D122 S123 S124 G125 T126

#### 4.2.19 Score per residue for model 19

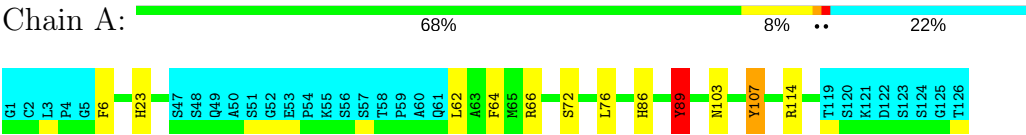
- Molecule 1: HCF C-terminal chain 1

Chain A:  66% 12% 22%

G1 C2 L3 P4 G5 S16 W26 Y39 Q46 S47 S48 Q49 A50 S51 G52 E53 P54 K55 S56 S57 T58 P59 A60 Q61 S80 Y89 A94 R98 I99 R102 K105 R114 W115 L116 Q117 E118 T119 S120 K121 D122 S123 S124 G125 T126

4.2.20 Score per residue for model 20

- Molecule 1: HCF C-terminal chain 1



## 5 Refinement protocol and experimental data overview

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

Of the 80 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
CYANA	refinement	
UNIO	structure solution	
Opalp	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)	2m26_cs.str
Number of chemical shift lists	1
Total number of shifts	1311
Number of shifts mapped to atoms	1311
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

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

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 (average) 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	0.63±0.01	0±0/776 (0.0±0.0%)	1.06±0.05	1±1/1060 (0.1±0.1%)
All	All	0.63	0/15520 (0.0%)	1.06	18/21200 (0.1%)

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	0.8±0.7
All	All	0	17

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	41	VAL	CA-CB-CG1	12.61	129.81	110.90	1	2
1	A	68	TYR	CB-CG-CD2	-7.12	116.72	121.00	1	1
1	A	67	VAL	CG1-CB-CG2	6.97	122.06	110.90	9	1
1	A	102	ARG	NE-CZ-NH2	-6.32	117.14	120.30	10	1
1	A	102	ARG	NE-CZ-NH1	5.90	123.25	120.30	1	1
1	A	66	ARG	NE-CZ-NH2	-5.83	117.38	120.30	4	1
1	A	89	TYR	CB-CG-CD2	-5.76	117.55	121.00	20	1
1	A	107	TYR	CB-CG-CD2	-5.73	117.56	121.00	12	1
1	A	114	ARG	NE-CZ-NH1	5.67	123.14	120.30	16	2
1	A	98	ARG	CD-NE-CZ	5.58	131.41	123.60	12	1
1	A	89	TYR	CA-CB-CG	5.52	123.90	113.40	20	1
1	A	98	ARG	NE-CZ-NH1	5.27	122.94	120.30	13	2
1	A	77	VAL	CA-CB-CG1	5.25	118.77	110.90	12	1
1	A	114	ARG	NE-CZ-NH2	-5.18	117.71	120.30	14	1
1	A	98	ARG	NE-CZ-NH2	-5.10	117.75	120.30	3	1

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	A	114	ARG	Sidechain	3
1	A	89	TYR	Sidechain	2
1	A	102	ARG	Sidechain	2
1	A	75	CYS	Peptide	2
1	A	68	TYR	Sidechain	2
1	A	107	TYR	Sidechain	1
1	A	46	GLN	Peptide	1
1	A	98	ARG	Sidechain	1
1	A	9	ALA	Peptide	1
1	A	66	ARG	Sidechain	1
1	A	86	HIS	Peptide	1

## 6.2 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 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	753	632	754	1±1
All	All	15060	12640	15080	25

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:41:VAL:HG12	1:A:68:TYR:HB2	0.72	1.60	9	2
1:A:116:LEU:HD22	1:A:116:LEU:H	0.69	1.47	7	2
1:A:116:LEU:H	1:A:116:LEU:HD22	0.65	1.51	10	2
1:A:92:LYS:HE3	1:A:116:LEU:HD21	0.53	1.81	9	1
1:A:100:ALA:HB3	1:A:107:TYR:CZ	0.52	2.40	12	1
1:A:102:ARG:HG2	1:A:107:TYR:CE1	0.51	2.40	15	1
1:A:94:ALA:HB2	1:A:116:LEU:HD23	0.49	1.84	9	4
1:A:62:LEU:HD13	1:A:64:PHE:CZ	0.47	2.44	20	1
1:A:36:ILE:HD12	1:A:36:ILE:N	0.44	2.27	15	1

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:12:ALA:HB3	1:A:27:GLU:HB2	0.44	1.89	14	1
1:A:26:TRP:CE3	1:A:99:ILE:HD13	0.42	2.49	16	3
1:A:14:LYS:HE2	1:A:14:LYS:HA	0.42	1.91	10	1
1:A:42:TYR:HB3	1:A:64:PHE:CG	0.42	2.50	17	1
1:A:41:VAL:CG1	1:A:68:TYR:HB2	0.41	2.39	1	1
1:A:92:LYS:HE3	1:A:116:LEU:HD22	0.41	1.93	4	1
1:A:102:ARG:HG2	1:A:107:TYR:CE2	0.40	2.51	10	1
1:A:94:ALA:HB2	1:A:116:LEU:HD13	0.40	1.93	2	1

## 6.3 Torsion angles

### 6.3.1 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	98/126 (78%)	81±3 (83±3%)	15±3 (15±3%)	2±2 (2±2%)	13	54
All	All	1960/2520 (78%)	1622 (83%)	298 (15%)	40 (2%)	13	54

All 22 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	71	PRO	5
1	A	31	VAL	3
1	A	108	GLY	3
1	A	33	SER	3
1	A	46	GLN	3
1	A	74	SER	2
1	A	62	LEU	2
1	A	72	SER	2
1	A	34	GLY	2
1	A	75	CYS	2
1	A	96	ILE	2
1	A	107	TYR	1
1	A	68	TYR	1
1	A	8	GLY	1

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	23	HIS	1
1	A	7	PRO	1
1	A	91	THR	1
1	A	32	THR	1
1	A	17	LYS	1
1	A	63	ALA	1
1	A	20	ASP	1
1	A	76	LEU	1

### 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	82/104 (79%)	69±3 (84±4%)	13±3 (16±4%)	6	44
All	All	1640/2080 (79%)	1383 (84%)	257 (16%)	6	44

All 55 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	114	ARG	19
1	A	26	TRP	12
1	A	62	LEU	11
1	A	90	THR	11
1	A	102	ARG	10
1	A	105	LYS	9
1	A	86	HIS	8
1	A	89	TYR	8
1	A	76	LEU	8
1	A	39	TYR	7
1	A	116	LEU	7
1	A	118	GLU	7
1	A	79	SER	6
1	A	103	ASN	6
1	A	45	ILE	6
1	A	18	SER	6
1	A	91	THR	6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	98	ARG	6
1	A	72	SER	6
1	A	6	PHE	6
1	A	107	TYR	5
1	A	31	VAL	5
1	A	23	HIS	5
1	A	80	SER	5
1	A	92	LYS	5
1	A	46	GLN	5
1	A	117	GLN	5
1	A	43	LEU	5
1	A	30	SER	5
1	A	38	GLU	5
1	A	74	SER	4
1	A	20	ASP	3
1	A	66	ARG	3
1	A	40	SER	3
1	A	33	SER	2
1	A	78	GLN	2
1	A	17	LYS	2
1	A	75	CYS	2
1	A	16	SER	2
1	A	104	GLU	2
1	A	82	LEU	2
1	A	24	LEU	2
1	A	64	PHE	1
1	A	25	THR	1
1	A	112	GLN	1
1	A	68	TYR	1
1	A	88	ASP	1
1	A	15	ILE	1
1	A	35	LYS	1
1	A	27	GLU	1
1	A	83	SER	1
1	A	69	CYS	1
1	A	67	VAL	1
1	A	65	MET	1
1	A	81	SER	1

### 6.3.3 RNA ⓘ

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

The completeness of assignment taking into account all chemical shift lists is 87% for the well-defined parts and 81% for the entire structure.

### 7.1 Chemical shift list 1

File name: 2m26\_cs.str

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

#### 7.1.1 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	1311
Number of shifts mapped to atoms	1311
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

#### 7.1.2 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	$0.01 \pm 0.07$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	105	$-0.08 \pm 0.15$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	85	$-0.41 \pm 0.20$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	102	$1.07 \pm 0.35$	Should be applied

#### 7.1.3 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 87%, i.e. 1028 atoms were assigned a chemical shift out of a possible 1176. 11 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	442/472 (94%)	184/187 (98%)	170/196 (87%)	88/89 (99%)
Sidechain	503/599 (84%)	304/357 (85%)	194/219 (89%)	5/23 (22%)

*Continued on next page...*

Continued from previous page...

	<b>Total</b>	<b><sup>1</sup>H</b>	<b><sup>13</sup>C</b>	<b><sup>15</sup>N</b>
Aromatic	83/105 (79%)	41/55 (75%)	40/46 (87%)	2/4 (50%)
Overall	1028/1176 (87%)	529/599 (88%)	404/461 (88%)	95/116 (82%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 81%, i.e. 1170 atoms were assigned a chemical shift out of a possible 1446. 11 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	<b>Total</b>	<b><sup>1</sup>H</b>	<b><sup>13</sup>C</b>	<b><sup>15</sup>N</b>
Backbone	514/606 (85%)	214/240 (89%)	198/252 (79%)	102/114 (89%)
Sidechain	573/735 (78%)	348/440 (79%)	218/268 (81%)	7/27 (26%)
Aromatic	83/105 (79%)	41/55 (75%)	40/46 (87%)	2/4 (50%)
Overall	1170/1446 (81%)	603/735 (82%)	456/566 (81%)	111/145 (77%)

#### 7.1.4 Statistically unusual chemical shifts [i](#)

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	7	PRO	CA	4.19	71.13 – 55.53	-37.9
1	A	26	TRP	NE1	115.62	139.19 – 119.59	-7.0
1	A	102	ARG	HB3	-0.15	3.17 – 0.37	-6.8

#### 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:

