Full wwPDB NMR Structure Validation Report (i)

Feb 24, 2021 – 10:03 AM EST

PDB ID	:	7KLO
Title	:	Solution structure of the PHD1 domain of histone demethylase KDM5A
Authors	:	Longbotham, E.J.; Kelly, M.J.S.; Fujimori, D.G.
Deposited on	:	2020-10-30

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

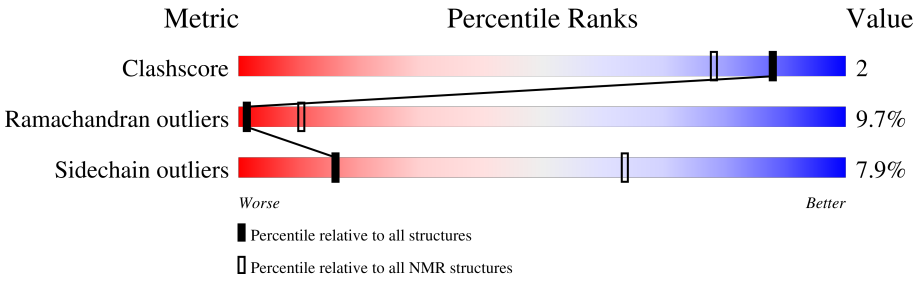
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.17.1
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.17.1

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 84%.

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 >=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	59	69% 19% 12%

2 Ensemble composition and analysis (i)

This entry contains 20 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:293-A:344 (52)	0.48	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 2 clusters. No single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 884 atoms, of which 424 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Lysine-specific demethylase 5A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	59	882	288	424	74	88	8	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	286	GLY	-	expression tag	UNP P29375

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

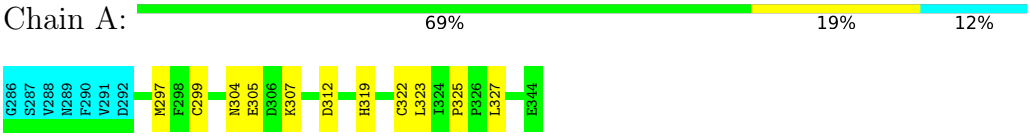
Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	2	2	2

4 Residue-property plots

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: Lysine-specific demethylase 5A

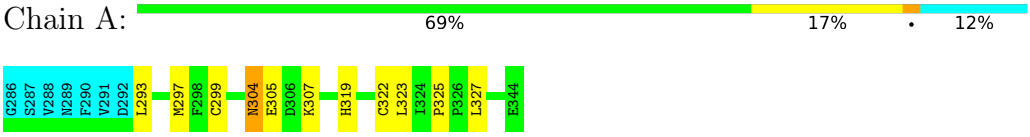


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

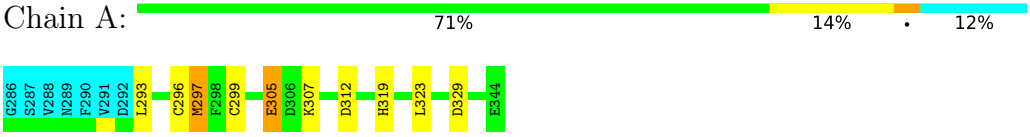
4.2.1 Score per residue for model 1 (medoid)

• Molecule 1: Lysine-specific demethylase 5A



4.2.2 Score per residue for model 2

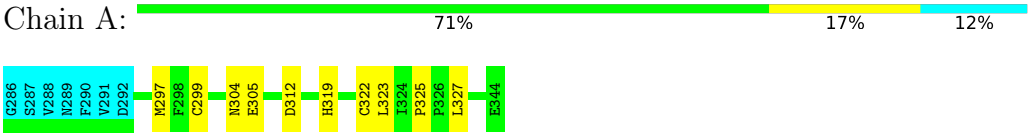
• Molecule 1: Lysine-specific demethylase 5A



4.2.3 Score per residue for model 3

• Molecule 1: Lysine-specific demethylase 5A

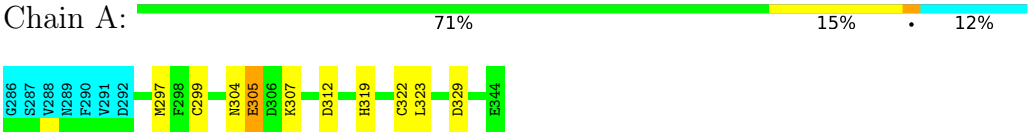
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Lysine-specific demethylase 5A

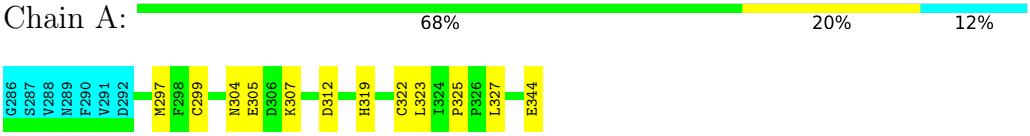
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Lysine-specific demethylase 5A

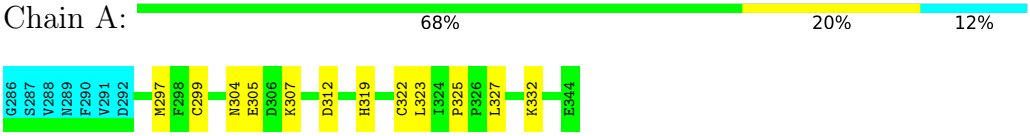
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Lysine-specific demethylase 5A

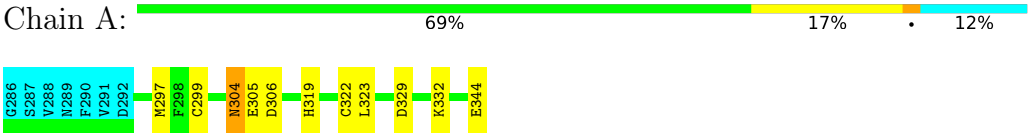
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Lysine-specific demethylase 5A

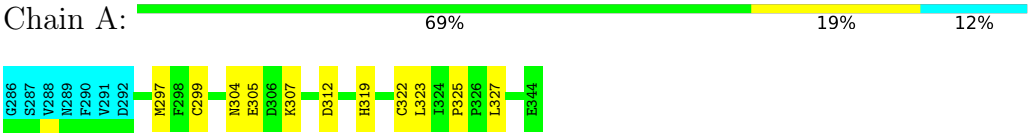
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Lysine-specific demethylase 5A

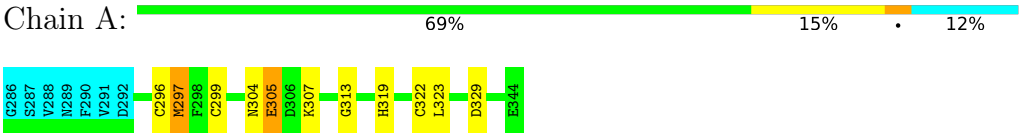
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Lysine-specific demethylase 5A

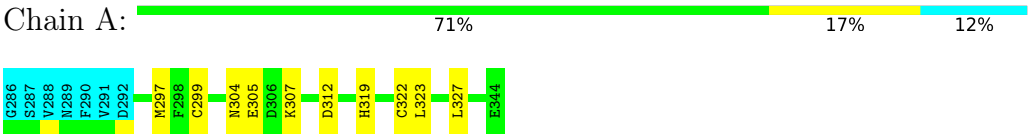
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Lysine-specific demethylase 5A

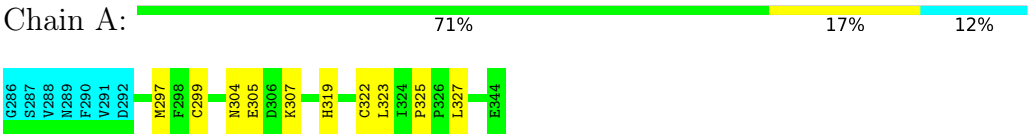
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Lysine-specific demethylase 5A

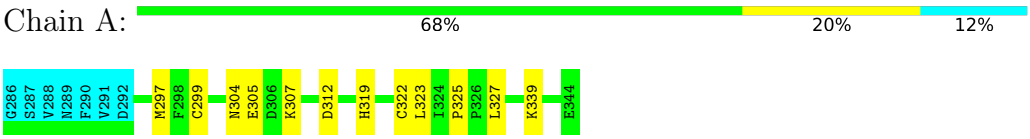
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Lysine-specific demethylase 5A

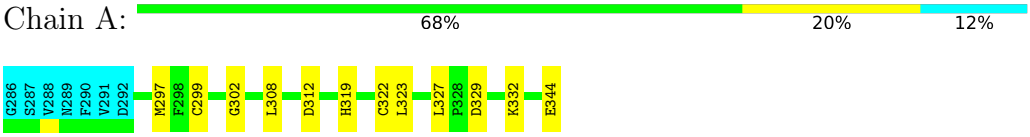
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Lysine-specific demethylase 5A

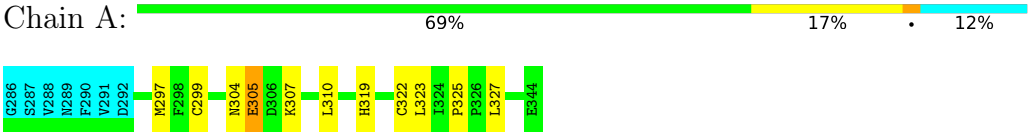
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Lysine-specific demethylase 5A

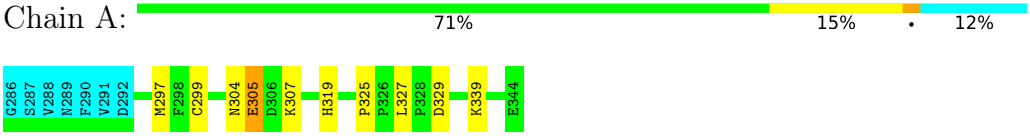
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Lysine-specific demethylase 5A

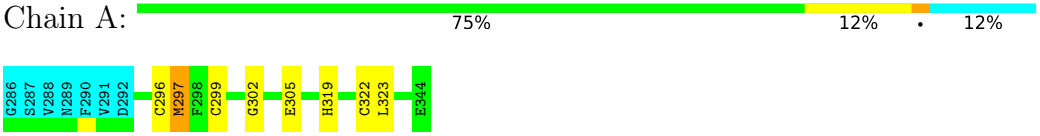
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Lysine-specific demethylase 5A

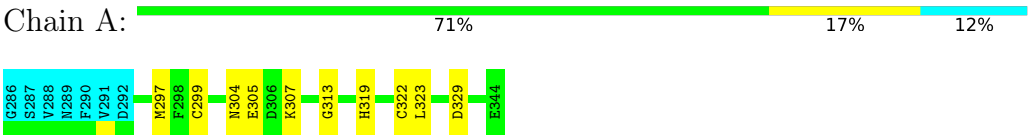
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Lysine-specific demethylase 5A

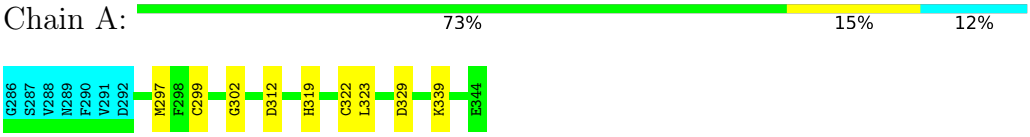
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Lysine-specific demethylase 5A

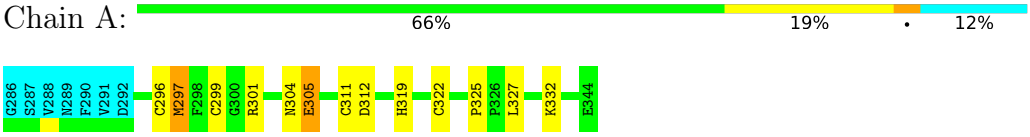
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Lysine-specific demethylase 5A

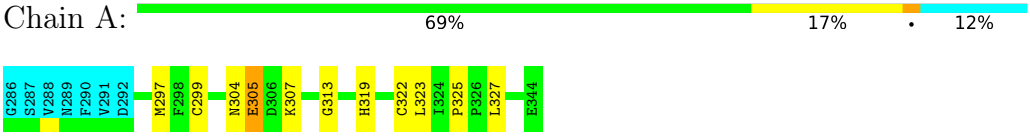
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Lysine-specific demethylase 5A

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 100 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
ARIA	refinement	2.3.2
ARIA	structure calculation	2.3.2

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	635
Number of shifts mapped to atoms	635
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	84%

6 Model quality

6.1 Standard geometry

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

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	407	377	377	2±1
All	All	8180	7540	7540	30

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:305:GLU:HA	1:A:319:HIS:HE2	0.47	1.70	20	3
1:A:319:HIS:HB2	1:A:322:CYS:SG	0.46	2.51	17	18
1:A:296:CYS:SG	1:A:297:MET:N	0.41	2.94	9	4
1:A:305:GLU:HA	1:A:319:HIS:NE2	0.41	2.31	19	3
1:A:304:ASN:O	1:A:306:ASP:N	0.41	2.54	7	1
1:A:305:GLU:N	1:A:319:HIS:HE2	0.41	2.14	2	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	51/59 (86%)	36±1 (72±2%)	10±1 (19±2%)	5±1 (10±2%)	1 10
All	All	1020/1180 (86%)	730 (72%)	191 (19%)	99 (10%)	1 10

All 11 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	297	MET	20
1	A	299	CYS	20
1	A	305	GLU	16
1	A	304	ASN	15
1	A	312	ASP	11
1	A	329	ASP	8
1	A	313	GLY	3
1	A	302	GLY	3
1	A	308	LEU	1
1	A	301	ARG	1
1	A	311	CYS	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	47/53 (89%)	43±1 (92±3%)	4±1 (8±3%)	16 63
All	All	940/1060 (89%)	866 (92%)	74 (8%)	16 63

All 11 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	323	LEU	18
1	A	307	LYS	14
1	A	327	LEU	13
1	A	325	PRO	11
1	A	332	LYS	4
1	A	344	GLU	3
1	A	305	GLU	3
1	A	339	LYS	3
1	A	293	LEU	2
1	A	304	ASN	2
1	A	310	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 monosaccharides in this entry.

6.6 Ligand geometry

Of 2 ligands modelled in this entry, 2 are monoatomic - 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 84% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list*

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	635
Number of shifts mapped to atoms	635
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	58	-0.37 ± 0.34	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	54	0.01 ± 0.33	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	52	-0.15 ± 0.61	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 84%, i.e. 517 atoms were assigned a chemical shift out of a possible 619. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	196/250 (78%)	97/99 (98%)	52/104 (50%)	47/47 (100%)
Sidechain	272/316 (86%)	169/188 (90%)	103/117 (88%)	0/11 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	49/53 (92%)	25/28 (89%)	23/23 (100%)	1/2 (50%)
Overall	517/619 (84%)	291/315 (92%)	178/244 (73%)	48/60 (80%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 83%, i.e. 572 atoms were assigned a chemical shift out of a possible 692. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	218/285 (76%)	108/113 (96%)	58/118 (49%)	52/54 (96%)
Sidechain	296/345 (86%)	183/204 (90%)	113/129 (88%)	0/12 (0%)
Aromatic	58/62 (94%)	30/33 (91%)	27/27 (100%)	1/2 (50%)
Overall	572/692 (83%)	321/350 (92%)	198/274 (72%)	53/68 (78%)

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	335	TRP	NE1	115.33	139.19 – 119.59	-7.2

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: