Full wwPDB NMR Structure Validation Report (i)

Nov 7, 2023 – 02:53 PM EST

PDB ID	:	6E86
BMRB ID	:	30502
Title	:	Solution structure of ZZZ3 ZZ domain in complex with histone H3K4ac peptide
Authors	:	Zhang, Y.; Kutateladze, T.G.
Deposited on	:	2018-07-27

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:

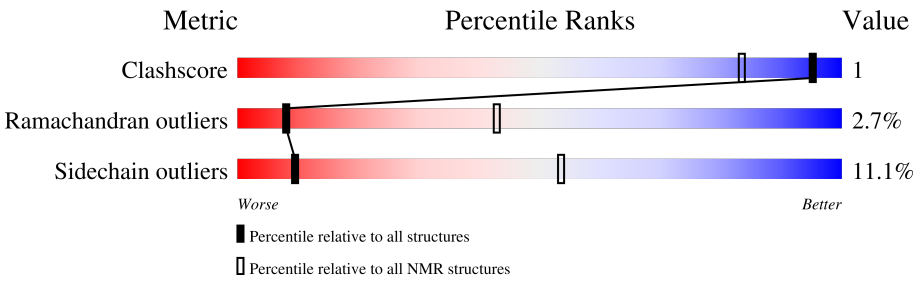
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
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 76%.

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	83% 8% • • 5%
2	A	8	25% 12% 62%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 7 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:814-B:874, A:1-A:3 (64)	0.40	7

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 4 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1101 atoms, of which 526 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ZZ-type zinc finger-containing protein 3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	B	64	957	309	452	88	100	8	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	811	GLY	-	expression tag	UNP Q8IYH5
B	812	PRO	-	expression tag	UNP Q8IYH5
B	813	LEU	-	expression tag	UNP Q8IYH5
B	814	GLY	-	expression tag	UNP Q8IYH5
B	815	SER	-	expression tag	UNP Q8IYH5

- Molecule 2 is a protein called H3K4ac.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	A	8	142	39	74	16	13	0

- 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

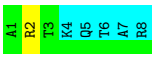
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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: ZZ-type zinc finger-containing protein 3

Chain B: 83% 8% • • 5%

• Molecule 2: H3K4ac

Chain A: 25% 12% 62%



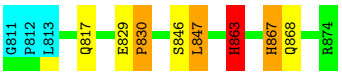
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

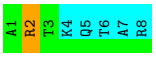
• Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 83% 6% 5% • 5%



• Molecule 2: H3K4ac

Chain A: 25% 12% 62%



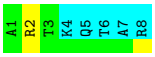
4.2.2 Score per residue for model 2

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B:

- Molecule 2: H3K4ac

Chain A:



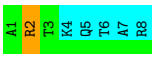
4.2.3 Score per residue for model 3

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B:

- Molecule 2: H3K4ac

Chain A:



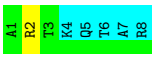
4.2.4 Score per residue for model 4

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B:

- Molecule 2: H3K4ac

Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 83% 9% • 5%

G811 P812 L813 Q817 P830 C838 L847 H858 H863 H867 Q868 R874

- Molecule 2: H3K4ac

Chain A: 25% 12% 62%

A1 R2 T3 K4 Q5 T6 A7 R8

4.2.6 Score per residue for model 6

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 81% 9% • • 5%

G811 P812 L813 Q817 K822 P830 R835 W836 S854 H863 H867 Q868 R874

- Molecule 2: H3K4ac

Chain A: 25% 12% 62%

A1 R2 T3 K4 Q5 T6 A7 R8

4.2.7 Score per residue for model 7 (medoid)

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 83% 9% • 5%

G811 P812 L813 Q817 K822 P830 L847 H863 H867 Q868 L869 R874

- Molecule 2: H3K4ac

Chain A: 25% 12% 62%

A1 R2 T3 K4 Q5 T6 A7 R8

4.2.8 Score per residue for model 8

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 81% 8% 6% 5%

G811 P812 L813 Q817 K822 P830 R835 C838 L847 H863 H867 Q868 L869 R874

- Molecule 2: H3K4ac

Chain A: 25% 12% 62%

A1 R2 T3 K4 Q5 T6 A7 R8

4.2.9 Score per residue for model 9

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 84% 9% • 5%

G811 P812 L813 Q817 K822 P830 H863 H867 Q868 L869 R874

- Molecule 2: H3K4ac

Chain A: 25% 12% 62%

A1 R2 T3 K4 Q5 T6 A7 R8

4.2.10 Score per residue for model 10

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 83% 6% 5% • 5%

G811 P812 L813 Q817 K822 C823 P830 L847 H863 H867 Q868 R874

- Molecule 2: H3K4ac

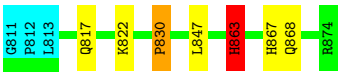
Chain A: 25% 12% 62%

A1 R2 T3 K4 Q5 T6 A7 R8

4.2.11 Score per residue for model 11

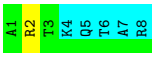
- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 84% 8% • • 5%



- Molecule 2: H3K4ac

Chain A: 25% 12% 62%



4.2.12 Score per residue for model 12

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 81% 11% • 5%

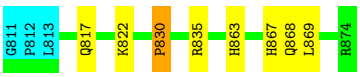
- Molecule 2: H3K4ac

Chain A: 25% 12% 62%

4.2.13 Score per residue for model 13

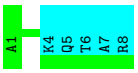
- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 83% 11% • 5%



- Molecule 2: H3K4ac

Chain A: 38% 62%



4.2.14 Score per residue for model 14

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 84% 9% • 5%

G811 P812 L813 Q817 K822 P830 H863 H867 Q868 L869 R874

- Molecule 2: H3K4ac

Chain A: 25% 12% 62%

A1 R2 T3 K4 Q5 T6 A7 R8

4.2.15 Score per residue for model 15

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 84% 6% • • 5%

G811 P812 L813 Q817 P830 R835 L847 H863 H867 Q868 R874

- Molecule 2: H3K4ac

Chain A: 25% 12% 62%

A1 R2 T3 K4 Q5 T6 A7 R8

4.2.16 Score per residue for model 16

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B: 78% 11% 5% • 5%

G811 P812 L813 Q817 H818 V819 P830 R835 L847 L857 H863 H867 Q868 L869 I872 Y873 R874

- Molecule 2: H3K4ac

Chain A: 25% 12% 62%

A1 R2 T3 K4 Q5 T6 A7 R8

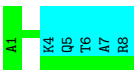
4.2.17 Score per residue for model 17

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B:

- Molecule 2: H3K4ac

Chain A:



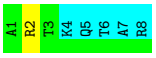
4.2.18 Score per residue for model 18

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B:

- Molecule 2: H3K4ac

Chain A:



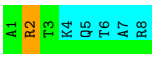
4.2.19 Score per residue for model 19

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B:

- Molecule 2: H3K4ac

Chain A:



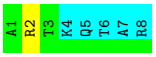
4.2.20 Score per residue for model 20

- Molecule 1: ZZ-type zinc finger-containing protein 3

Chain B:

- Molecule 2: H3K4ac

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, 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	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	666
Number of shifts mapped to atoms	666
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	76%

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	0.67±0.01	0±0/499 (0.0± 0.0%)	1.13±0.04	3±2/674 (0.4± 0.2%)
2	A	0.79±0.03	0±0/22 (0.0± 0.0%)	1.58±0.12	0±1/28 (1.6± 2.6%)
All	All	0.67	0/10420 (0.0%)	1.15	62/14040 (0.4%)

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
1	B	0.0±0.0	1.8±0.4
All	All	0	36

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	835	ARG	NE-CZ-NH1	7.16	123.88	120.30	18	11
1	B	835	ARG	NE-CZ-NH2	7.15	123.88	120.30	12	9
1	B	830	PRO	CA-N-CD	-6.39	102.55	111.50	17	10
1	B	835	ARG	CD-NE-CZ	5.66	131.53	123.60	15	4
1	B	863	HIS	CB-CG-CD2	-5.62	113.39	130.80	10	13
1	B	863	HIS	CA-CB-CG	5.57	123.06	113.60	6	3
2	A	2	ARG	NE-CZ-NH1	5.37	122.98	120.30	11	4
2	A	2	ARG	NE-CZ-NH2	5.33	122.97	120.30	19	3
2	A	2	ARG	CD-NE-CZ	5.28	131.00	123.60	11	2
1	B	838	CYS	CA-CB-SG	-5.28	104.49	114.00	8	3

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	B	863	HIS	Sidechain	20
1	B	867	HIS	Sidechain	16

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	486	429	429	1±1
All	All	10220	9120	9120	19

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:847:LEU:HD22	1:B:863:HIS:CD2	0.75	2.16	11	8
1:B:847:LEU:HD22	1:B:863:HIS:HD2	0.65	1.49	18	4
1:B:847:LEU:HD22	1:B:863:HIS:ND1	0.47	2.25	1	4
1:B:847:LEU:HD22	1:B:863:HIS:CE1	0.45	2.46	10	1
1:B:847:LEU:HD22	1:B:863:HIS:HD1	0.43	1.73	8	1
1:B:836:TRP:CD1	1:B:854:SER:HB3	0.42	2.49	6	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	B	60/64 (94%)	51±1 (85±2%)	7±1 (12±2%)	2±1 (3±1%)	8 44
2	A	2/8 (25%)	2±0 (90±20%)	0±0 (5±15%)	0±0 (5±15%)	4 25
All	All	1240/1440 (86%)	1061 (86%)	146 (12%)	33 (3%)	8 43

All 7 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	830	PRO	20
1	B	846	SER	3
1	B	829	GLU	3
1	B	857	LEU	3
2	A	2	ARG	2
1	B	858	HIS	1
1	B	819	VAL	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	57/59 (97%)	51±1 (90±2%)	6±1 (10±2%)	11 56
2	A	2/5 (40%)	1±0 (60±20%)	1±0 (40±20%)	0 5
All	All	1180/1280 (92%)	1049 (89%)	131 (11%)	9 53

All 14 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	817	GLN	20
1	B	867	HIS	18
2	A	2	ARG	16
1	B	868	GLN	14
1	B	863	HIS	13
1	B	822	LYS	13
1	B	830	PRO	10
1	B	869	LEU	10
1	B	847	LEU	9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	B	857	LEU	3
1	B	872	ILE	2
1	B	829	GLU	1
1	B	845	MET	1
1	B	823	CYS	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	4	2	10,11,12	0.59±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	4	2	7,12,14	0.91±0.16	0±0 (0±3%)

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	4	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	4	ALY	CE-NZ-CH	3.31	127.65	122.56	12	1

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 76% for the well-defined parts and 72% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *xeasy.str.5*

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	666
Number of shifts mapped to atoms	666
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	62	0.00 ± 0.50	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	58	-0.20 ± 0.26	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	57	-1.30 ± 0.48	Should be applied

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 76%, i.e. 631 atoms were assigned a chemical shift out of a possible 825. 0 out of 6 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	238/316 (75%)	122/128 (95%)	60/128 (47%)	56/60 (93%)
Sidechain	347/433 (80%)	235/277 (85%)	107/140 (76%)	5/16 (31%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	46/76 (61%)	27/40 (68%)	18/30 (60%)	1/6 (17%)
Overall	631/825 (76%)	384/445 (86%)	185/298 (62%)	62/82 (76%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 72%, i.e. 662 atoms were assigned a chemical shift out of a possible 919. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	245/350 (70%)	126/142 (89%)	62/142 (44%)	57/66 (86%)
Sidechain	371/493 (75%)	252/316 (80%)	114/157 (73%)	5/20 (25%)
Aromatic	46/76 (61%)	27/40 (68%)	18/30 (60%)	1/6 (17%)
Overall	662/919 (72%)	405/498 (81%)	194/329 (59%)	63/92 (68%)

7.1.4 Statistically unusual chemical shifts (i)

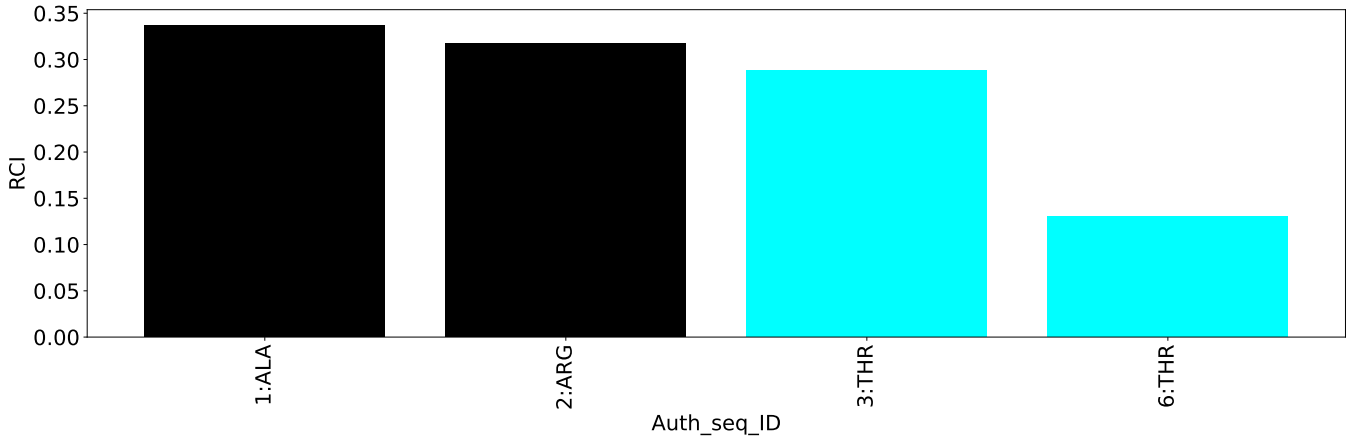
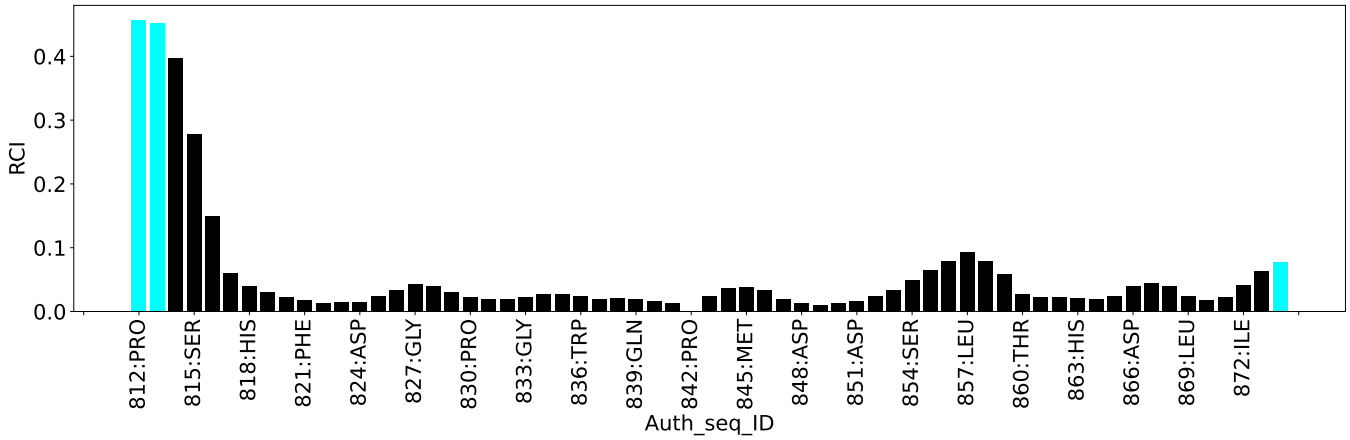
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	B	835	ARG	HB2	-0.27	0.52 – 3.08	-8.1
1	B	871	PRO	HB3	-0.46	0.25 – 3.76	-7.0
1	B	830	PRO	CG	34.20	21.69 – 32.72	6.3
1	B	840	ASP	HB3	1.24	1.32 – 4.00	-5.3
1	B	854	SER	HB3	2.41	2.49 – 5.20	-5.3

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: