Full wwPDB NMR Structure Validation Report (i)

Apr 21, 2024 – 12:02 AM EDT

PDB ID	:	2M89
BMRB ID	:	19235
Title	:	Solution structure of the Aha1 dimer from Colwellia psychrerythraea
Authors	:	Rossi, P.; Sgourakis, N.G.; Shi, L.; Liu, G.; Barbieri, C.M.; Lee, H.; Grant, T.D.; Luft, J.R.; Xiao, R.; Acton, T.B.; Montelione, G.T.; Snell, E.H.; Baker, D.; Lange, O.A.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2013-05-09

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

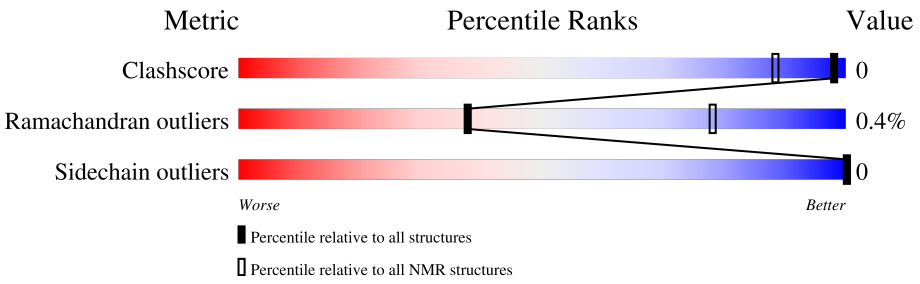
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR, SOLUTION SCATTERING

The overall completeness of chemical shifts assignment is 43%.

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	148	90% • 9%
1	B	148	90% • 9%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *structural features*.

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:134, B:1-B:134 (268)	0.95	6

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 and 5 single-model clusters were found.

Cluster number	Models
1	2, 6, 8
2	7, 9
Single-model clusters	1; 3; 4; 5; 10

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 4204 atoms, of which 2074 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Aha1 domain protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	134	2102	682	1037	183	194	6	0
1	B	134	2102	682	1037	183	194	6	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	147	LEU	-	expression tag	UNP Q484T9
A	148	GLU	-	expression tag	UNP Q484T9
B	147	LEU	-	expression tag	UNP Q484T9
B	148	GLU	-	expression tag	UNP Q484T9

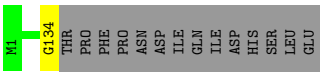
4 Residue-property plots (i)

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.

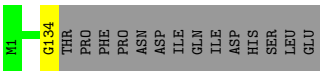
• Molecule 1: Aha1 domain protein

Chain A: 90% • 9%



• Molecule 1: Aha1 domain protein

Chain B: 90% • 9%



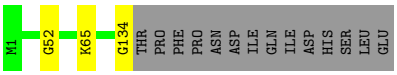
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

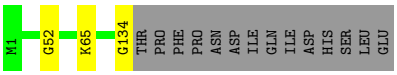
• Molecule 1: Aha1 domain protein

Chain A: 89% • 9%



• Molecule 1: Aha1 domain protein

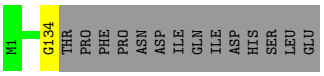
Chain B: 89% • 9%



4.2.2 Score per residue for model 2

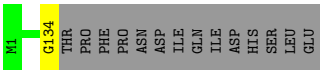
- Molecule 1: Aha1 domain protein

Chain A: 90% • 9%



- Molecule 1: Aha1 domain protein

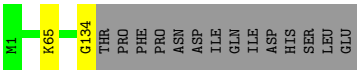
Chain B: 90% • 9%



4.2.3 Score per residue for model 3

- Molecule 1: Aha1 domain protein

Chain A: 89% • 9%



- Molecule 1: Aha1 domain protein

Chain B: 88% • 9%

4.2.4 Score per residue for model 4

- Molecule 1: Aha1 domain protein

Chain A: 87% • 9%

- Molecule 1: Aha1 domain protein

Chain B: 87% • 9%

4.2.5 Score per residue for model 5

• Molecule 1: Aha1 domain protein

Chain A: 89% • 9%

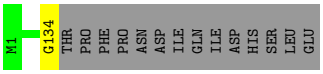
• Molecule 1: Aha1 domain protein

Chain B: 89% • 9%

4.2.6 Score per residue for model 6 (medoid)

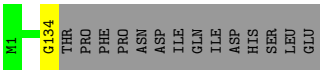
• Molecule 1: Aha1 domain protein

Chain A: 90% • 9%



• Molecule 1: Aha1 domain protein

Chain B: 90% • 9%



4.2.7 Score per residue for model 7

• Molecule 1: Aha1 domain protein

Chain A: 89% • 9%

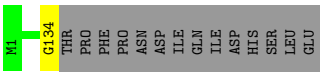
• Molecule 1: Aha1 domain protein

Chain B: 89% • 9%

4.2.8 Score per residue for model 8

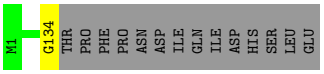
- Molecule 1: Aha1 domain protein

Chain A: 90% • 9%



- Molecule 1: Aha1 domain protein

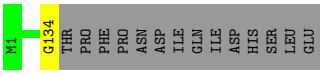
Chain B: 90% • 9%



4.2.9 Score per residue for model 9

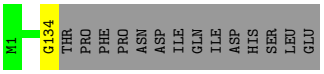
- Molecule 1: Aha1 domain protein

Chain A: 90% • 9%



- Molecule 1: Aha1 domain protein

Chain B: 90% • 9%



4.2.10 Score per residue for model 10

- Molecule 1: Aha1 domain protein

Chain A: 88% • 9%

- Molecule 1: Aha1 domain protein

Chain B: 87% • 9%

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics*.

Of the 40000 calculated structures, 10 were deposited, based on the following criterion: *target function*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
CS-ROSETTA	structure solution	
CS-ROSETTA	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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1669
Number of shifts mapped to atoms	1556
Number of unparsed shifts	0
Number of shifts with mapping errors	113
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	43%

6 Model quality (i)

6.1 Standard geometry (i)

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	0.74±0.00	1±0/1094 (0.1± 0.0%)	0.43±0.01	0±0/1482 (0.0± 0.0%)
1	B	0.74±0.00	1±0/1094 (0.1± 0.0%)	0.43±0.01	0±0/1482 (0.0± 0.0%)
All	All	0.74	20/21880 (0.1%)	0.43	0/29640 (0.0%)

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	B	134	GLY	C-O	-14.62	1.00	1.23	7	10
1	A	134	GLY	C-O	-14.59	1.00	1.23	9	10

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	1065	1037	1042	0±1
1	B	1065	1037	1042	1±1
All	All	21300	20740	20840	12

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:B:111:MET:HA	1:B:111:MET:HE2	0.56	1.78	4	1
1:A:111:MET:HA	1:A:111:MET:HE2	0.53	1.80	4	1
1:A:102:SER:O	1:A:103:ASN:HB2	0.46	2.10	7	4
1:B:102:SER:O	1:B:103:ASN:HB2	0.46	2.10	7	5
1:B:105:HIS:CD2	1:B:105:HIS:C	0.41	2.94	10	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	132/148 (89%)	124±2 (94±2%)	8±2 (6±2%)	0±1 (0±1%)	38	78
1	B	132/148 (89%)	124±3 (94±2%)	8±2 (6±2%)	0±1 (0±1%)	38	78
All	All	2640/2960 (89%)	2477 (94%)	153 (6%)	10 (0%)	38	78

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	65	LYS	4
1	B	65	LYS	4
1	A	52	GLY	1
1	B	52	GLY	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	115/129 (89%)	115±0 (100±0%)	0±0 (0±0%)	100	100
1	B	115/129 (89%)	115±0 (100±0%)	0±0 (0±0%)	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	2300/2580 (89%)	2300 (100%)	0 (0%)	100 100

There are no protein residues with a non-rotameric sidechain to report.

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

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 (i)

The completeness of assignment taking into account all chemical shift lists is 43% for the well-defined parts and 43% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_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	1669
Number of shifts mapped to atoms	1556
Number of unparsed shifts	0
Number of shifts with mapping errors	113
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 113 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	135	THR	H	9.534	.	1
1	A	135	THR	HA	4.562	.	1
1	A	135	THR	HB	4.66	.	1
1	A	135	THR	HG21	1.243	.	1
1	A	135	THR	HG22	1.243	.	1
1	A	135	THR	HG23	1.243	.	1
1	A	135	THR	C	172.831	.	1
1	A	135	THR	CA	58.293	.	1
1	A	135	THR	CB	68.347	.	1
1	A	135	THR	CG2	21.854	.	1
1	A	135	THR	N	111.716	.	1
1	A	136	PRO	HA	4.652	.	1
1	A	136	PRO	HB2	1.619	.	2
1	A	136	PRO	HB3	1.454	.	2

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	136	PRO	HG2	2.056	.	2
1	A	136	PRO	HG3	1.856	.	2
1	A	136	PRO	HD2	3.504	.	2
1	A	136	PRO	HD3	3.388	.	2
1	A	136	PRO	C	178.296	.	1
1	A	136	PRO	CA	62.636	.	1
1	A	136	PRO	CB	30.611	.	1
1	A	136	PRO	CG	26.448	.	1
1	A	136	PRO	CD	49.816	.	1
1	A	137	PHE	H	8.889	.	1
1	A	137	PHE	HA	5.078	.	1
1	A	137	PHE	HB2	3.42	.	2
1	A	137	PHE	HB3	3.361	.	2
1	A	137	PHE	HD2	6.59	.	3
1	A	137	PHE	CA	57.285	.	1
1	A	137	PHE	CB	39.289	.	1
1	A	137	PHE	N	123.486	.	1
1	A	138	PRO	HA	5.06	.	1
1	A	138	PRO	HB2	2.05	.	2
1	A	138	PRO	HB3	2.05	.	2
1	A	138	PRO	HD2	3.421	.	2
1	A	138	PRO	HD3	3.296	.	2
1	A	138	PRO	C	176.731	.	1
1	A	138	PRO	CA	64.115	.	1
1	A	138	PRO	CB	31.502	.	1
1	A	138	PRO	CD	48.947	.	1
1	A	139	ASN	H	9.581	.	1
1	A	139	ASN	HA	4.973	.	1
1	A	139	ASN	HB2	3.08	.	2
1	A	139	ASN	HB3	2.548	.	2
1	A	139	ASN	C	173.619	.	1
1	A	139	ASN	CA	53.278	.	1
1	A	139	ASN	CB	39.019	.	1
1	A	139	ASN	N	128.242	.	1
1	A	141	ILE	HG12	1.725	.	2
1	A	141	ILE	HG13	1.267	.	2
1	A	141	ILE	HG21	0.83	.	1
1	A	141	ILE	HG22	0.83	.	1
1	A	141	ILE	HG23	0.83	.	1
1	A	141	ILE	HD11	0.953	.	1
1	A	141	ILE	HD12	0.953	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	141	ILE	HD13	0.953	.	1
1	A	141	ILE	CG1	27.797	.	1
1	A	141	ILE	CG2	19.282	.	1
1	A	141	ILE	CD1	12.603	.	1
1	A	142	GLN	HG2	2.32	.	2
1	A	142	GLN	HG3	2.32	.	2
1	A	142	GLN	CG	33.72	.	1
1	A	143	ILE	HA	3.845	.	1
1	A	143	ILE	HB	1.716	.	1
1	A	143	ILE	HG21	0.831	.	1
1	A	143	ILE	HG22	0.831	.	1
1	A	143	ILE	HG23	0.831	.	1
1	A	143	ILE	HD11	0.967	.	1
1	A	143	ILE	HD12	0.967	.	1
1	A	143	ILE	HD13	0.967	.	1
1	A	143	ILE	C	175.729	.	1
1	A	143	ILE	CA	61.632	.	1
1	A	143	ILE	CB	39.569	.	1
1	A	143	ILE	CG2	17.083	.	1
1	A	143	ILE	CD1	12.373	.	1
1	A	144	ASP	H	8.057	.	1
1	A	144	ASP	HB2	2.74	.	2
1	A	144	ASP	HB3	2.7	.	2
1	A	144	ASP	C	173.987	.	1
1	A	144	ASP	CA	52.773	.	1
1	A	144	ASP	CB	41.66	.	1
1	A	144	ASP	N	125.257	.	1
1	A	146	SER	HA	4.51	.	1
1	A	146	SER	C	175.612	.	1
1	A	146	SER	CA	57.344	.	1
1	A	146	SER	CB	66.626	.	1
1	A	147	LEU	H	8.161	.	1
1	A	147	LEU	HA	4.442	.	1
1	A	147	LEU	HB2	1.63	.	2
1	A	147	LEU	HB3	1.57	.	2
1	A	147	LEU	HG	1.54	.	1
1	A	147	LEU	HD11	0.77	.	1
1	A	147	LEU	HD12	0.77	.	1
1	A	147	LEU	HD13	0.77	.	1
1	A	147	LEU	HD21	0.76	.	1
1	A	147	LEU	HD22	0.76	.	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	147	LEU	HD23	0.76	.	1
1	A	147	LEU	C	177.403	.	1
1	A	147	LEU	CA	54.365	.	1
1	A	147	LEU	CB	36.852	.	1
1	A	147	LEU	CG	26.77	.	1
1	A	147	LEU	CD2	24.1	.	1
1	A	147	LEU	N	110.805	.	1
1	A	148	GLU	H	8.554	.	1
1	A	148	GLU	HA	4.245	.	1
1	A	148	GLU	HB2	2.04	.	2
1	A	148	GLU	HB3	2.04	.	2
1	A	148	GLU	HG2	2.31	.	2
1	A	148	GLU	HG3	2.31	.	2
1	A	148	GLU	CA	58.466	.	1
1	A	148	GLU	CB	27.571	.	1
1	A	148	GLU	CG	36.01	.	1
1	A	148	GLU	N	119.472	.	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	138	-0.32 ± 0.22	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	128	0.24 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	138	-1.13 ± 0.12	Should be applied
^{15}N	131	0.48 ± 0.33	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 43%, i.e. 1556 atoms were assigned a chemical shift out of a possible 3616. 0 out of 32 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	647/1348 (48%)	264/552 (48%)	258/536 (48%)	125/260 (48%)
Sidechain	799/1886 (42%)	543/1226 (44%)	256/594 (43%)	0/66 (0%)
Aromatic	110/382 (29%)	57/200 (28%)	48/170 (28%)	5/12 (42%)
Overall	1556/3616 (43%)	864/1978 (44%)	562/1300 (43%)	130/338 (38%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 43%, i.e. 1556 atoms were assigned a chemical shift out of a possible 3616. 0 out of 32 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

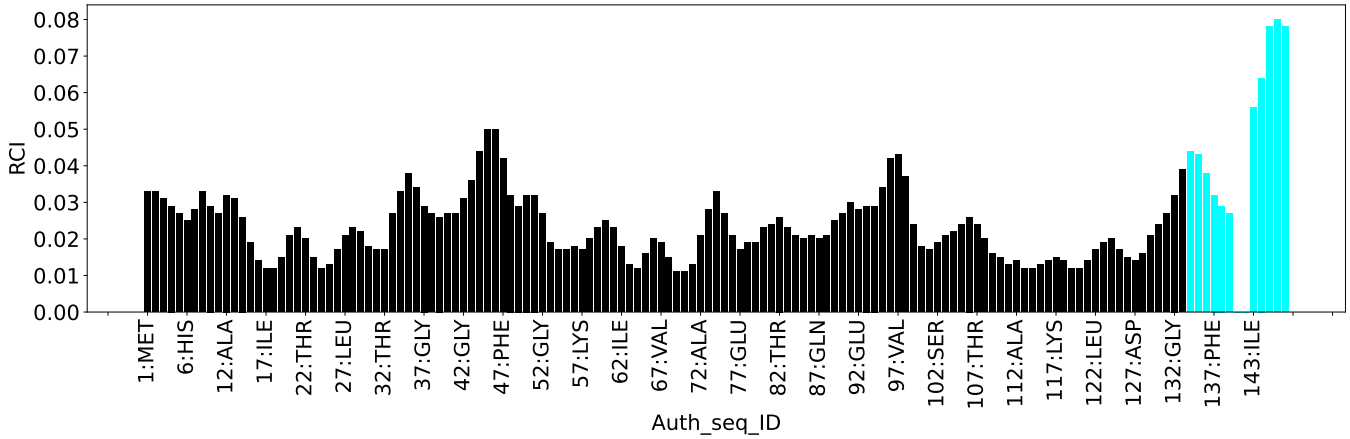
	Total	^{1}H	^{13}C	^{15}N
Backbone	647/1348 (48%)	264/552 (48%)	258/536 (48%)	125/260 (48%)
Sidechain	799/1886 (42%)	543/1226 (44%)	256/594 (43%)	0/66 (0%)
Aromatic	110/382 (29%)	57/200 (28%)	48/170 (28%)	5/12 (42%)
Overall	1556/3616 (43%)	864/1978 (44%)	562/1300 (43%)	130/338 (38%)

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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	64	ASN	H	4.55	5.28 – 11.36	-6.2
1	A	31	TRP	HB3	1.04	1.31 – 4.93	-5.8
1	A	111	MET	HG3	0.45	0.54 – 4.26	-5.2

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: