Full wwPDB NMR Structure Validation Report (i)

Nov 7, 2023 – 01:07 PM EST

PDB ID	:	6O7G
BMRB ID	:	30585
Title	:	Solution structure of MLL4 PHD6 domain in complex with histone H4K16ac peptide
Authors	:	Zhang, Y.; Kutateladze, T.G.
Deposited on	:	2019-03-07

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:

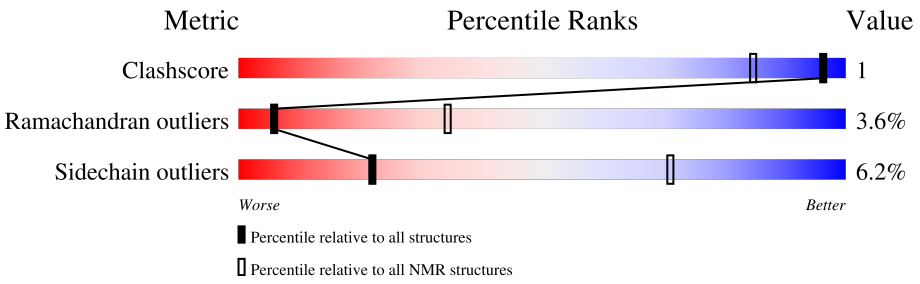
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 81%.

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	B	64	69% 8% • 16% 6%
2	A	13	15% 8% 38% 38%

2 Ensemble composition and analysis (i)

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	B:1504-B:1553, A:14-A:15, A:17-A:17 (53)	0.88	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, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15
2	4, 12

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 998 atoms, of which 473 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone-lysine N-methyltransferase 2D.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	B	60	897	293	425	77	94	8	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1499	GLY	-	expression tag	UNP O14686
B	1500	SER	-	expression tag	UNP O14686
B	1501	HIS	-	expression tag	UNP O14686
B	1502	MET	-	expression tag	UNP O14686

- Molecule 2 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	A	8	99	30	48	13	8	1

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
3	B	2	2	2

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

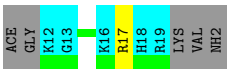
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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: Histone-lysine N-methyltransferase 2D

Chain B:

• Molecule 2: Histone H4

Chain A:



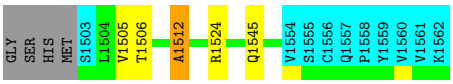
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

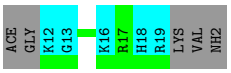
• Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B:



• Molecule 2: Histone H4

Chain A:



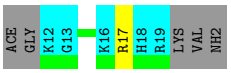
4.2.2 Score per residue for model 2

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 69% 8% • 16% 6%

- Molecule 2: Histone H4

Chain A: 15% 8% 38% 38%



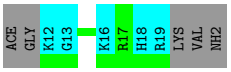
4.2.3 Score per residue for model 3

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 67% 9% • 16% 6%

- Molecule 2: Histone H4

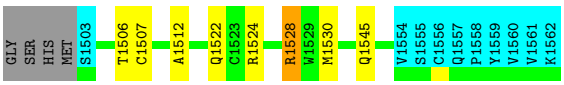
Chain A: 23% 38% 38%



4.2.4 Score per residue for model 4

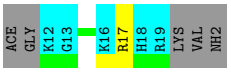
- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 66% 11% • 16% 6%



- Molecule 2: Histone H4

Chain A: 15% 8% 38% 38%



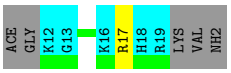
4.2.5 Score per residue for model 5

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 72% 5% • 16% 6%

- Molecule 2: Histone H4

Chain A: 15% 8% 38% 38%



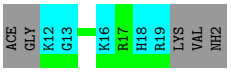
4.2.6 Score per residue for model 6

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 70% 6% • 16% 6%

- Molecule 2: Histone H4

Chain A: 23% 38% 38%



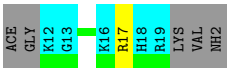
4.2.7 Score per residue for model 7

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 69% 8% • 16% 6%

- Molecule 2: Histone H4

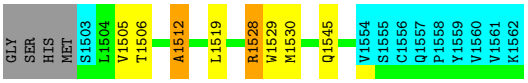
Chain A: 15% 8% 38% 38%



4.2.8 Score per residue for model 8

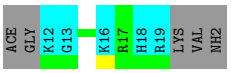
- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 66% 9% • 16% 6%



- Molecule 2: Histone H4

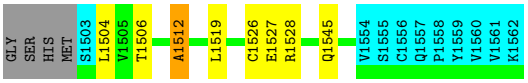
Chain A: 23% 38% 38%



4.2.9 Score per residue for model 9

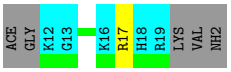
- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 66% 11% • 16% 6%



- Molecule 2: Histone H4

Chain A: 15% 8% 38% 38%



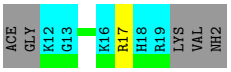
4.2.10 Score per residue for model 10

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 69% 8% • 16% 6%

- Molecule 2: Histone H4

Chain A: 15% 8% 38% 38%



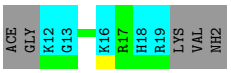
4.2.11 Score per residue for model 11

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 72% 6% 16% 6%

- Molecule 2: Histone H4

Chain A: 23% 38% 38%



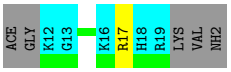
4.2.12 Score per residue for model 12

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 70% 8% 16% 6%

- Molecule 2: Histone H4

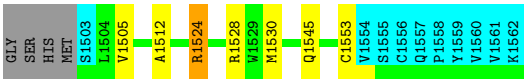
Chain A: 15% 8% 38% 38%



4.2.13 Score per residue for model 13

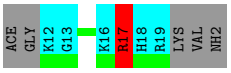
- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B: 67% 9% • 16% 6%



- Molecule 2: Histone H4

Chain A: 15% 8% 38% 38%



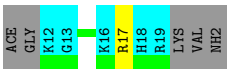
4.2.14 Score per residue for model 14

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B:

- Molecule 2: Histone H4

Chain A:



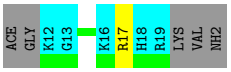
4.2.15 Score per residue for model 15

- Molecule 1: Histone-lysine N-methyltransferase 2D

Chain B:

- Molecule 2: Histone H4

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 100 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**
X-PLOR NIH	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	669
Number of shifts mapped to atoms	669
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	81%

6 Model quality (i)

6.1 Standard geometry (i)

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

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	B	0.74±0.01	0±0/406 (0.0± 0.0%)	1.11±0.03	1±1/553 (0.2± 0.1%)
2	A	0.89±0.04	0±0/19 (0.0± 0.0%)	1.64±0.23	1±0/23 (2.9± 2.0%)
All	All	0.75	0/6375 (0.0%)	1.14	27/8640 (0.3%)

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
2	A	0.0±0.0	0.1±0.2
All	All	0	1

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	B	1524	ARG	NE-CZ-NH1	9.05	124.82	120.30	1	5
2	A	17	ARG	NE-CZ-NH1	6.91	123.75	120.30	7	6
1	B	1524	ARG	NE-CZ-NH2	-6.46	117.07	120.30	1	4
1	B	1528	ARG	NE-CZ-NH2	5.70	123.15	120.30	4	3
2	A	17	ARG	NE-CZ-NH2	5.69	123.15	120.30	5	4
1	B	1506	THR	CA-CB-CG2	5.63	120.28	112.40	4	1
1	B	1528	ARG	NE-CZ-NH1	5.47	123.04	120.30	9	2
1	B	1552	ASP	CB-CG-OD1	5.07	122.86	118.30	15	2

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
2	A	17	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	B	396	347	347	1±0
2	A	20	21	21	0±0
All	All	6270	5520	5520	9

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 1.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:1506:THR:HG23	1:B:1512:ALA:HB2	0.71	1.62	9	9

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	B	50/64 (78%)	44±1 (89±3%)	4±2 (7±3%)	2±1 (4±1%)	5 32
2	A	3/13 (23%)	3±0 (93±13%)	0±0 (7±13%)	0±0 (0±0%)	100 100
All	All	795/1155 (69%)	708 (89%)	58 (7%)	29 (4%)	6 34

All 8 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	B	1512	ALA	15
1	B	1505	VAL	6
1	B	1504	LEU	2
1	B	1507	CYS	2
1	B	1515	VAL	1
1	B	1519	LEU	1
1	B	1536	SER	1
1	B	1553	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	B	44/57 (77%)	41±1 (94±3%)	3±1 (6±3%)	22	71
2	A	1/6 (17%)	1±0 (93±25%)	0±0 (7±25%)	20	68
All	All	675/945 (71%)	633 (94%)	42 (6%)	22	71

All 13 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	B	1545	GLN	14
1	B	1519	LEU	7
1	B	1530	MET	5
1	B	1528	ARG	4
1	B	1526	CYS	2
1	B	1537	LEU	2
1	B	1527	GLU	2
1	B	1534	CYS	1
1	B	1522	GLN	1
1	B	1524	ARG	1
2	A	17	ARG	1
1	B	1516	GLU	1
1	B	1552	ASP	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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
2	ALY	A	16	2	10,11,12	0.49±0.04	0±0 (0±0%)

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
2	ALY	A	16	2	7,12,14	0.92±0.19	0±0 (1±4%)

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
2	ALY	A	16	2	-	0±0,9,10,12	-

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	A	16	ALY	CE-NZ-CH	3.19	127.47	122.56	4	2

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 81% for the well-defined parts and 76% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_output*

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	669
Number of shifts mapped to atoms	669
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 (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	47	-0.41 ± 0.28	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	55	0.02 ± 0.33	None needed (< 0.5 ppm)
$^{13}C'$	41	0.06 ± 0.14	None needed (< 0.5 ppm)
^{15}N	54	-0.17 ± 0.90	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 81%, i.e. 553 atoms were assigned a chemical shift out of a possible 679. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	222/264 (84%)	101/107 (94%)	75/106 (71%)	46/51 (90%)
Sidechain	291/353 (82%)	196/228 (86%)	93/114 (82%)	2/11 (18%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	40/62 (65%)	23/32 (72%)	16/26 (62%)	1/4 (25%)
Overall	553/679 (81%)	320/367 (87%)	184/246 (75%)	49/66 (74%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 76%, i.e. 661 atoms were assigned a chemical shift out of a possible 872. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	262/333 (79%)	120/135 (89%)	88/134 (66%)	54/64 (84%)
Sidechain	353/461 (77%)	237/298 (80%)	113/146 (77%)	3/17 (18%)
Aromatic	46/78 (59%)	27/40 (68%)	18/33 (55%)	1/5 (20%)
Overall	661/872 (76%)	384/473 (81%)	219/313 (70%)	58/86 (67%)

7.1.4 Statistically unusual chemical shifts (i)

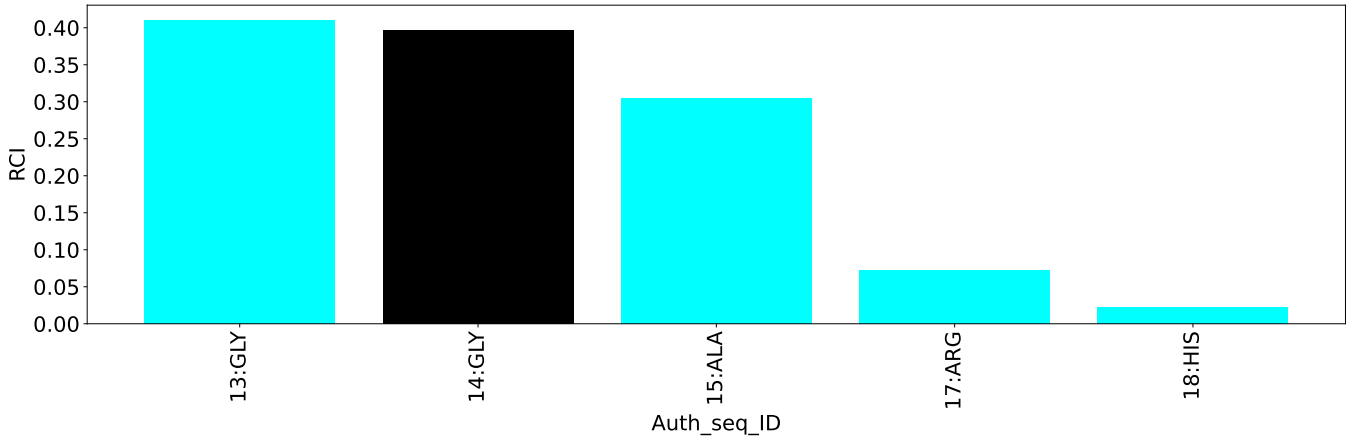
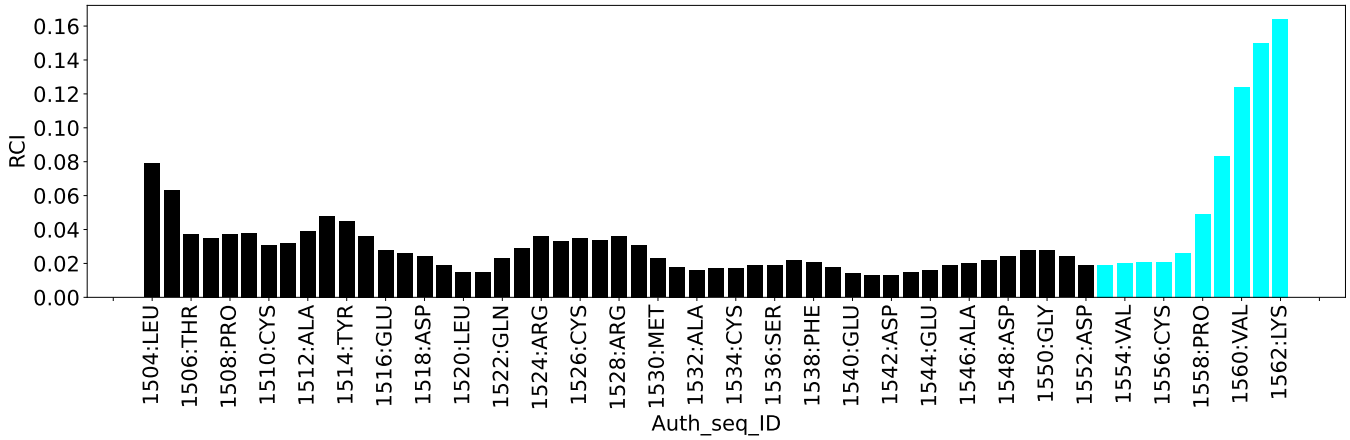
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	B	1522	GLN	HG2	0.50	1.01 – 3.62	-6.9

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:



Random coil index (RCI) for chain B: