wwPDB NMR Structure Validation Summary Report

Jun 13, 2020 – 10:56 pm BST

PDB ID	:	6OCV
Title	:	Solution structure of the H-NOX protein from Shewanella woodyi in the Fe(II)CO ligation state
Authors	:	Chen, C.Y.; Lee, W.; Montfort, W.R.
Deposited on	:	2019-03-25

This is a wwPDB NMR Structure Validation Summary 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 (1)) were used in the production of this report:

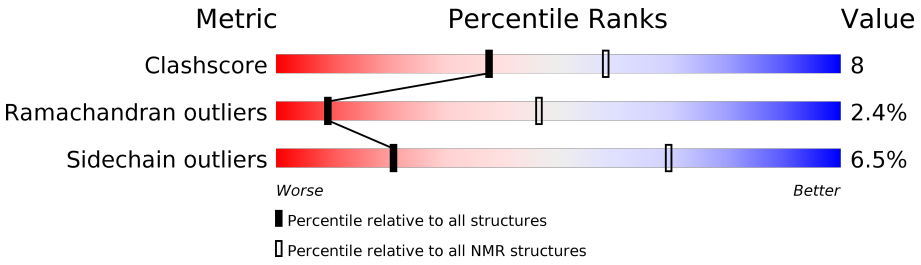
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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	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	190	79% 14% • 6%

2 Ensemble composition and analysis

This entry contains 20 models. Model 19 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:1-A:179 (179)	0.19	19

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 6 clusters and 1 single-model cluster was found.

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

3 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3091 atoms, of which 1516 are hydrogens and 0 are deuteriums.

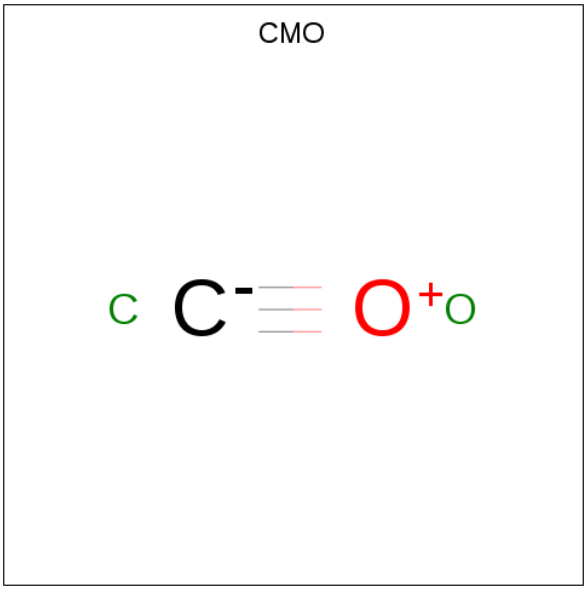
- Molecule 1 is a protein called Heme NO binding domain protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	190	3016	978	1486	239	305	8	0

There are 8 discrepancies between the modelled and reference sequences:

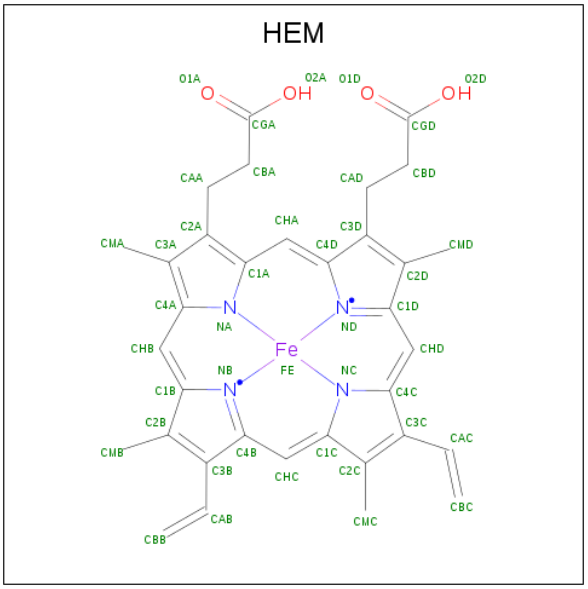
Chain	Residue	Modelled	Actual	Comment	Reference
A	183	GLY	-	expression tag	UNP B1KIH6
A	184	ALA	-	expression tag	UNP B1KIH6
A	185	GLU	-	expression tag	UNP B1KIH6
A	186	ASN	-	expression tag	UNP B1KIH6
A	187	LEU	-	expression tag	UNP B1KIH6
A	188	TYR	-	expression tag	UNP B1KIH6
A	189	PHE	-	expression tag	UNP B1KIH6
A	190	GLN	-	expression tag	UNP B1KIH6

- Molecule 2 is CARBON MONOXIDE (three-letter code: CMO) (formula: CO).



Mol	Chain	Residues	Atoms		
			Total	C	O
2	A	1	2	1	1

- Molecule 3 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula: $C_{34}H_{32}FeN_4O_4$).



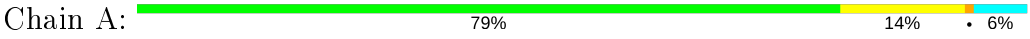
Mol	Chain	Residues	Atoms					
			Total	C	Fe	H	N	O
3	A	1	73	34	1	30	4	4

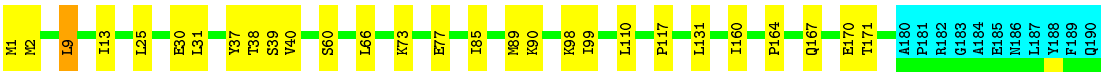
4 Residue-property plots (i)

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: Heme NO binding domain protein

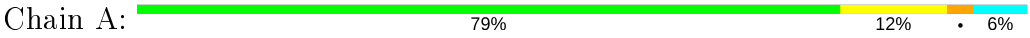
Chain A: 

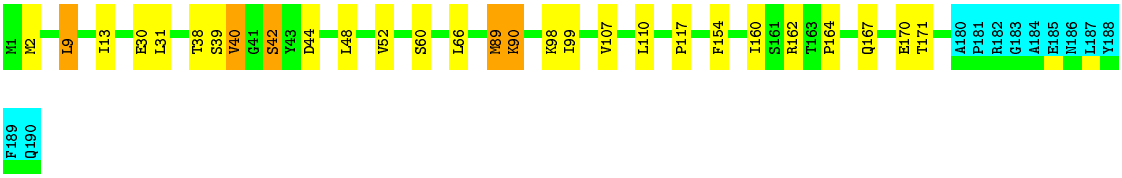


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 19. Colouring as in section 4.1 above.

- Molecule 1: Heme NO binding domain protein

Chain A: 



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
X-PLOR NIH	refinement	
X-PLOR NIH	structure calculation	

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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	2267
Number of shifts mapped to atoms	2267
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

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

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	1440	1404	1404	20±6
3	A	43	30	30	11±4
All	All	29700	28680	28680	475

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 8.

5 of 173 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:99:ILE:HD11	3:A:202:HEM:C3B	1.28	1.64	11	13
1:A:99:ILE:HD11	3:A:202:HEM:C2B	1.17	1.74	19	10
3:A:202:HEM:HHA	3:A:202:HEM:HBD2	1.08	1.23	8	6
1:A:99:ILE:HD11	3:A:202:HEM:C4B	1.08	1.83	12	3
3:A:202:HEM:HBA2	3:A:202:HEM:HMA2	0.99	1.26	16	2

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	178/190 (94%)	165±2 (93±1%)	9±1 (5±1%)	4±1 (2±1%)	9 46
All	All	3560/3800 (94%)	3298 (93%)	175 (5%)	87 (2%)	9 46

5 of 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	117	PRO	19
1	A	30	GLU	15
1	A	110	LEU	14
1	A	39	SER	12
1	A	40	VAL	12

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	164/172 (95%)	153±2 (94±1%)	11±2 (6±1%)	21 69
All	All	3280/3440 (95%)	3068 (94%)	212 (6%)	21 69

5 of 35 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	9	LEU	20
1	A	31	LEU	20
1	A	2	MET	20
1	A	66	LEU	20
1	A	90	LYS	16

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

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
3	HEM	A	202	1	27,50,50	2.09±0.02	0±0 (0±0%)
2	CMO	A	201	-	0,1,1	0.00±0.00	-

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
3	HEM	A	202	1	17,82,82	1.10±0.02	0±0 (0±0%)
2	CMO	A	201	-	-	-	-

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	HEM	A	202	1	-	0±0,6,54,54	-

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

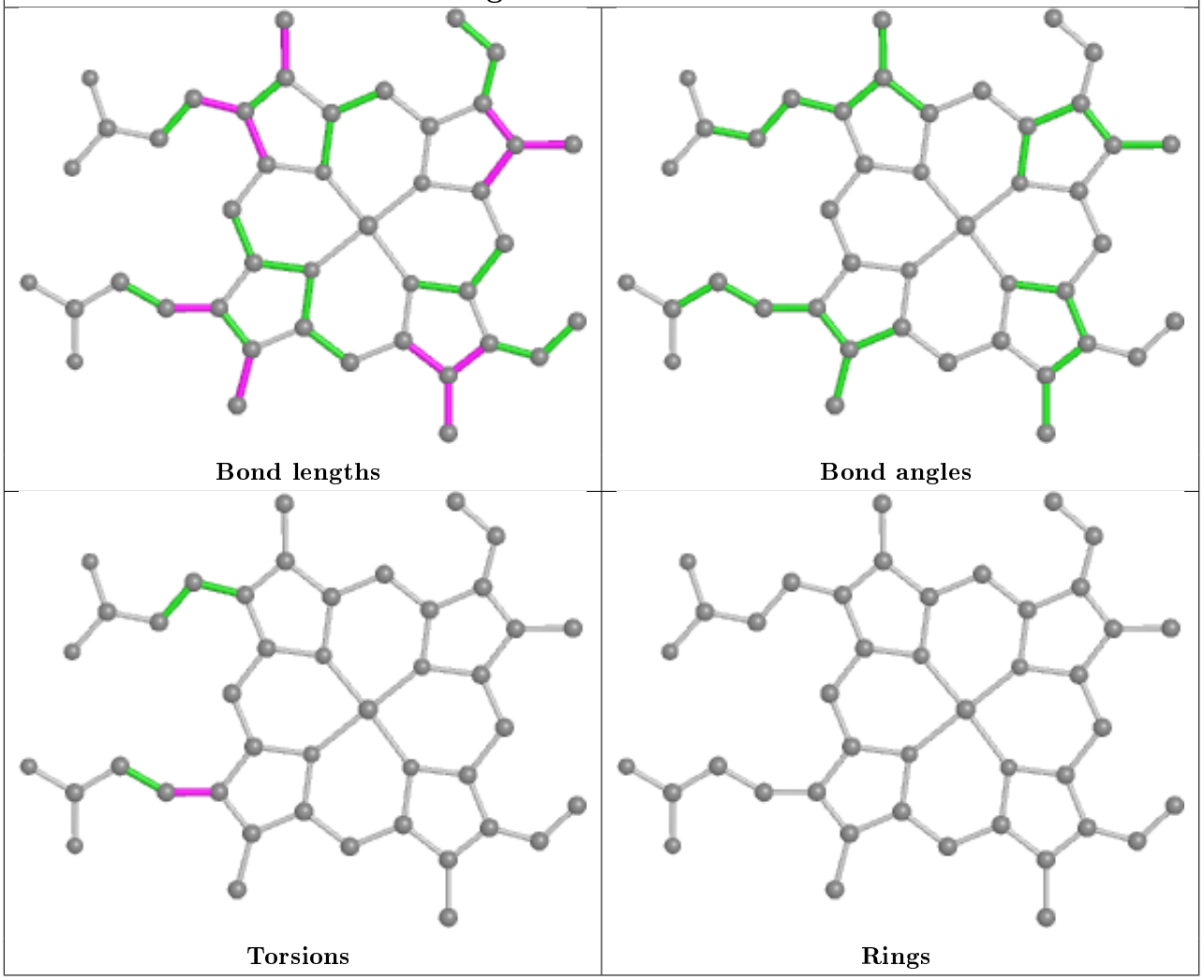
All unique torsion outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
3	A	202	HEM	C1A-C2A-CAA-CBA	1
3	A	202	HEM	C3A-C2A-CAA-CBA	1

There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand HEM A 202



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

The completeness of assignment taking into account all chemical shift lists is 87% for the well-defined parts and 86% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *SwHNOX_Native_chemical_shift*

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	2267
Number of shifts mapped to atoms	2267
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	28

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	187	-0.27 ± 0.05	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	174	0.23 ± 0.08	None needed (< 0.5 ppm)
$^{13}C'$	185	0.04 ± 0.10	None needed (< 0.5 ppm)
^{15}N	179	0.21 ± 0.33	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 87%, i.e. 1930 atoms were assigned a chemical shift out of a possible 2208. 0 out of 34 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	872/883 (99%)	348/352 (99%)	354/358 (99%)	170/173 (98%)
Sidechain	939/1133 (83%)	592/660 (90%)	346/437 (79%)	1/36 (3%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	119/192 (62%)	92/104 (88%)	26/84 (31%)	1/4 (25%)
Overall	1930/2208 (87%)	1032/1116 (92%)	726/879 (83%)	172/213 (81%)

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	146	LEU	HD12	-3.10	2.16 – -0.64	-13.8
1	A	146	LEU	HD11	-3.10	2.16 – -0.64	-13.8
1	A	146	LEU	HD13	-3.10	2.16 – -0.64	-13.8
1	A	188	TYR	CE1	133.43	124.14 – 111.74	12.5
1	A	70	PHE	CD2	117.83	137.34 – 125.84	-12.0
1	A	188	TYR	CE2	133.43	124.68 – 111.18	11.5
1	A	70	PHE	CD1	117.83	137.63 – 125.43	-11.2
1	A	78	LEU	HD11	-2.38	2.16 – -0.64	-11.2
1	A	78	LEU	HD12	-2.38	2.16 – -0.64	-11.2
1	A	78	LEU	HD13	-2.38	2.16 – -0.64	-11.2
1	A	179	ARG	HD2	0.92	4.27 – 1.97	-9.6
1	A	146	LEU	HD21	-1.62	2.14 – -0.66	-8.4
1	A	146	LEU	HD22	-1.62	2.14 – -0.66	-8.4
1	A	146	LEU	HD23	-1.62	2.14 – -0.66	-8.4
1	A	132	HIS	CE1	117.97	149.70 – 125.30	-8.0
1	A	99	ILE	HG23	-1.32	2.13 – -0.57	-7.8
1	A	99	ILE	HG22	-1.32	2.13 – -0.57	-7.8
1	A	99	ILE	HG21	-1.32	2.13 – -0.57	-7.8
1	A	116	LEU	HD11	-1.18	2.16 – -0.64	-6.9
1	A	116	LEU	HD12	-1.18	2.16 – -0.64	-6.9
1	A	116	LEU	HD13	-1.18	2.16 – -0.64	-6.9
1	A	146	LEU	HG	-0.63	3.16 – -0.14	-6.5
1	A	121	CYS	H	4.08	11.75 – 5.05	-6.4
1	A	142	PHE	HE2	5.04	8.69 – 5.49	-6.4
1	A	142	PHE	HE1	5.04	8.69 – 5.49	-6.4
1	A	136	LYS	CE	47.10	46.00 – 37.80	6.3
1	A	74	LEU	HG	-0.42	3.16 – -0.14	-5.8
1	A	116	LEU	HB2	-0.17	3.32 – -0.08	-5.3

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: