



# Full wwPDB NMR Structure Validation Report ⓘ

Mar 6, 2018 – 10:47 PM EST

PDB ID : 5N9Q  
Title : Structure of A. thaliana RCD1(468-567)  
Authors : Tossavainen, H.; Hellman, M.; Vainonen, J.; Kangasjarvi, J.; Permi, P.  
Deposited on : 2017-02-27

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-20030736  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030736

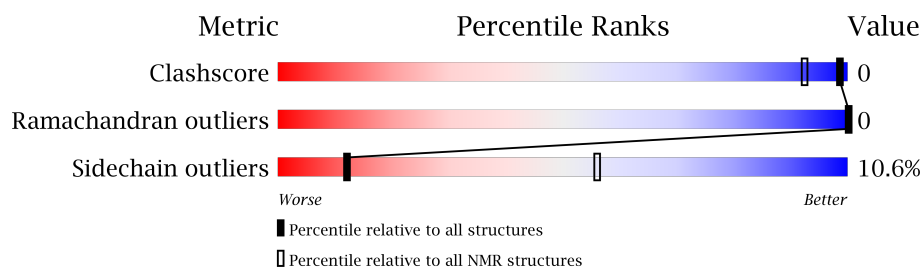
# 1 Overall quality at a glance

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	128	

## 2 Ensemble composition and analysis

This entry contains 15 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:513-A:562 (50)	0.33	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 4 clusters and 3 single-model clusters were found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1960 atoms, of which 983 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Inactive poly [ADP-ribose] polymerase RCD1.

Mol	Chain	Residues	Atoms						Trace
1	A	128	Total	C	H	N	O	S	0
			1960	606	983	176	191	4	

There are 6 discrepancies between the modelled and reference sequences:

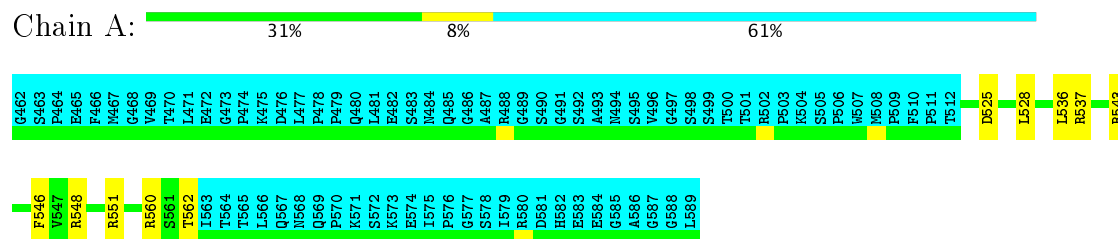
Chain	Residue	Modelled	Actual	Comment	Reference
A	462	GLY	-	expression tag	UNP Q8RY59
A	463	SER	-	expression tag	UNP Q8RY59
A	464	PRO	-	expression tag	UNP Q8RY59
A	465	GLU	-	expression tag	UNP Q8RY59
A	466	PHE	-	expression tag	UNP Q8RY59
A	467	MET	-	expression tag	UNP Q8RY59

## 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: Inactive poly [ADP-ribose] polymerase RCD1

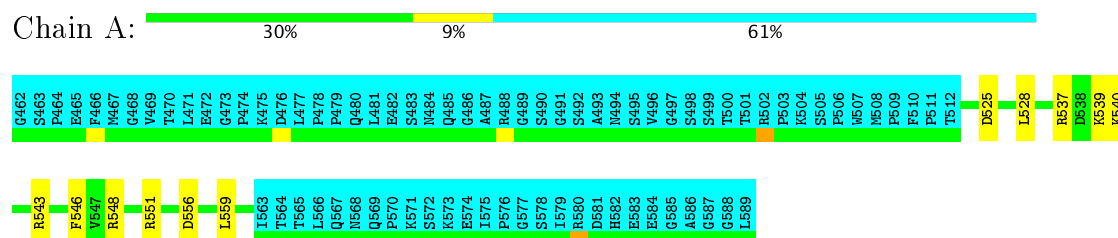


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

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

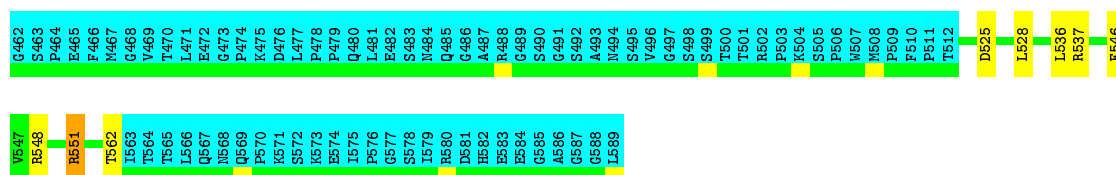
- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1



#### 4.2.2 Score per residue for model 2

- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1

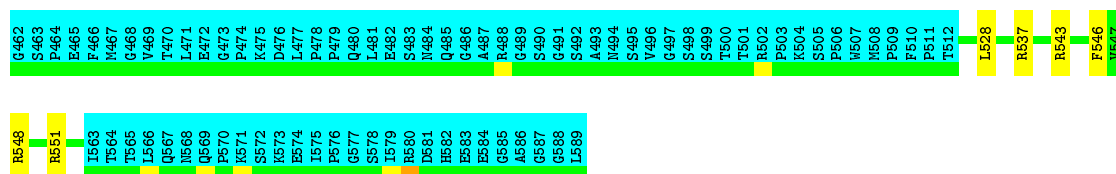




### 4.2.3 Score per residue for model 3

- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1

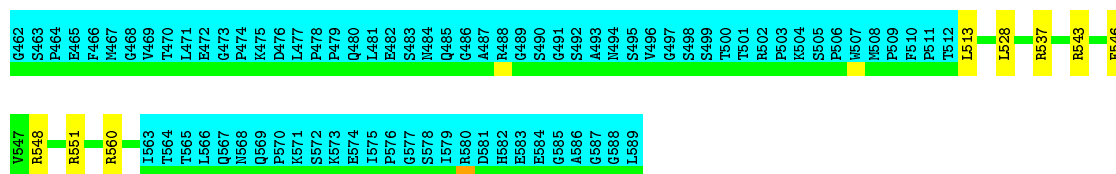
Chain A: 34% 5% 61%



### 4.2.4 Score per residue for model 4

- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1

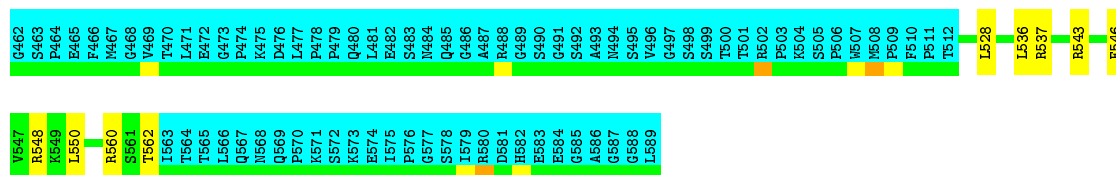
Chain A: 33% 6% 61%



### 4.2.5 Score per residue for model 5

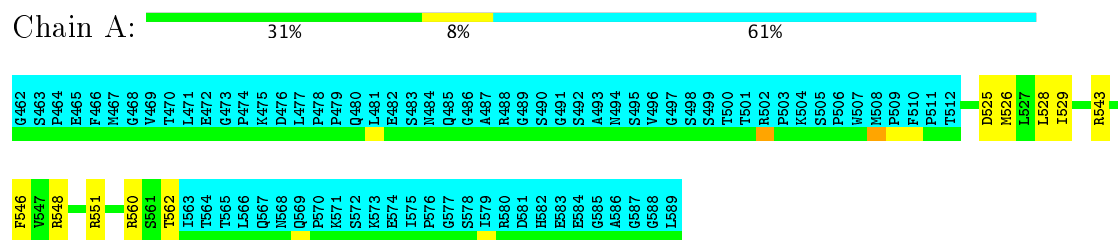
- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1

Chain A: 32% 7% 61%



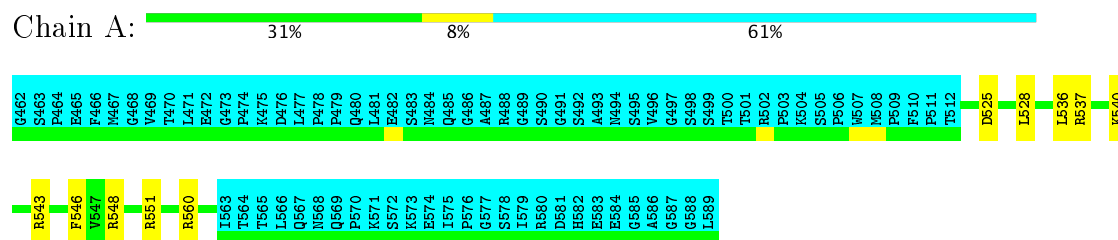
### 4.2.6 Score per residue for model 6

- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1



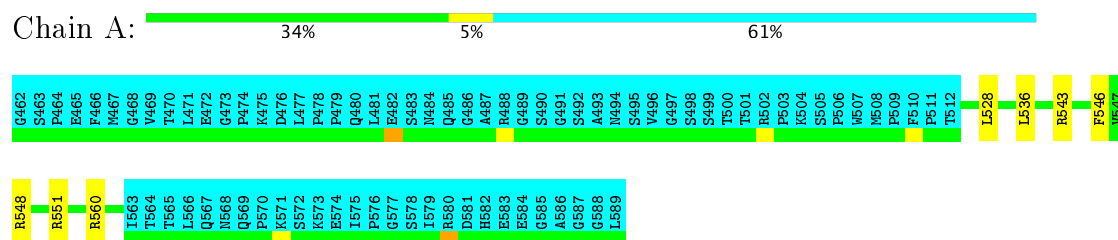
#### 4.2.7 Score per residue for model 7

- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1



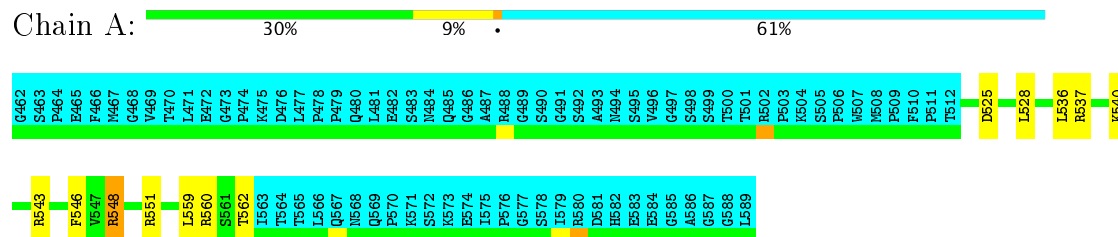
#### 4.2.8 Score per residue for model 8

- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1



#### 4.2.9 Score per residue for model 9

- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1



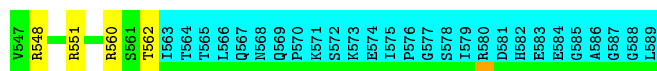
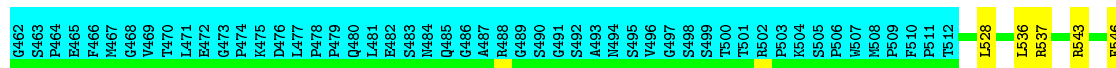






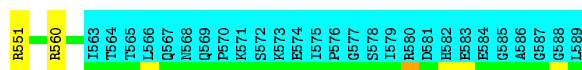
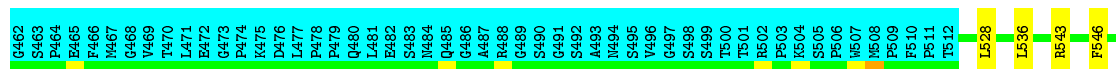
#### 4.2.14 Score per residue for model 14

- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1



#### 4.2.15 Score per residue for model 15

- Molecule 1: Inactive poly [ADP-ribose] polymerase RCD1



## 5 Refinement protocol and experimental data overview

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

Of the 300 calculated structures, 15 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 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 7 of this report.

Chemical shift file(s)	5n9q_cs.cif
Number of chemical shift lists	1
Total number of shifts	1434
Number of shifts mapped to atoms	1434
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

### 6.1 Standard geometry

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	0.70±0.01	0±0/410 (0.0±0.0%)	1.21±0.04	4±1/550 (0.8±0.2%)
All	All	0.70	0/6150 (0.0%)	1.21	64/8250 (0.8%)

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±0.3
All	All	0	2

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	543	ARG	NE-CZ-NH1	9.77	125.19	120.30	14	14
1	A	551	ARG	NE-CZ-NH1	9.46	125.03	120.30	2	13
1	A	560	ARG	NE-CZ-NH1	9.02	124.81	120.30	9	11
1	A	548	ARG	NE-CZ-NH1	8.55	124.58	120.30	4	12
1	A	537	ARG	NE-CZ-NH1	8.21	124.40	120.30	10	11
1	A	537	ARG	NE-CZ-NH2	-5.81	117.40	120.30	14	1
1	A	543	ARG	NE-CZ-NH2	-5.64	117.48	120.30	4	2

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	533	TYR	Sidechain	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group	Models (Total)
1	A	537	ARG	Sidechain	1

## 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	406	427	427	0±0
All	All	6090	6405	6405	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:526:MET:HA	1:A:529:ILE:HD12	0.47	1.85	6	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	50/128 (39%)	48±1 (97±2%)	2±1 (3±2%)	0±0 (0±0%)	100	100
All	All	750/1920 (39%)	727 (97%)	23 (3%)	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	44/108 (41%)	39±2 (89±3%)	5±2 (11±3%)	12 56
All	All	660/1620 (41%)	590 (89%)	70 (11%)	12 56

All 14 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	546	PHE	15
1	A	528	LEU	15
1	A	536	LEU	11
1	A	562	THR	8
1	A	525	ASP	8
1	A	540	LYS	3
1	A	559	LEU	2
1	A	551	ARG	2
1	A	558	LEU	1
1	A	513	LEU	1
1	A	539	LYS	1
1	A	556	ASP	1
1	A	548	ARG	1
1	A	550	LEU	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 [i](#)

The completeness of assignment taking into account all chemical shift lists is 87% for the well-defined parts and 82% for the entire structure.

### 7.1 Chemical shift list 1

File name: 5n9q\_cs.cif

Chemical shift list name: *RST.bmr.v3.1*

#### 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	1434
Number of shifts mapped to atoms	1434
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 [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	120	$-0.17 \pm 0.13$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	109	$0.23 \pm 0.11$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	114	$-0.12 \pm 0.07$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	114	$-0.38 \pm 0.21$	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 87%, i.e. 571 atoms were assigned a chemical shift out of a possible 659. 11 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	242/250 (97%)	98/100 (98%)	96/100 (96%)	48/50 (96%)
Sidechain	307/376 (82%)	191/218 (88%)	112/135 (83%)	4/23 (17%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	22/33 (67%)	15/18 (83%)	7/14 (50%)	0/1 (0%)
Overall	571/659 (87%)	304/336 (90%)	215/249 (86%)	52/74 (70%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 82%, i.e. 1264 atoms were assigned a chemical shift out of a possible 1545. 18 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	565/620 (91%)	224/246 (91%)	234/256 (91%)	107/118 (91%)
Sidechain	658/855 (77%)	407/506 (80%)	243/306 (79%)	8/43 (19%)
Aromatic	41/70 (59%)	25/38 (66%)	15/29 (52%)	1/3 (33%)
Overall	1264/1545 (82%)	656/790 (83%)	492/591 (83%)	116/164 (71%)

#### 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:

