W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

Full wwPDB NMR Structure Validation Report ⓘ

Feb 13, 2017 – 01:10 am GMT

PDB ID	:	2MMY
Title	:	Solution structure of the RNA recognition motif of human TAF15
Authors	:	Kashyap, M.; Ganguly, A.K.; Bhavesh, N.S.
Deposited on	:	2014-03-22

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 ⓘ symbol.

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

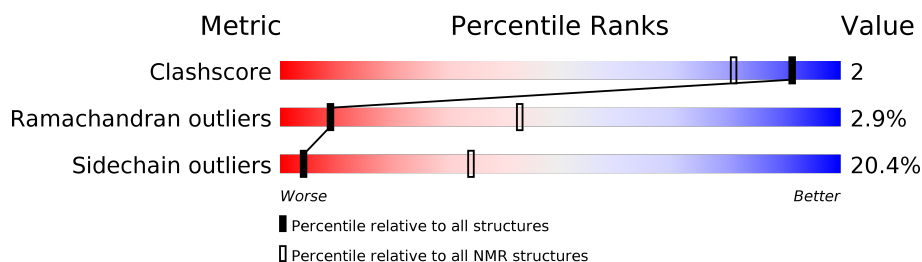|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (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 92%.

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	Percentile Ranks	Value
Clashscore		2
Ramachandran outliers		2.9%
Sidechain outliers		20.4%

Worse   *Better*

▮ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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 >=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	97	73% 14% • 11%

2   Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 17 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:233-A:318 (86)	0.17	17

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. No single-model clusters were found.

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

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 1485 atoms, of which 737 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called TATA-binding protein-associated factor 2N.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	97	1485	472	737	127	147	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	227	GLY	-	EXPRESSION TAG	UNP Q92804
A	228	SER	-	EXPRESSION TAG	UNP Q92804
A	229	HIS	-	EXPRESSION TAG	UNP Q92804
A	230	MET	-	EXPRESSION TAG	UNP Q92804

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.

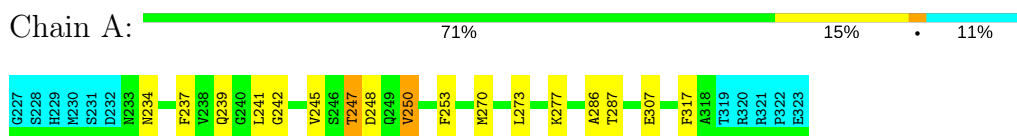• Molecule 1: TATA-binding protein-associated factor 2N
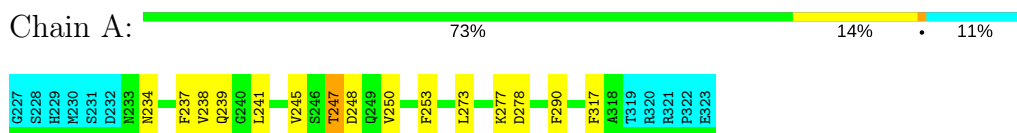
4.2 Scores per residue for each member of the ensemble
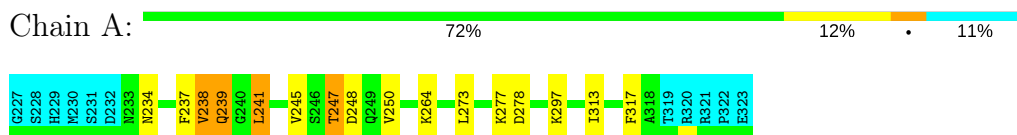
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

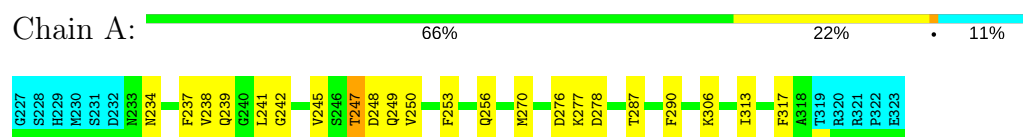• Molecule 1: TATA-binding protein-associated factor 2N

4.2.2 Score per residue for model 2

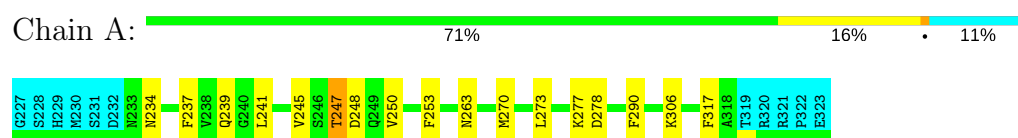• Molecule 1: TATA-binding protein-associated factor 2N

4.2.3 Score per residue for model 3

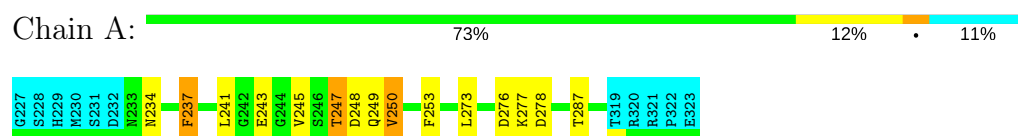• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:

4.2.5 Score per residue for model 5

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:

4.2.6 Score per residue for model 6

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:

4.2.7 Score per residue for model 7

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:

4.2.8 Score per residue for model 8

• Molecule 1: TATA-binding protein-associated factor 2N

4.2.9 Score per residue for model 9

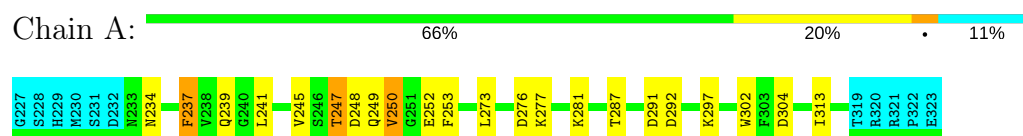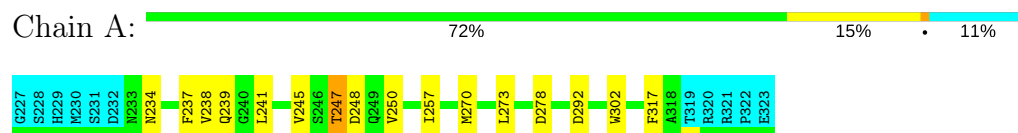• Molecule 1: TATA-binding protein-associated factor 2N

4.2.10 Score per residue for model 10

• Molecule 1: TATA-binding protein-associated factor 2N

4.2.11 Score per residue for model 11

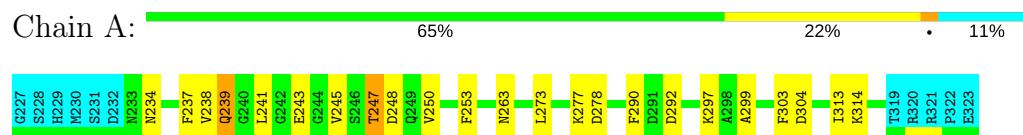• Molecule 1: TATA-binding protein-associated factor 2N

4.2.12 Score per residue for model 12

• Molecule 1: TATA-binding protein-associated factor 2N

4.2.13 Score per residue for model 13

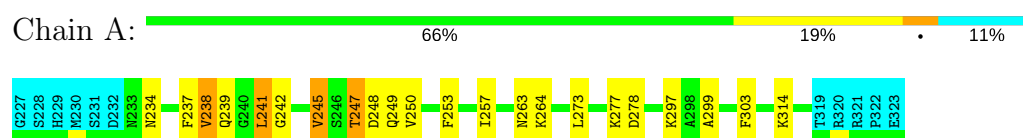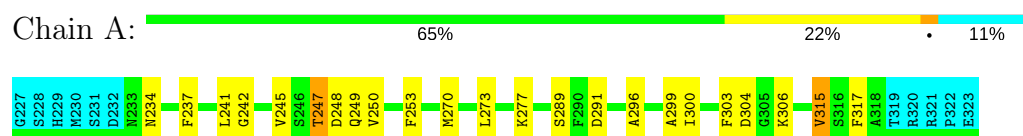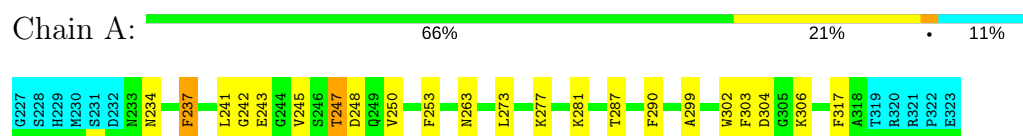• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:

4.2.14 Score per residue for model 14

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:
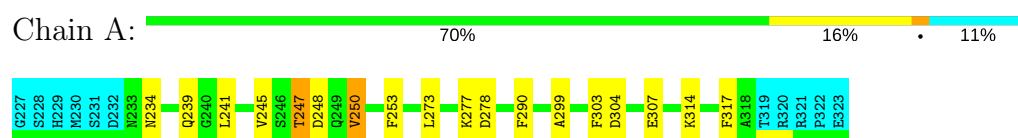
4.2.15 Score per residue for model 15

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:

4.2.16 Score per residue for model 16

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:

4.2.17 Score per residue for model 17 (medoid)

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:

4.2.18 Score per residue for model 18

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:
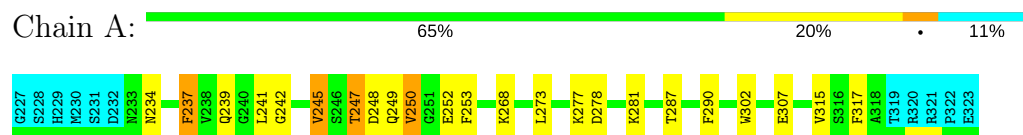
4.2.19 Score per residue for model 19

• Molecule 1: TATA-binding protein-associated factor 2N

Chain A:
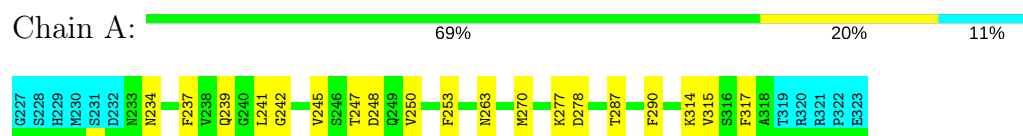
4.2.20 Score per residue for model 20
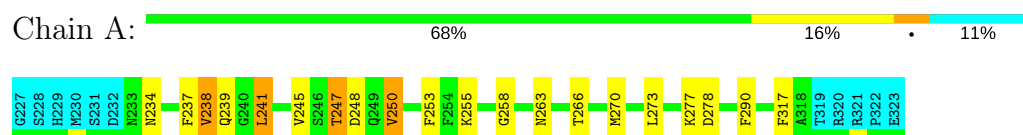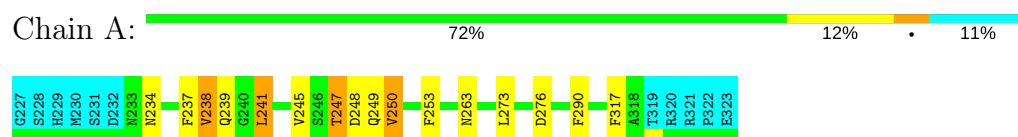
• Molecule 1: TATA-binding protein-associated factor 2N
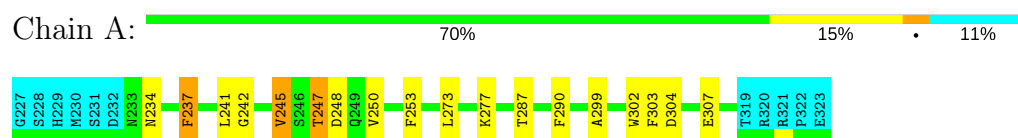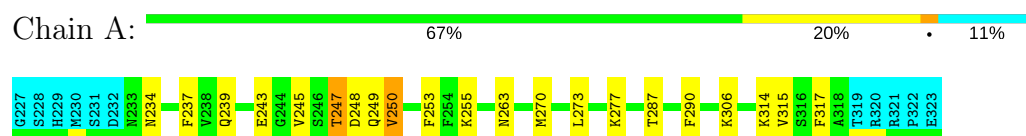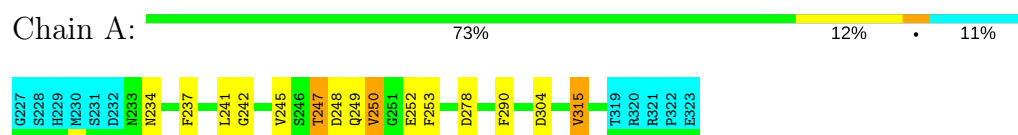
Chain A:
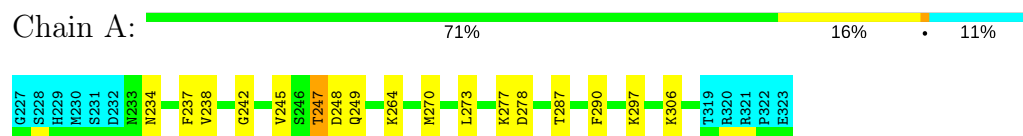
5 Refinement protocol and experimental data overview ⓘ

The models were refined using the following method: *distance geometry, 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	
CYANA	geometry optimization	
CYANA	refinement	
CNS	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)	2mmy_cs.str
Number of chemical shift lists	1
Total number of shifts	1145
Number of shifts mapped to atoms	1145
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	92%

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

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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 ⓘ

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	660	655	654	2±1
All	All	13200	13100	13080	46

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:238:VAL:HB	1:A:241:LEU:HD11	0.59	1.75	2	4
1:A:299:ALA:O	1:A:303:PHE:HB3	0.51	2.05	15	6
1:A:304:ASP:HB2	1:A:315:VAL:HG12	0.48	1.84	10	2
1:A:250:VAL:O	1:A:253:PHE:HB3	0.47	2.09	6	17
1:A:237:PHE:HA	1:A:287:THR:HG22	0.47	1.87	18	6
1:A:241:LEU:HD12	1:A:286:ALA:HB3	0.45	1.88	1	1
1:A:241:LEU:HD13	1:A:250:VAL:HG11	0.43	1.90	1	1
1:A:249:GLN:O	1:A:252:GLU:HG2	0.43	2.13	17	1
1:A:239:GLN:O	1:A:313:ILE:HA	0.42	2.15	3	3
1:A:242:GLY:O	1:A:245:VAL:HB	0.42	2.15	9	3
1:A:296:ALA:O	1:A:300:ILE:HG13	0.41	2.16	10	1
1:A:255:LYS:HD2	1:A:258:GLY:O	0.40	2.16	12	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	86/97 (89%)	78±1 (90±1%)	6±1 (7±1%)	3±1 (3±1%)	9	43
All	All	1720/1940 (89%)	1550 (90%)	120 (7%)	50 (3%)	9	43

All 4 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	248	ASP	20
1	A	247	THR	20
1	A	242	GLY	7
1	A	243	GLU	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	71/81 (88%)	57±3 (80±4%)	15±3 (20±4%)	4	34
All	All	1420/1620 (88%)	1130 (80%)	290 (20%)	4	34

All 38 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	245	VAL	20
1	A	234	ASN	20
1	A	247	THR	19
1	A	237	PHE	18
1	A	277	LYS	17
1	A	273	LEU	17

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	241	LEU	17
1	A	278	ASP	13
1	A	317	PHE	13
1	A	239	GLN	13
1	A	290	PHE	13
1	A	250	VAL	11
1	A	270	MET	9
1	A	249	GLN	9
1	A	238	VAL	8
1	A	263	ASN	8
1	A	306	LYS	6
1	A	304	ASP	5
1	A	315	VAL	5
1	A	302	TRP	5
1	A	297	LYS	5
1	A	314	LYS	5
1	A	287	THR	4
1	A	276	ASP	4
1	A	307	GLU	4
1	A	281	LYS	3
1	A	292	ASP	3
1	A	264	LYS	3
1	A	291	ASP	2
1	A	252	GLU	2
1	A	257	ILE	2
1	A	255	LYS	1
1	A	256	GLN	1
1	A	268	LYS	1
1	A	313	ILE	1
1	A	266	THR	1
1	A	289	SER	1
1	A	243	GLU	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

6.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 92% for the well-defined parts and 89% for the entire structure.

7.1 Chemical shift list 1

File name: 2mmy_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	1145
Number of shifts mapped to atoms	1145
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 ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	97	-0.12 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_\beta$	85	-0.17 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	96	-0.02 ± 0.09	None needed (< 0.5 ppm)
^{15}N	91	0.32 ± 0.45	None needed (< 0.5 ppm)

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 92%, i.e. 936 atoms were assigned a chemical shift out of a possible 1021. 5 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	420/422 (100%)	167/168 (99%)	171/172 (99%)	82/82 (100%)
Sidechain	433/508 (85%)	273/297 (92%)	152/192 (79%)	8/19 (42%)

Continued on next page...

Continued from previous page...

	Total	**¹H**	**¹³C**	**¹⁵N**
Aromatic	83/91 (91%)	43/49 (88%)	39/39 (100%)	1/3 (33%)
Overall	936/1021 (92%)	483/514 (94%)	362/403 (90%)	91/104 (88%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 89%, i.e. 1029 atoms were assigned a chemical shift out of a possible 1159. 5 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**¹H**	**¹³C**	**¹⁵N**
Backbone	469/475 (99%)	185/189 (98%)	193/194 (99%)	91/92 (99%)
Sidechain	473/585 (81%)	298/344 (87%)	167/216 (77%)	8/25 (32%)
Aromatic	87/99 (88%)	45/53 (85%)	41/41 (100%)	1/5 (20%)
Overall	1029/1159 (89%)	528/586 (90%)	401/451 (89%)	100/122 (82%)

7.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

7.1.5 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: