Full wwPDB NMR Structure Validation Report (i)

Nov 9, 2024 – 09:52 AM EST

PDB ID	:	5IPO
BMRB ID	:	30033
Title	:	Solution Structure of Hge36: Scorpine-like Peptide from Hadrurus Gertschi
Authors	:	Flores-Solis, D.; Rodriguez De La Vega, R.; del Rio-Portilla, F.
Deposited on	:	2016-03-09

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:

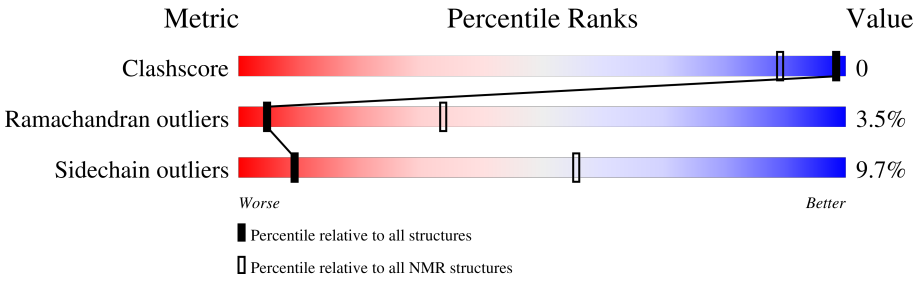
MolProbity	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 83%.

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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	48	77% 8% 15%

2 Ensemble composition and analysis

This entry contains 20 models. Model 6 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:5-A:45 (41)	0.61	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 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 728 atoms, of which 362 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hge-scorpine.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	48	728	226	362	68	65	7	0

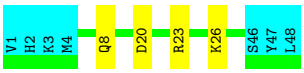
4 Residue-property plots

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: Hge-scorpine

Chain A:



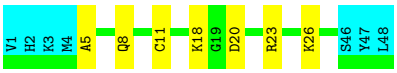
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Hge-scorpine

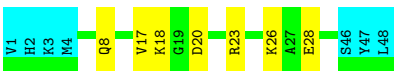
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Hge-scorpine

Chain A:



4.2.3 Score per residue for model 3

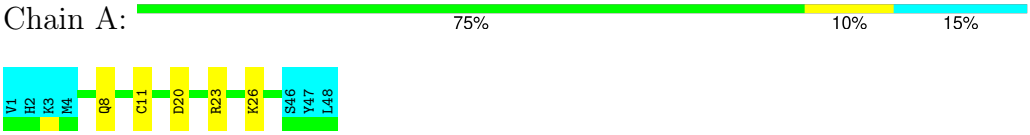
• Molecule 1: Hge-scorpine

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: Hge-scorpine

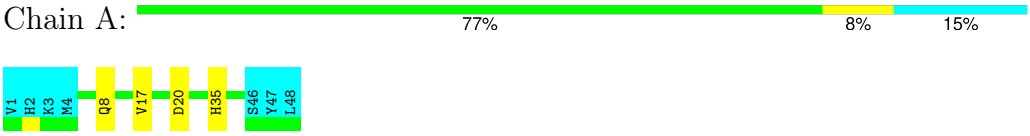
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Hge-scorpine

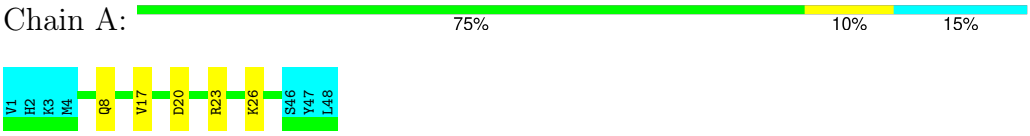
Chain A:



4.2.6 Score per residue for model 6 (medoid)

• Molecule 1: Hge-scorpine

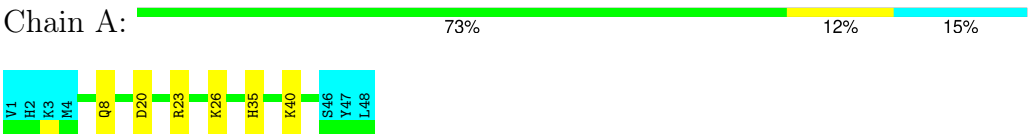
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Hge-scorpine

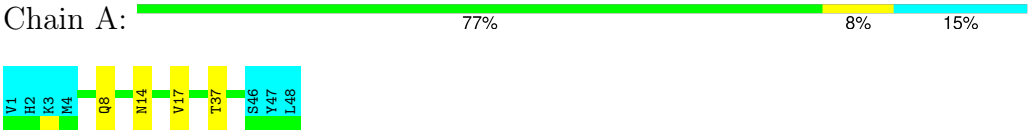
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Hge-scorpine

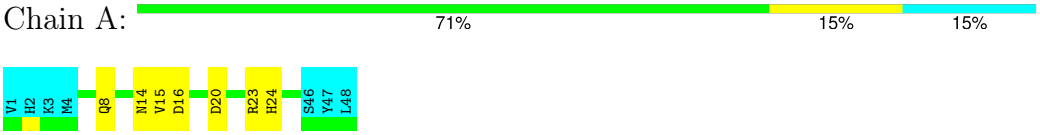
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Hge-scorpine

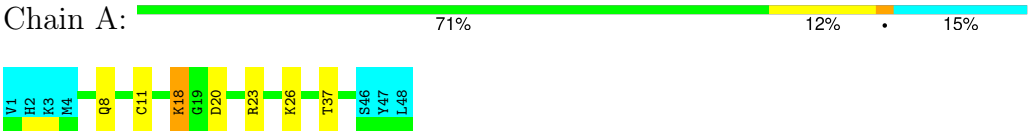
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Hge-scorpine

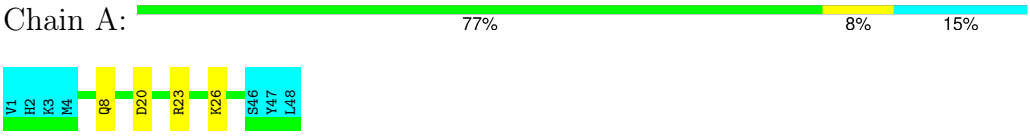
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Hge-scorpine

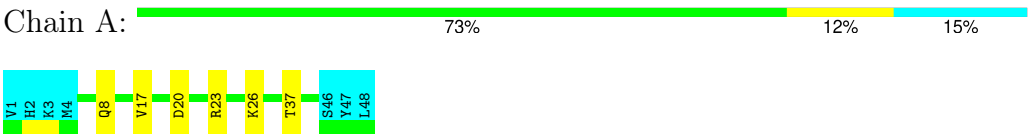
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Hge-scorpine

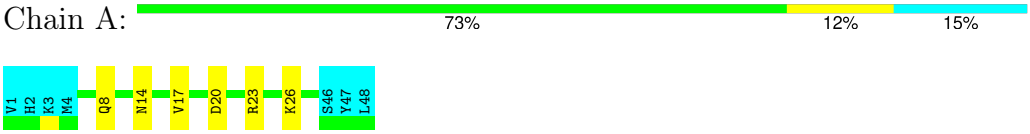
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Hge-scorpine

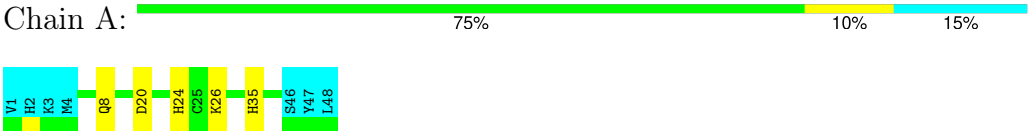
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Hge-scorpine

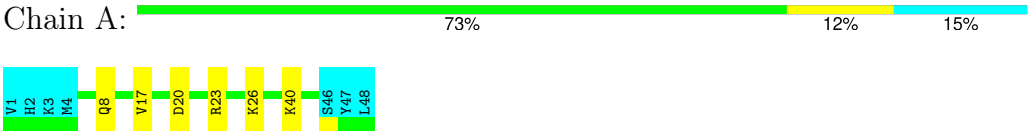
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Hge-scorpine

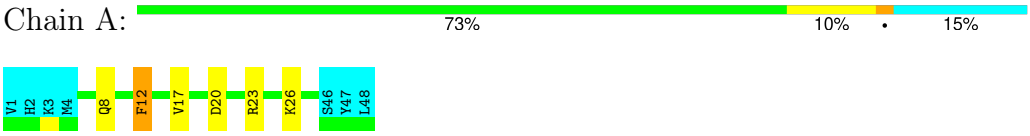
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Hge-scorpine

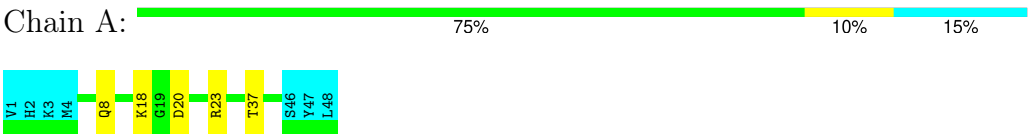
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Hge-scorpine

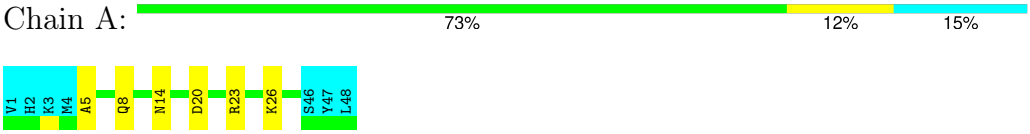
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Hge-scorpine

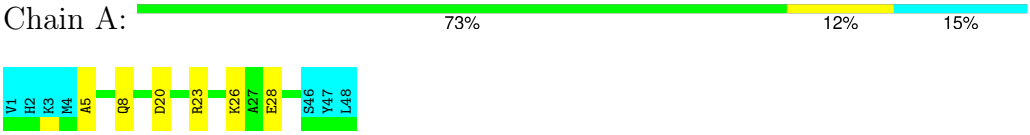
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Hge-scorpine

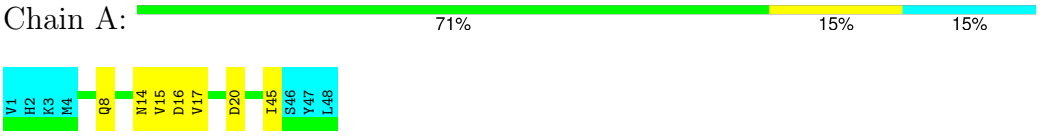
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Hge-scorpine

Chain A:



5 Refinement protocol and experimental data overview

Of the ? calculated structures, 20 were deposited, based on the following criterion: *?*.

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	520
Number of shifts mapped to atoms	519
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	83%

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	0.62±0.01	0±0/310 (0.0± 0.0%)	1.14±0.04	1±1/412 (0.3± 0.1%)
All	All	0.62	0/6200 (0.0%)	1.14	24/8240 (0.3%)

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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	23	ARG	NE-CZ-NH1	10.36	125.48	120.30	16	15
1	A	17	VAL	CA-CB-CG2	6.98	121.37	110.90	15	7
1	A	23	ARG	NE-CZ-NH2	5.85	123.22	120.30	17	1
1	A	12	PHE	CB-CG-CD1	-5.64	116.85	120.80	16	1

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	305	297	297	0±0
All	All	6100	5940	5940	2

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:15:VAL:HG12	1:A:16:ASP:H	0.48	1.69	9	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	41/48 (85%)	35±2 (85±4%)	5±2 (11±5%)	1±1 (4±2%)	5	34
All	All	820/960 (85%)	699 (85%)	92 (11%)	29 (4%)	5	34

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	20	ASP	19
1	A	5	ALA	4
1	A	18	LYS	4
1	A	35	HIS	2

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	33/40 (82%)	30±1 (90±3%)	3±1 (10±3%)	9	55
All	All	660/800 (82%)	596 (90%)	64 (10%)	9	55

All 15 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	8	GLN	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	26	LYS	15
1	A	37	THR	5
1	A	14	ASN	5
1	A	11	CYS	3
1	A	17	VAL	3
1	A	28	GLU	2
1	A	18	LYS	2
1	A	40	LYS	2
1	A	24	HIS	2
1	A	7	ASN	1
1	A	22	LYS	1
1	A	35	HIS	1
1	A	12	PHE	1
1	A	45	ILE	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 oligosaccharides 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

The completeness of assignment taking into account all chemical shift lists is 83% for the well-defined parts and 82% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *Starfile.str*

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	520
Number of shifts mapped to atoms	519
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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

- No matching atom found in the structure. All 1 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	35	HIS	HD1	8.659	0.020	.

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	48	-0.24 ± 0.18	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	43	0.75 ± 0.37	Should be checked
$^{13}C'$	48	-0.05 ± 0.25	None needed (< 0.5 ppm)
^{15}N	47	-1.36 ± 0.58	Should be applied

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 83%, i.e. 428 atoms were assigned a chemical shift out of a possible 514. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	208/208 (100%)	86/86 (100%)	82/82 (100%)	40/40 (100%)
Sidechain	208/272 (76%)	169/173 (98%)	36/86 (42%)	3/13 (23%)
Aromatic	12/34 (35%)	12/18 (67%)	0/14 (0%)	0/2 (0%)
Overall	428/514 (83%)	267/277 (96%)	118/182 (65%)	43/55 (78%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 82%, i.e. 508 atoms were assigned a chemical shift out of a possible 620. 0 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	242/243 (100%)	99/100 (99%)	96/96 (100%)	47/47 (100%)
Sidechain	249/327 (76%)	203/210 (97%)	43/103 (42%)	3/14 (21%)
Aromatic	17/50 (34%)	17/26 (65%)	0/21 (0%)	0/3 (0%)
Overall	508/620 (82%)	319/336 (95%)	139/220 (63%)	50/64 (78%)

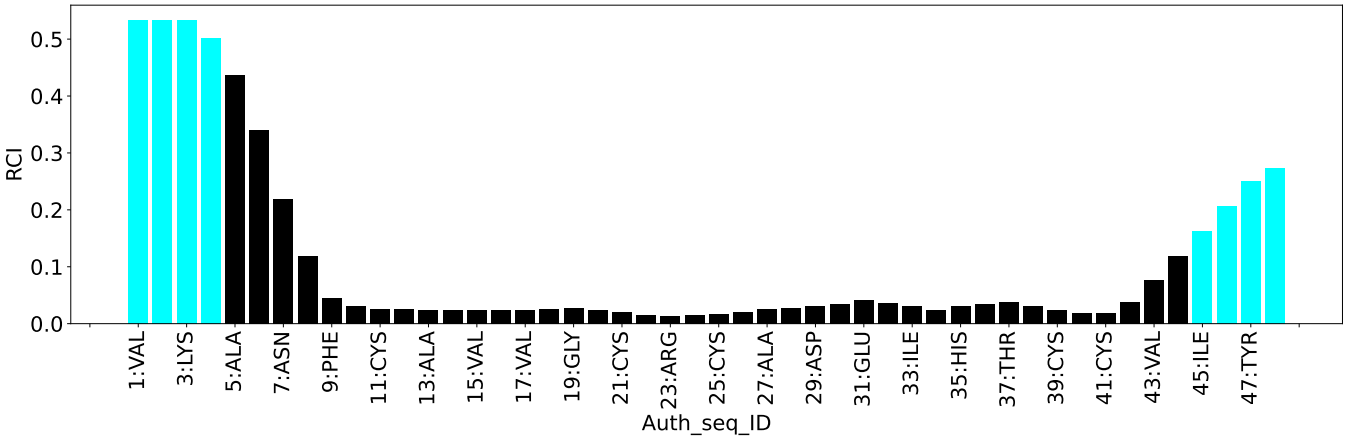
7.1.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

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.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:


RCI
0.0
0.1
0.2
0.3
0.4
0.5
1:VAL
3:LYS
5:ALA
7:ASN
9:PHE
11:CYS
13:ALA
15:VAL
17:VAL
19:GLY
21:CYS
23:ARG
25:CYS
27:ALA
29:ASP
31:GLU
33:ILE
35:HIS
37:THR
39:CYS
41:CYS
43:VAL
45:ILE
47:TYR
Auth_seq_ID