



# Full wwPDB NMR Structure Validation Report ⓘ

Aug 31, 2017 – 12:13 AM EDT

PDB ID : 5MML  
Title : HYL-20k  
Authors : Hexnerova, R.  
Deposited on : unknown

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	:	rb-20029824
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20029824

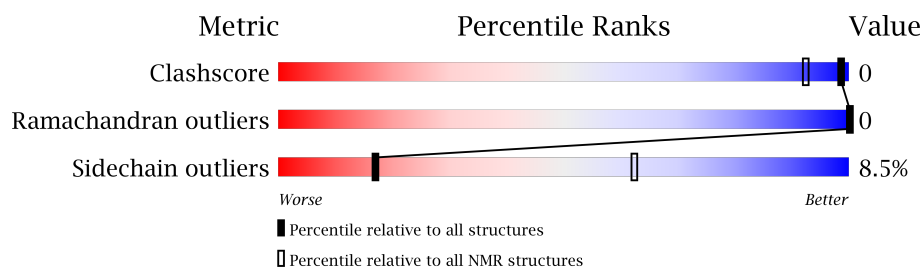
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*


The overall completeness of chemical shifts assignment is 59%.

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	16	

## 2 Ensemble composition and analysis

This entry contains 30 models. Model 11 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:2-A:14 (13)	0.18	11

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 5 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 5, 8, 10, 11, 12, 13, 14, 15, 21, 25, 28, 29, 30
2	3, 4, 6, 7, 9, 16, 19, 22
3	17, 18, 27
4	2, 23
5	24, 26
Single-model clusters	20

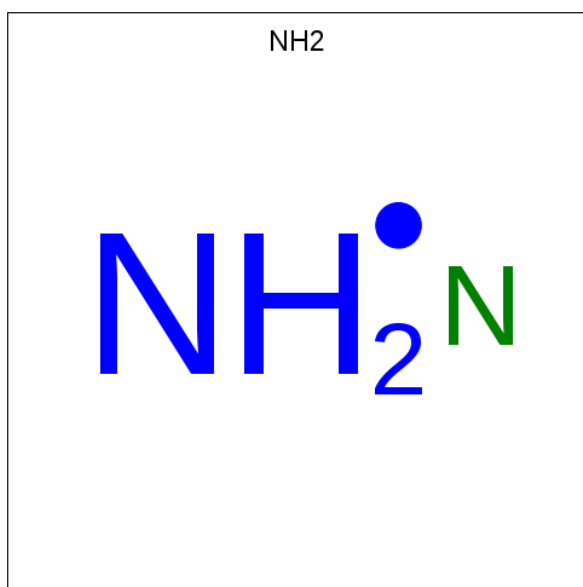
### 3 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 292 atoms, of which 163 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS.

Mol	Chain	Residues	Atoms					Trace
1	A	16	Total	C	H	N	O	0
			289	88	161	22	18	

- Molecule 2 is AMINO GROUP (three-letter code: NH2) (formula: H<sub>2</sub>N).



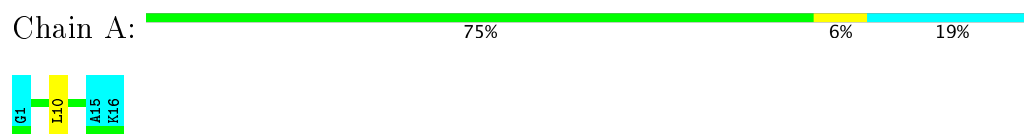
Mol	Chain	Residues	Atoms		
2	A	1	Total	H	N
			3	2	1

## 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: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

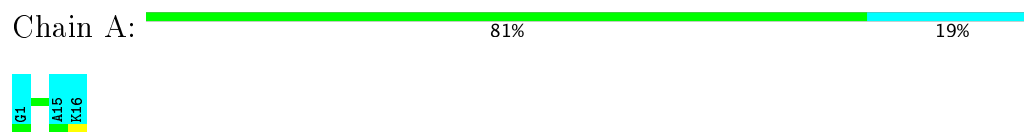


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

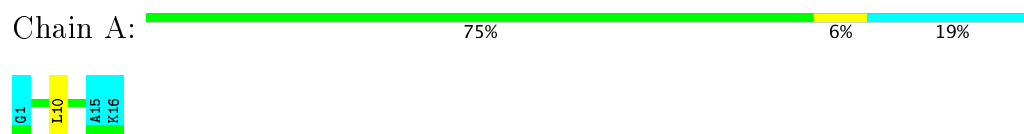
#### 4.2.1 Score per residue for model 1

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS




#### 4.2.2 Score per residue for model 2

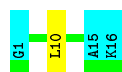
- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS



### 4.2.3 Score per residue for model 3

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

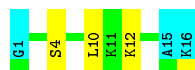
Chain A:  75% 6% 19%



### 4.2.4 Score per residue for model 4


- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

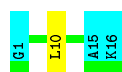
Chain A:  63% 19% 19%



### 4.2.5 Score per residue for model 5


- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  75% 6% 19%



### 4.2.6 Score per residue for model 6

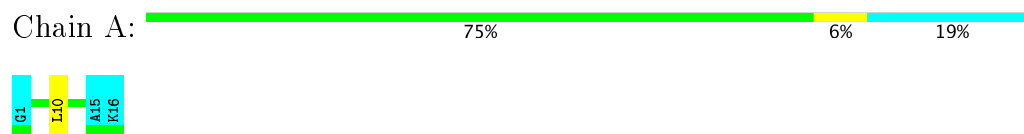
- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  81% 19%



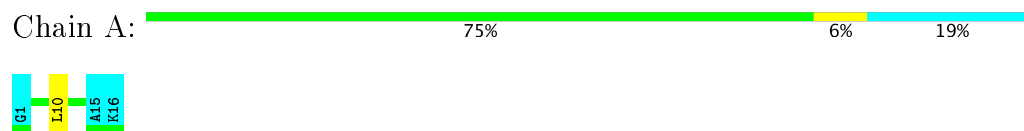
### 4.2.7 Score per residue for model 7

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS



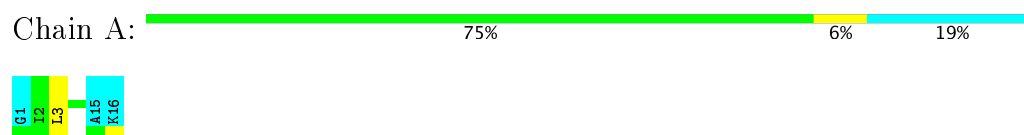
#### 4.2.8 Score per residue for model 8

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS



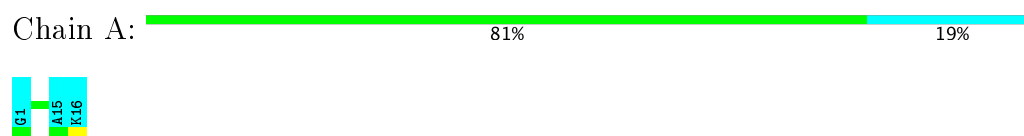
#### 4.2.9 Score per residue for model 9

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS



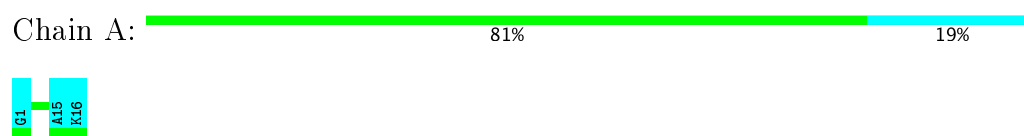
#### 4.2.10 Score per residue for model 10

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS




#### 4.2.11 Score per residue for model 11 (medoid)

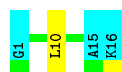
- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS



#### 4.2.12 Score per residue for model 12

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  75% 6% 19%



#### 4.2.13 Score per residue for model 13

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  75% 6% 19%



#### 4.2.14 Score per residue for model 14

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

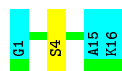
Chain A:  69% 13% 19%



#### 4.2.15 Score per residue for model 15

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  75% 6% 19%



#### 4.2.16 Score per residue for model 16

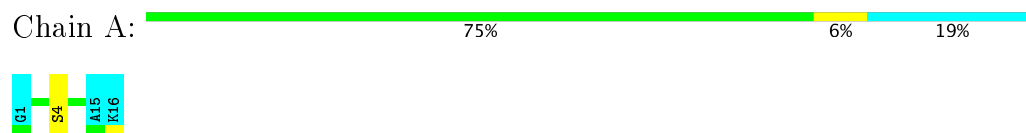
- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS





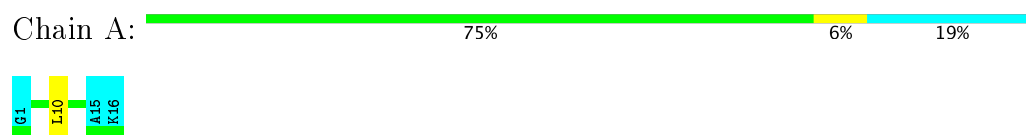
#### 4.2.17 Score per residue for model 17

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS



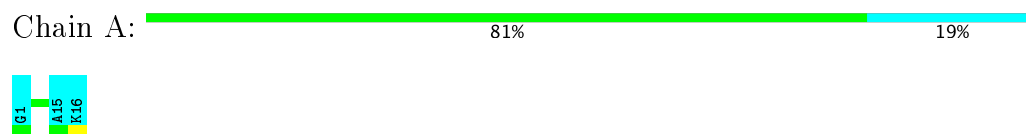
#### 4.2.18 Score per residue for model 18

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS



#### 4.2.19 Score per residue for model 19

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS




#### 4.2.20 Score per residue for model 20

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS



#### 4.2.21 Score per residue for model 21

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

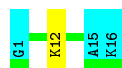
Chain A:  75% 6% 19%



#### 4.2.22 Score per residue for model 22


- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

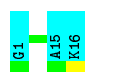
Chain A:  75% 6% 19%



#### 4.2.23 Score per residue for model 23

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

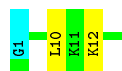
Chain A:  81% 19%



#### 4.2.24 Score per residue for model 24


- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  69% 13% 19%



#### 4.2.25 Score per residue for model 25

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  81% 19%



#### 4.2.26 Score per residue for model 26

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

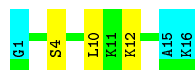
Chain A:  63% 19% 19%



#### 4.2.27 Score per residue for model 27


- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  63% 19% 19%



#### 4.2.28 Score per residue for model 28

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  81% 19%



#### 4.2.29 Score per residue for model 29


- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

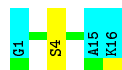
Chain A:  75% 6% 19%



#### 4.2.30 Score per residue for model 30

- Molecule 1: GLY-ILE-LEU-SER-SER-LEU-TRP-LYS-LYS-LEU-LYS-LYS-ILE-ILE-ALA-LYS

Chain A:  75% 6% 19%



## 5 Refinement protocol and experimental data overview ⓘ

Of the ? calculated structures, 30 were deposited, based on the following criterion: ?.

The authors did not provide any information on software used for structure solution, optimization or 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)	5mml_cs.cif
Number of chemical shift lists	1
Total number of shifts	157
Number of shifts mapped to atoms	157
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	59%

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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NH2

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 [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	110	138	138	0±0
2	A	1	2	0	0±0
All	All	3330	4200	4140	1

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:12:LYS:HZ3	2:A:101:NH2:N	0.44	2.10	24	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	13/16 (81%)	13±0 (100±0%)	0±0 (0±0%)	0±0 (0±0%)	100	100
All	All	390/480 (81%)	390 (100%)	0 (0%)	0 (0%)	100	100

There are no Ramachandran outliers.

### 6.3.2 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	13/14 (93%)	12±1 (92±8%)	1±1 (8±8%)	17	62
All	All	390/420 (93%)	357 (92%)	33 (8%)	17	62

All 6 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	10	LEU	16
1	A	4	SER	8
1	A	3	LEU	3
1	A	12	LYS	3
1	A	6	LEU	2
1	A	9	LYS	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

Of 1 ligands modelled in this entry, 1 is modelled with single atom - leaving 0 for Mogul analysis.

## 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 59% for the well-defined parts and 59% for the entire structure.

### 7.1 Chemical shift list 1

File name: 5mml\_cs.cif

Chemical shift list name: *lb10.star*

#### 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	157
Number of shifts mapped to atoms	157
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

No chemical shift referencing corrections were calculated (not enough data).

#### 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 59%, i.e. 114 atoms were assigned a chemical shift out of a possible 192. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	39/65 (60%)	26/26 (100%)	0/26 (0%)	13/13 (100%)
Sidechain	68/115 (59%)	68/68 (100%)	0/43 (0%)	0/4 (0%)
Aromatic	7/12 (58%)	6/6 (100%)	0/5 (0%)	1/1 (100%)
Overall	114/192 (59%)	100/100 (100%)	0/74 (0%)	14/18 (78%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 59%, i.e. 130 atoms were assigned a chemical shift out of a possible 222. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	46/80 (58%)	31/32 (97%)	0/32 (0%)	15/16 (94%)
Sidechain	77/130 (59%)	77/77 (100%)	0/48 (0%)	0/5 (0%)
Aromatic	7/12 (58%)	6/6 (100%)	0/5 (0%)	1/1 (100%)
Overall	130/222 (59%)	114/115 (99%)	0/85 (0%)	16/22 (73%)

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

There are no statistically unusual chemical shifts.

#### 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:

