



# Full wwPDB NMR Structure Validation Report ⓘ

Apr 21, 2024 – 05:07 AM EDT

PDB ID : 2LY1  
BMRB ID : 18211  
Title : NMR structure of the second and third lotus domains of tudor domain-containing protein 7 (NMR ensemble overlay for Lotus #2)  
Authors : Cui, G.; Botuyan, M.; Mer, G.  
Deposited on : 2012-09-10

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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) 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)  
wwPDB-RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
wwPDB-ShiftChecker : v1.2  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.36.2

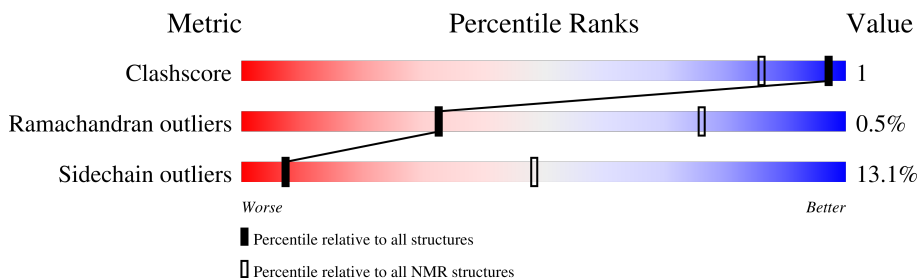
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 96%.

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	180	

## 2 Ensemble composition and analysis i

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:257-A:309, A:318-A:323 (59)	0.28	15
2	A:361-A:426 (66)	0.32	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 1 single-model cluster was found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2852 atoms, of which 1421 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Tudor domain-containing protein 7.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	180	2852	917	1421	238	270	6	0

There are 2 discrepancies between the modelled and reference sequences:

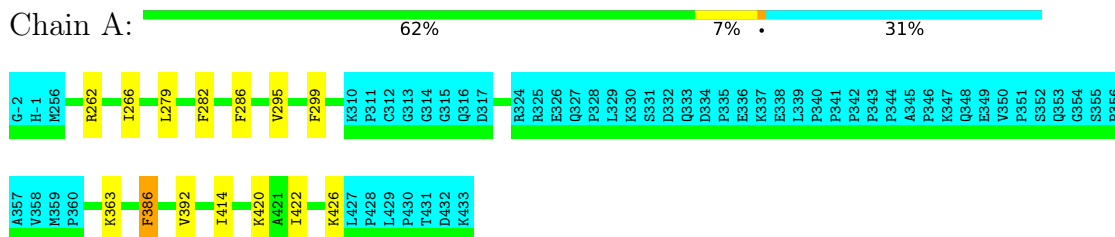
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP Q8K1H1
A	-1	HIS	-	expression tag	UNP Q8K1H1

## 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: Tudor domain-containing protein 7

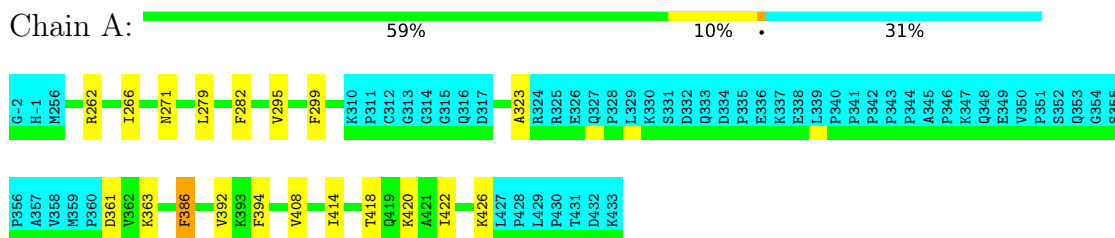


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

#### 4.2.1 Score per residue for model 1

- Molecule 1: Tudor domain-containing protein 7



#### 4.2.2 Score per residue for model 2

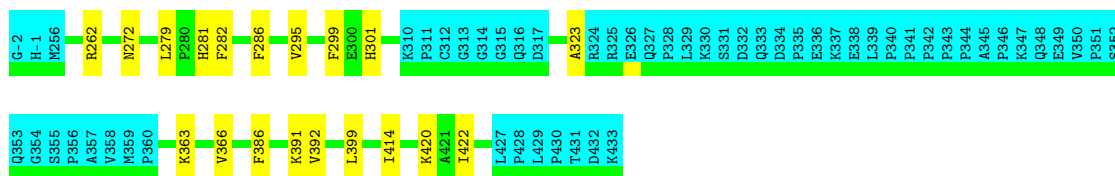
- Molecule 1: Tudor domain-containing protein 7





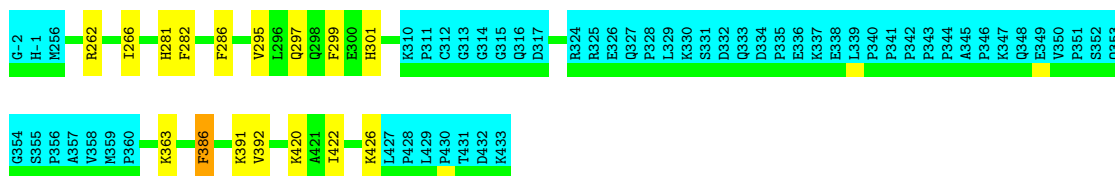
### 4.2.3 Score per residue for model 3

- Molecule 1: Tudor domain-containing protein 7



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

- Molecule 1: Tudor domain-containing protein 7



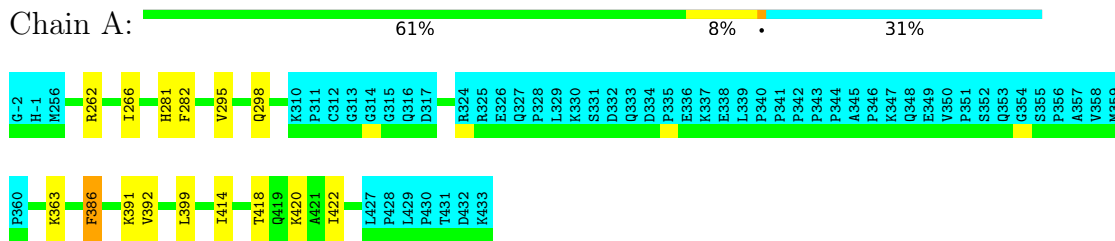
### 4.2.5 Score per residue for model 5

- Molecule 1: Tudor domain-containing protein 7



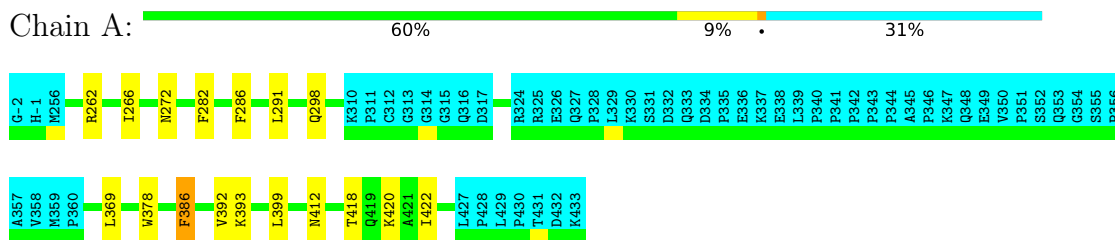
### 4.2.6 Score per residue for model 6

- Molecule 1: Tudor domain-containing protein 7



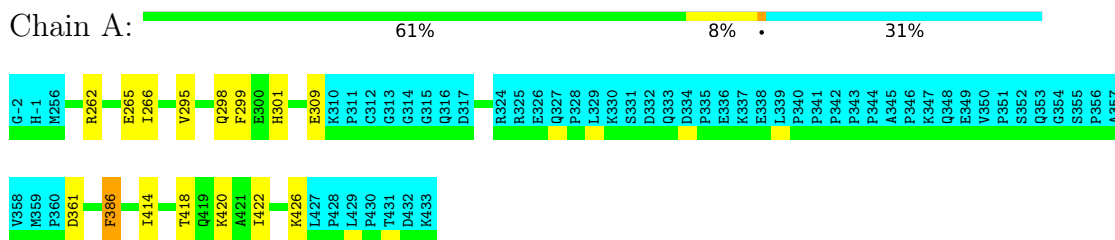
#### 4.2.7 Score per residue for model 7

- Molecule 1: Tudor domain-containing protein 7



#### 4.2.8 Score per residue for model 8

- Molecule 1: Tudor domain-containing protein 7



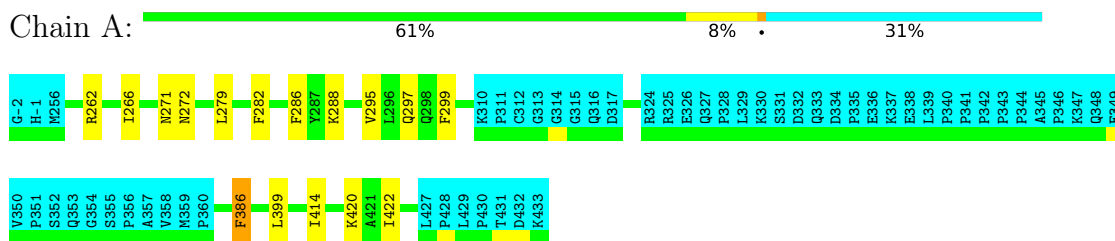
#### 4.2.9 Score per residue for model 9

- Molecule 1: Tudor domain-containing protein 7



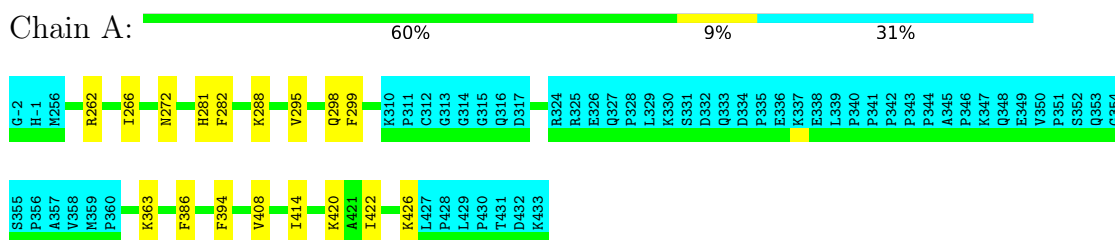
#### 4.2.10 Score per residue for model 10

- Molecule 1: Tudor domain-containing protein 7



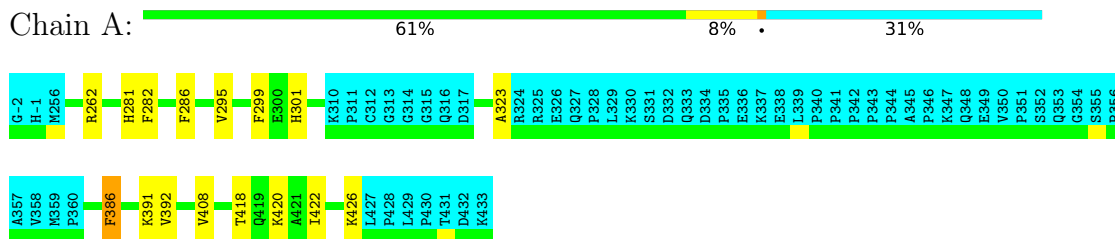
#### 4.2.11 Score per residue for model 11

- Molecule 1: Tudor domain-containing protein 7



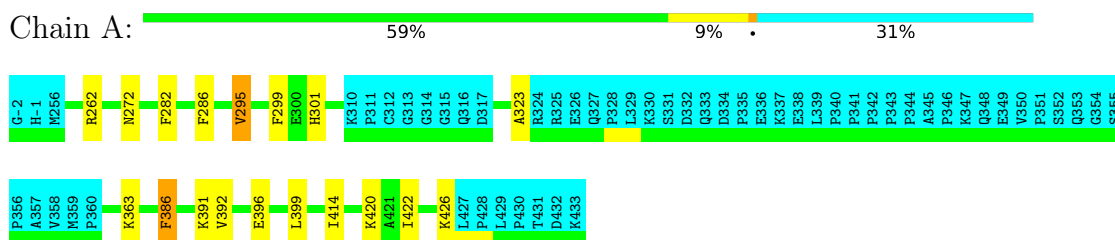
#### 4.2.12 Score per residue for model 12

- Molecule 1: Tudor domain-containing protein 7



#### 4.2.13 Score per residue for model 13

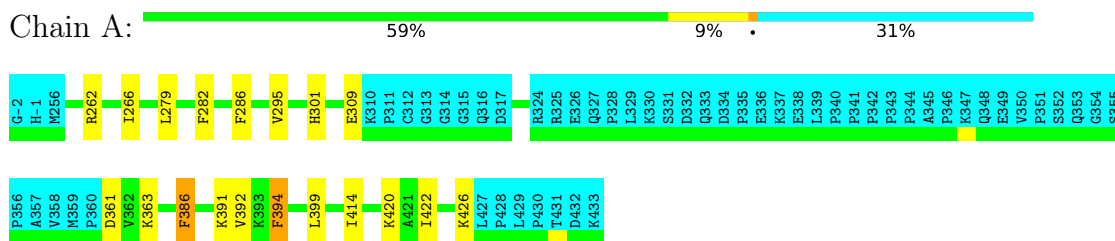
- Molecule 1: Tudor domain-containing protein 7





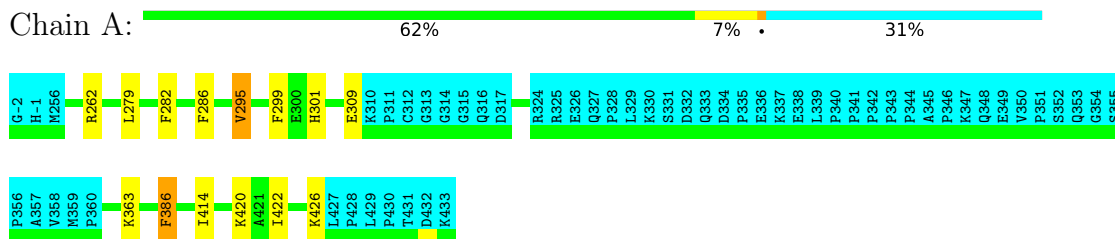
## 4.2.14 Score per residue for model 14

- Molecule 1: Tudor domain-containing protein 7



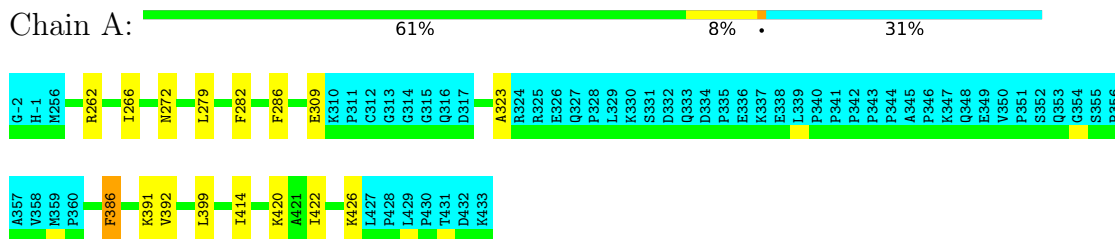
## 4.2.15 Score per residue for model 15

- Molecule 1: Tudor domain-containing protein 7



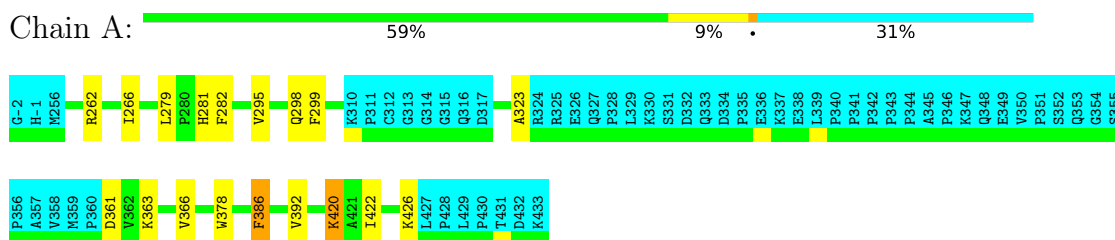
## 4.2.16 Score per residue for model 16

- Molecule 1: Tudor domain-containing protein 7



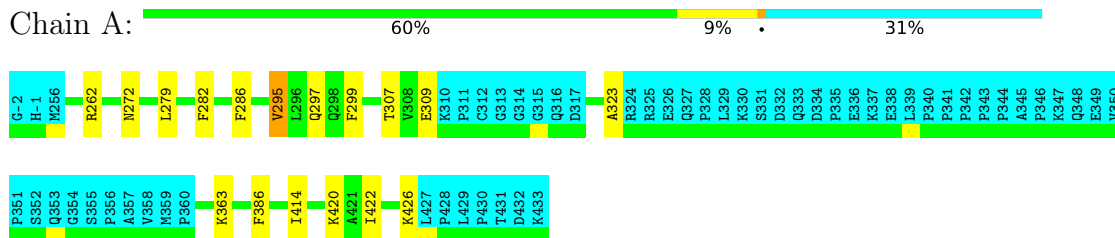
## 4.2.17 Score per residue for model 17

- Molecule 1: Tudor domain-containing protein 7



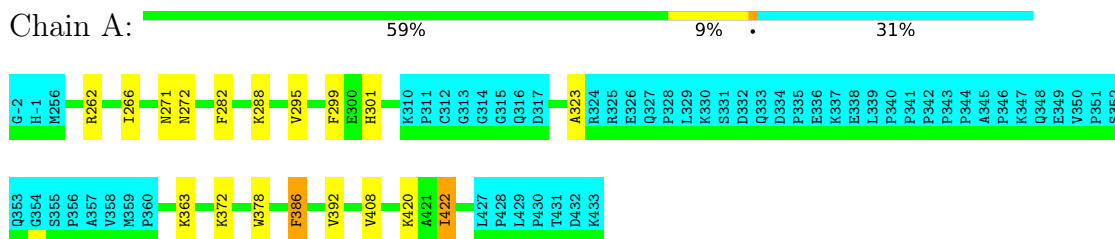
### 4.2.18 Score per residue for model 18

- Molecule 1: Tudor domain-containing protein 7



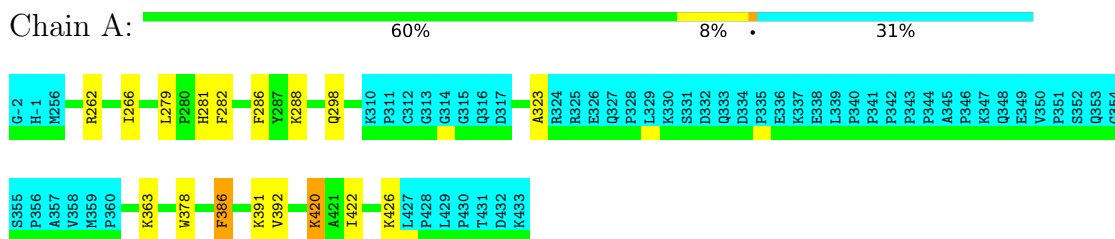
### 4.2.19 Score per residue for model 19

- Molecule 1: Tudor domain-containing protein 7



### 4.2.20 Score per residue for model 20

- Molecule 1: Tudor domain-containing protein 7



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing*.

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
CYANA	structure solution	
Amber	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	2329
Number of shifts mapped to atoms	2329
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	96%

## 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	0.67±0.00	0±0/1046 ( 0.0± 0.0%)	1.02±0.01	1±1/1417 ( 0.1± 0.0%)
All	All	0.67	0/20920 ( 0.0%)	1.02	21/28340 ( 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	0.1±0.4
All	All	0	3

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	386	PHE	CB-CG-CD2	-7.91	115.27	120.80	4	16
1	A	386	PHE	CB-CG-CD1	6.19	125.13	120.80	4	5

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	394	PHE	Sidechain	3

### 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	1019	1011	1011	1±1
All	All	20380	20220	20220	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 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:378:TRP:CZ3	1:A:420:LYS:HE3	0.64	2.27	17	2
1:A:386:PHE:CD2	1:A:394:PHE:CZ	0.57	2.92	14	1
1:A:378:TRP:CZ2	1:A:422:ILE:HD11	0.49	2.42	7	1
1:A:295:VAL:HG12	1:A:299:PHE:CE1	0.48	2.44	2	14
1:A:378:TRP:CZ3	1:A:420:LYS:CE	0.46	2.98	20	2
1:A:378:TRP:CD2	1:A:422:ILE:HD11	0.45	2.47	19	1
1:A:378:TRP:CE3	1:A:420:LYS:HE3	0.43	2.46	17	2
1:A:298:GLN:HB3	1:A:302:TRP:CE2	0.40	2.51	9	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	125/180 (69%)	120±1 (96±1%)	4±2 (3±1%)	1±1 (1±1%)	32	76
All	All	2500/3600 (69%)	2407 (96%)	80 (3%)	13 (1%)	32	76

All 2 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	323	ALA	10
1	A	271	ASN	3

### 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	111/159 (70%)	96±2 (87±2%)	15±2 (13±2%)	<b>7</b> 48
All	All	2220/3180 (70%)	1929 (87%)	291 (13%)	<b>7</b> 48

All 34 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	262	ARG	20
1	A	386	PHE	20
1	A	420	LYS	20
1	A	282	PHE	19
1	A	422	ILE	18
1	A	266	ILE	15
1	A	426	LYS	15
1	A	392	VAL	14
1	A	363	LYS	13
1	A	286	PHE	12
1	A	279	LEU	11
1	A	414	ILE	11
1	A	391	LYS	10
1	A	281	HIS	9
1	A	301	HIS	9
1	A	399	LEU	9
1	A	418	THR	8
1	A	272	ASN	8
1	A	298	GLN	7
1	A	408	VAL	6
1	A	309	GLU	6
1	A	295	VAL	6
1	A	361	ASP	5
1	A	288	LYS	5
1	A	297	GLN	4
1	A	396	GLU	2
1	A	366	VAL	2
1	A	291	LEU	1
1	A	369	LEU	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	393	LYS	1
1	A	412	ASN	1
1	A	265	GLU	1
1	A	307	THR	1
1	A	372	LYS	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 96% for the well-defined parts and 95% 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	2329
Number of shifts mapped to atoms	2329
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

#### 7.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	180	$-0.08 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	170	$0.30 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	175	$-0.04 \pm 0.07$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	157	$0.07 \pm 0.23$	None needed ( $< 0.5$ ppm)

#### 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 96%, i.e. 1673 atoms were assigned a chemical shift out of a possible 1746. 0 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	616/620 (99%)	248/250 (99%)	250/250 (100%)	118/120 (98%)
Sidechain	899/949 (95%)	612/616 (99%)	274/304 (90%)	13/29 (45%)

*Continued on next page...*



Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	158/177 (89%)	79/87 (91%)	76/83 (92%)	3/7 (43%)
Overall	1673/1746 (96%)	939/953 (99%)	600/637 (94%)	134/156 (86%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 95%, i.e. 2329 atoms were assigned a chemical shift out of a possible 2447. 0 out of 27 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	859/872 (99%)	347/351 (99%)	355/360 (99%)	157/161 (98%)
Sidechain	1308/1391 (94%)	888/900 (99%)	401/446 (90%)	19/45 (42%)
Aromatic	162/184 (88%)	81/91 (89%)	78/85 (92%)	3/8 (38%)
Overall	2329/2447 (95%)	1316/1342 (98%)	834/891 (94%)	179/214 (84%)

#### 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.

List Id	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	420	LYS	HE3	1.26	1.92 – 3.89	-8.4
1	A	260	GLN	HG3	0.33	0.91 – 3.68	-7.1
1	A	264	LYS	CG	32.40	19.35 – 30.45	6.8
1	A	260	GLN	HE21	4.36	5.02 – 9.43	-6.5
1	A	262	ARG	HD3	1.45	1.81 – 4.39	-6.4
1	A	395	PRO	HD3	1.75	1.76 – 5.48	-5.0

#### 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. If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:

