Full wwPDB NMR Structure Validation Report (i)

Apr 21, 2024 – 05:25 PM EDT

PDB ID	:	2LKT
BMRB ID	:	18012
Title	:	Solution structure of N-terminal domain of human TIG3 in 2 M UREA
Authors	:	Wang, L.; Yu, W.; Xia, B.
Deposited on	:	2011-10-21

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:

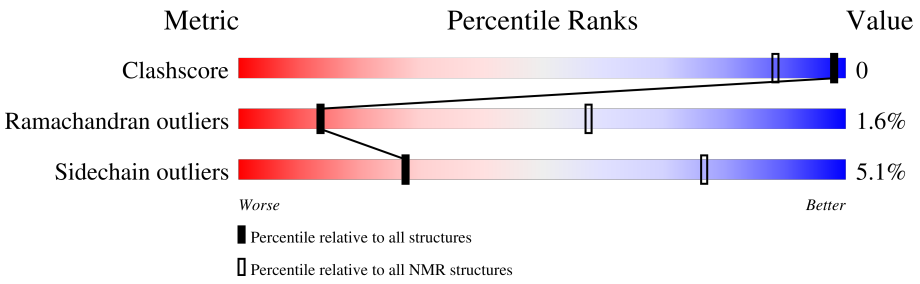
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

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	A	125	77% • • 18%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 19 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *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	A:7-A:39, A:57-A:125 (102)	0.87	19

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 4 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1948 atoms, of which 961 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Retinoic acid receptor responder protein 3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	125	1948	623	961	171	187	6	0

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: Retinoic acid receptor responder protein 3

Chain A: 77% • • 18%

M1 A2 S3 P4 H5 Q6 R18 L26 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R64 R75 Y84 R87 S125

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: Retinoic acid receptor responder protein 3

Chain A: 74% 6% • 18%

M1 A2 S3 P4 H5 Q6 R18 L26 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R64 Y84 R87 Q102 F116 L120 R121 S125

4.2.2 Score per residue for model 2

• Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 74% 6% • 18%

M1 A2 S3 P4 H5 Q6 L19 L26 D30 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R64 R75 D81 Y84 Q85 P86 R87 R121 S125

4.2.3 Score per residue for model 3

• Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 74% 7% • 18%

M1 A2 S3 P4 H5 Q6 R18 H23 L26 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R64 Y84 R87 K105 R111 R121 Y122 S125

4.2.4 Score per residue for model 4

• Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 74% 7% • 18%

M1 A2 S3 P4 H5 Q6 L26 Y27 D30 G31 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 H82 E83 Y84 R87 E90 R111 N112 C113 S125

4.2.5 Score per residue for model 5

• Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 74% 6% • 18%

M1 A2 S3 P4 H5 Q6 D12 L19 L26 Y27 G31 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R64 Y84 R87 R121 S125

4.2.6 Score per residue for model 6

• Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 74% 8% 18%

M1 A2 S3 P4 H5 Q6 D12 L26 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R64 D81 Q85 P86 R87 C113 F116 L120 R121 S125

4.2.7 Score per residue for model 7

• Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 77% • • 18%

M1 A2 S3 P4 H5 Q6 R18 Y27 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R62 E63 R64 Q85 P86 R87 S125

4.2.8 Score per residue for model 8

- Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 74% 6% • 18%

M1 A2 S3 P4 H5 Q6 K9 R18 L19 H23 L26 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R75 Y84 R87 R111 S125

4.2.9 Score per residue for model 9

- Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 75% 6% • 18%

M1 A2 S3 P4 H5 Q6 E7 L26 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 D81 Y84 R87 E90 F116 L120 S125

4.2.10 Score per residue for model 10

- Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 74% 6% • 18%

M1 A2 S3 P4 H5 Q6 K9 L13 L26 Y27 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R64 S79 Y84 R87 Y122 S125

4.2.11 Score per residue for model 11

- Molecule 1: Retinoic acid receptor responder protein 3

Chain A: 77% • • 18%

M1 A2 S3 P4 H5 Q6 R18 L19 L26 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R64 Y84 R87 S125

4.2.12 Score per residue for model 12

- Molecule 1: Retinoic acid receptor responder protein 3

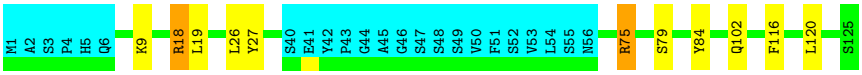
Chain A: 72% 9% • 18%

M1 A2 S3 P4 H5 Q6 K9 R18 L19 L26 S40 E41 Y42 P43 G44 A45 G46 S47 S48 S49 V50 F51 S52 V53 L54 S55 N56 R62 E63 R64 Y84 R87 F116 L120 R121 S125

4.2.13 Score per residue for model 13

- Molecule 1: Retinoic acid receptor responder protein 3

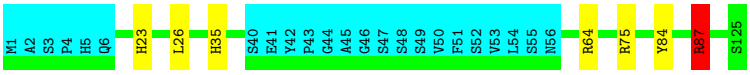
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Retinoic acid receptor responder protein 3

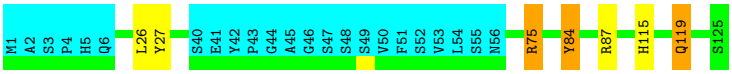
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Retinoic acid receptor responder protein 3

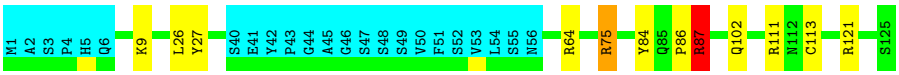
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Retinoic acid receptor responder protein 3

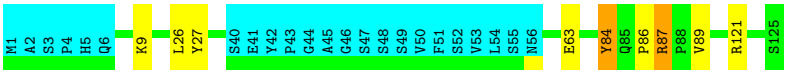
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Retinoic acid receptor responder protein 3

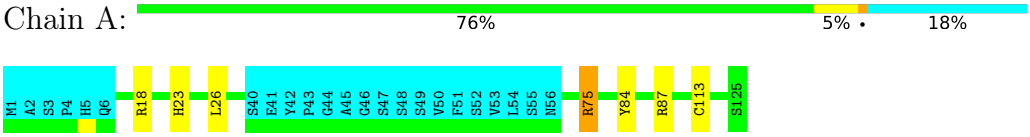
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Retinoic acid receptor responder protein 3

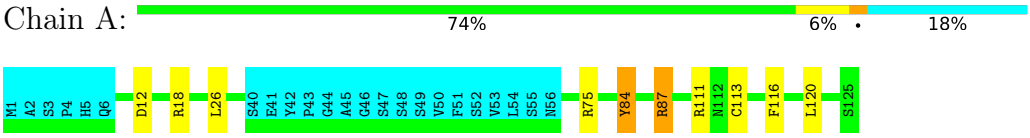
Chain A:



4.2.19 Score per residue for model 19 (medoid)

- Molecule 1: Retinoic acid receptor responder protein 3

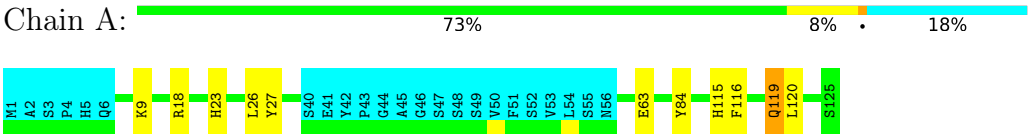
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Retinoic acid receptor responder protein 3

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
Amber	refinement	9.0
CYANA	structure solution	3.0
DYANA	structure solution	3.0
ProcheckNMR	geometry optimization	
TALOS	geometry optimization	

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	1284
Number of shifts mapped to atoms	1284
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)

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	0±0/844 (0.0± 0.0%)	1.04±0.03	3±2/1139 (0.3± 0.1%)
All	All	0.74	0/16880 (0.0%)	1.04	64/22780 (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
1	A	0.0±0.0	1.4±0.9
All	All	0	27

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	87	ARG	NE-CZ-NH1	8.31	124.46	120.30	1	16
1	A	75	ARG	NE-CZ-NH1	7.25	123.93	120.30	18	7
1	A	64	ARG	NE-CZ-NH2	-7.23	116.69	120.30	1	11
1	A	111	ARG	NE-CZ-NH2	-7.18	116.71	120.30	19	4
1	A	64	ARG	NE-CZ-NH1	6.57	123.58	120.30	5	11
1	A	111	ARG	NE-CZ-NH1	6.42	123.51	120.30	4	5
1	A	75	ARG	NE-CZ-NH2	-6.29	117.16	120.30	2	4
1	A	62	ARG	NE-CZ-NH1	5.69	123.15	120.30	7	1
1	A	87	ARG	CD-NE-CZ	5.31	131.04	123.60	14	1
1	A	121	ARG	NE-CZ-NH1	5.30	122.95	120.30	1	3
1	A	18	ARG	NE-CZ-NH1	5.12	122.86	120.30	20	1

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	A	84	TYR	Peptide	15
1	A	27	TYR	Sidechain	9
1	A	121	ARG	Sidechain	3

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	824	811	811	0±1
All	All	16480	16220	16220	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 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:116:PHE:CZ	1:A:120:LEU:HD11	0.54	2.37	1	7
1:A:115:HIS:CD2	1:A:119:GLN:HE22	0.42	2.32	20	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	101/125 (81%)	91±2 (90±2%)	9±2 (8±2%)	2±1 (2±1%)	13	57
All	All	2020/2500 (81%)	1816 (90%)	171 (8%)	33 (2%)	13	57

All 13 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	87	ARG	13
1	A	63	GLU	3
1	A	121	ARG	2
1	A	19	LEU	2
1	A	30	ASP	2
1	A	31	GLY	2
1	A	79	SER	2
1	A	86	PRO	2
1	A	84	TYR	1
1	A	105	LYS	1
1	A	85	GLN	1
1	A	7	GLU	1
1	A	18	ARG	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	90/109 (83%)	85±1 (95±1%)	5±1 (5±1%)	27 77
All	All	1800/2180 (83%)	1709 (95%)	91 (5%)	27 77

All 22 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	26	LEU	19
1	A	18	ARG	9
1	A	84	TYR	7
1	A	9	LYS	7
1	A	75	ARG	6
1	A	23	HIS	5
1	A	87	ARG	5
1	A	113	CYS	5
1	A	19	LEU	4
1	A	102	GLN	3
1	A	81	ASP	3
1	A	85	GLN	3
1	A	12	ASP	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	122	TYR	2
1	A	90	GLU	2
1	A	119	GLN	2
1	A	121	ARG	1
1	A	82	HIS	1
1	A	13	LEU	1
1	A	62	ARG	1
1	A	35	HIS	1
1	A	89	VAL	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)

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 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: *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	1284
Number of shifts mapped to atoms	1284
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	8

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	117	-0.16 ± 0.09	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	106	0.25 ± 0.10	None needed (< 0.5 ppm)
$^{13}C'$	99	0.16 ± 0.11	None needed (< 0.5 ppm)
^{15}N	96	-0.32 ± 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 81%, i.e. 1149 atoms were assigned a chemical shift out of a possible 1420. 0 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	474/506 (94%)	198/206 (96%)	189/204 (93%)	87/96 (91%)
Sidechain	651/791 (82%)	433/513 (84%)	213/245 (87%)	5/33 (15%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	24/123 (20%)	23/60 (38%)	0/58 (0%)	1/5 (20%)
Overall	1149/1420 (81%)	654/779 (84%)	402/507 (79%)	93/134 (69%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 76%, i.e. 1281 atoms were assigned a chemical shift out of a possible 1682. 0 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	535/619 (86%)	223/252 (88%)	216/250 (86%)	96/117 (82%)
Sidechain	718/914 (79%)	476/595 (80%)	236/284 (83%)	6/35 (17%)
Aromatic	28/149 (19%)	27/73 (37%)	0/70 (0%)	1/6 (17%)
Overall	1281/1682 (76%)	726/920 (79%)	452/604 (75%)	103/158 (65%)

7.1.4 Statistically unusual chemical shifts (i)

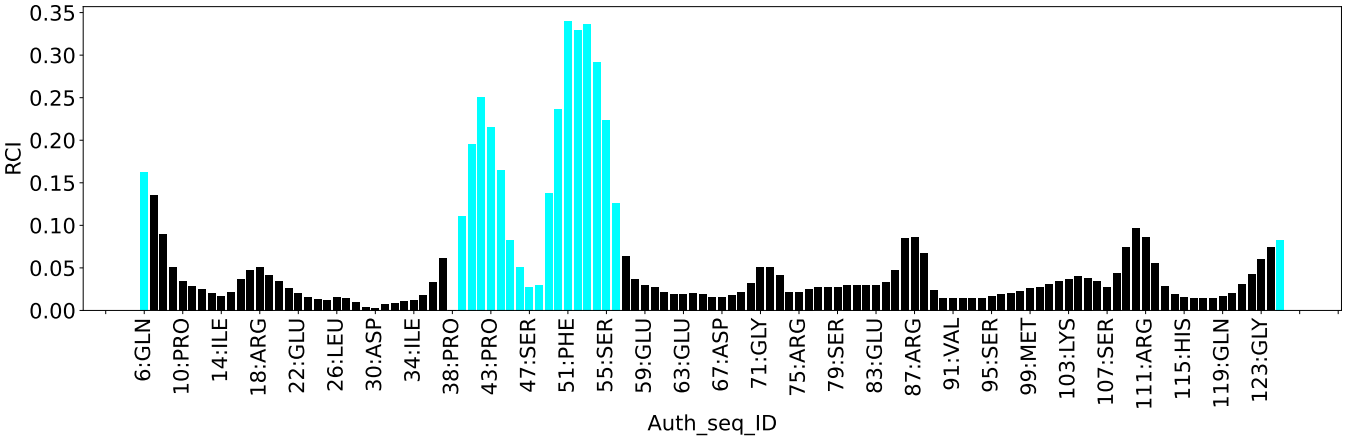
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	30	ASP	C	124.90	168.07 – 184.82	-30.8
1	A	15	GLU	HB3	0.09	0.95 – 3.05	-9.1
1	A	114	GLU	HG3	0.42	1.20 – 3.30	-8.7
1	A	47	SER	CB	75.81	56.28 – 71.32	8.0
1	A	15	GLU	HG2	0.67	1.24 – 3.30	-7.8
1	A	107	SER	CB	74.79	56.28 – 71.32	7.3
1	A	15	GLU	HG3	0.74	1.20 – 3.30	-7.2
1	A	121	ARG	HG3	-0.09	0.15 – 2.94	-5.9

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:


RCI
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
6:GLN
10:PRO
14:ILE
18:ARG
22:GLU
26:LEU
30:ASP
34:ILE
38:PRO
43:PRO
47:SER
51:PHE
55:SER
59:GLU
63:GLU
67:ASP
71:GLY
75:ARG
79:SER
83:GLU
87:ARG
91:VAL
95:SER
99:MET
103:LYS
107:SER
111:ARG
115:HIS
119:GLN
123:GLY
Auth_seq_ID