Full wwPDB NMR Structure Validation Report (i)

May 31, 2020 – 10:07 pm BST

PDB ID	:	6EZ4
Title	:	NMR structure of the C-terminal domain of the human RPAP3 protein
Authors	:	Fabre, P.; Chagot, M.E.; Bragantini, B.; Manival, X.; Quinternet, M.
Deposited on	:	2017-11-14

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 following versions of software and data (see references (1)) were used in the production of this report:

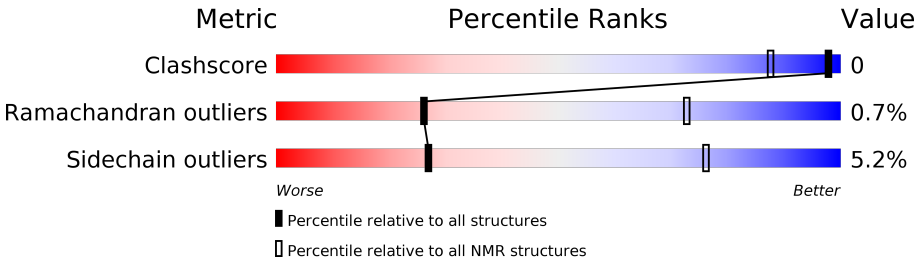
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 92%.

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	135	84% 8% 8%

2 Ensemble composition and analysis

This entry contains 20 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *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:542-A:665 (124)	0.22	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 3 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2250 atoms, of which 1137 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called RNA polymerase II-associated protein 3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	135	2250	724	1137	182	202	5	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	531	GLY	-	expression tag	UNP Q9H6T3
A	532	PRO	-	expression tag	UNP Q9H6T3
A	533	HIS	-	expression tag	UNP Q9H6T3
A	534	MET	-	expression tag	UNP Q9H6T3

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: RNA polymerase II-associated protein 3

Chain A:

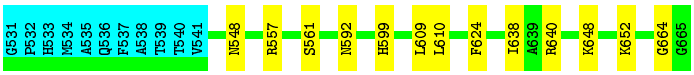
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: RNA polymerase II-associated protein 3

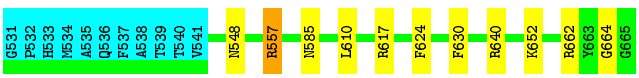
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: RNA polymerase II-associated protein 3

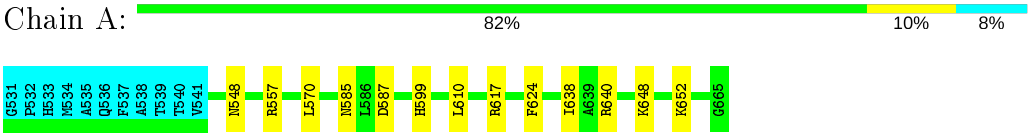
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: RNA polymerase II-associated protein 3

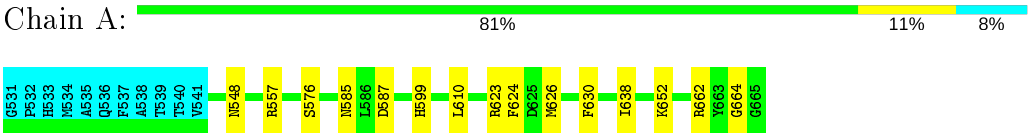
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: RNA polymerase II-associated protein 3

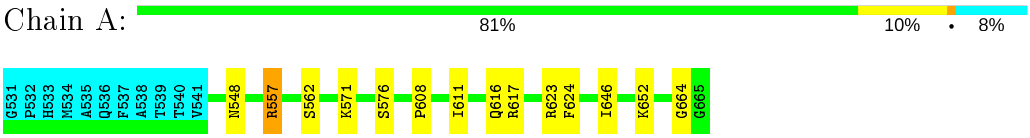
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: RNA polymerase II-associated protein 3

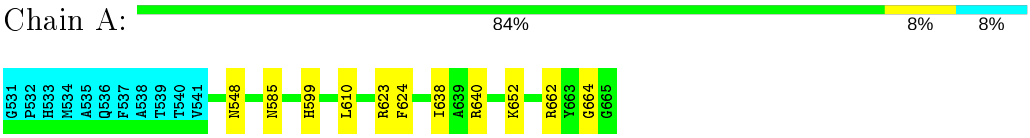
Chain A:



4.2.6 Score per residue for model 6 (medoid)

• Molecule 1: RNA polymerase II-associated protein 3

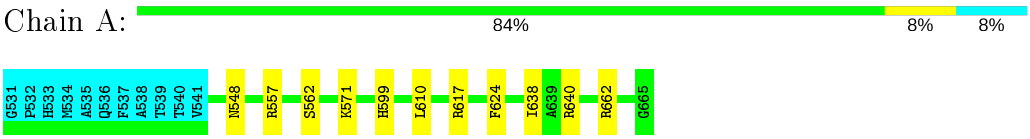
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: RNA polymerase II-associated protein 3

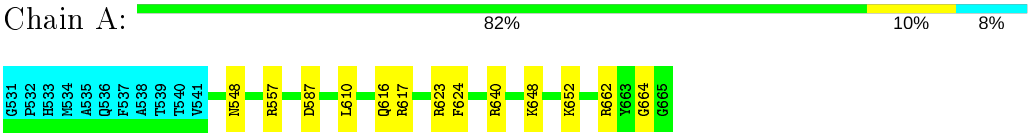
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: RNA polymerase II-associated protein 3

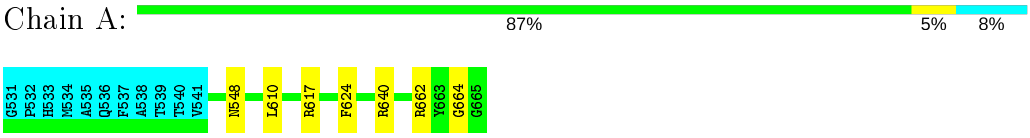
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: RNA polymerase II-associated protein 3

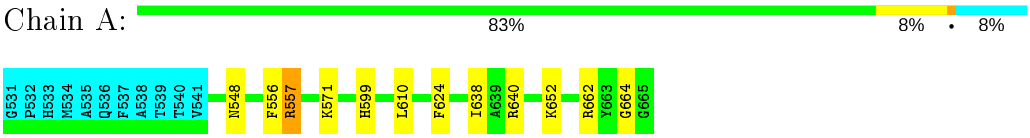
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: RNA polymerase II-associated protein 3

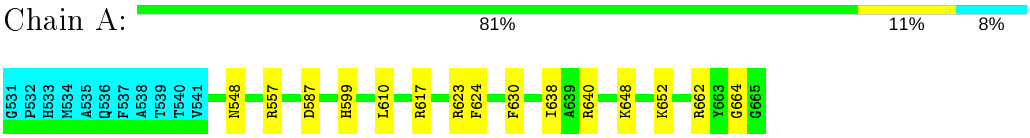
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: RNA polymerase II-associated protein 3

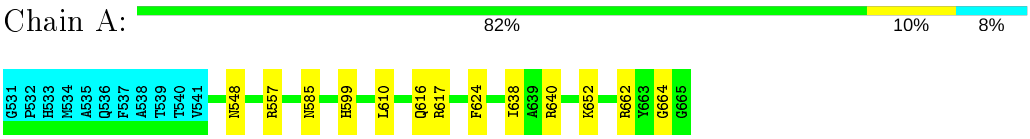
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: RNA polymerase II-associated protein 3

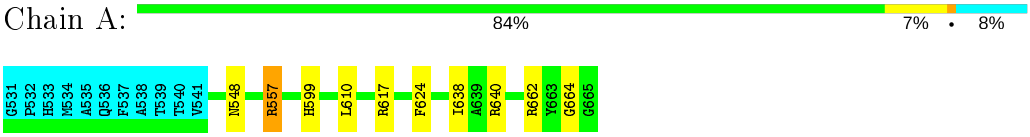
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: RNA polymerase II-associated protein 3

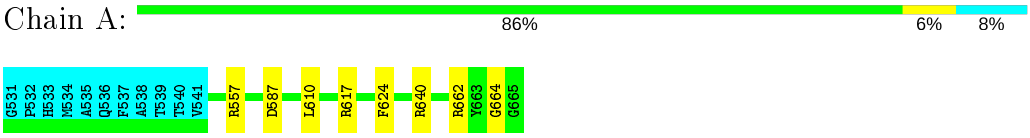
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: RNA polymerase II-associated protein 3

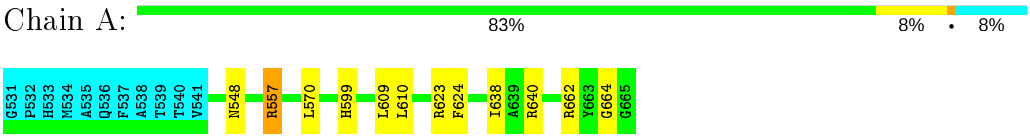
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: RNA polymerase II-associated protein 3

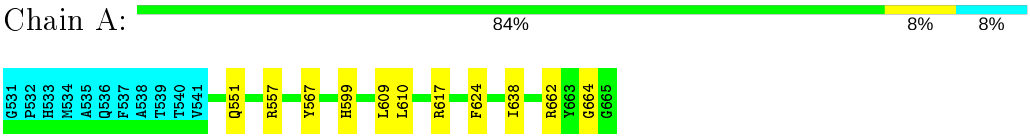
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: RNA polymerase II-associated protein 3

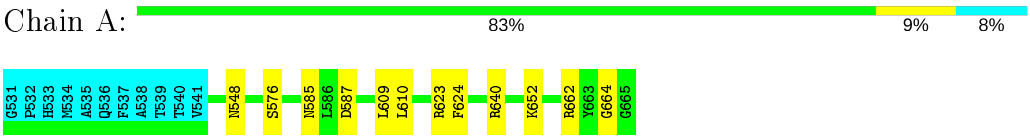
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: RNA polymerase II-associated protein 3

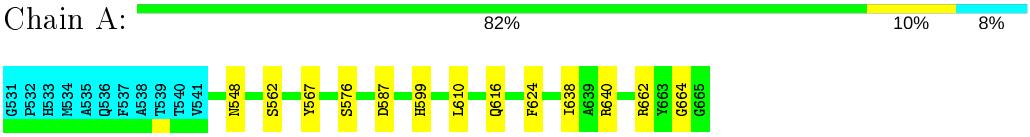
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: RNA polymerase II-associated protein 3

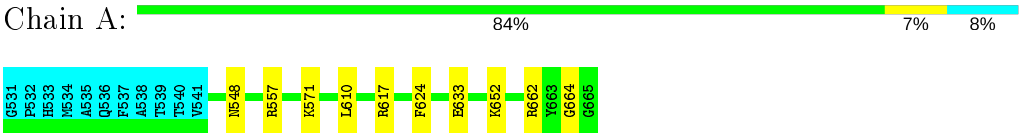
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: RNA polymerase II-associated protein 3

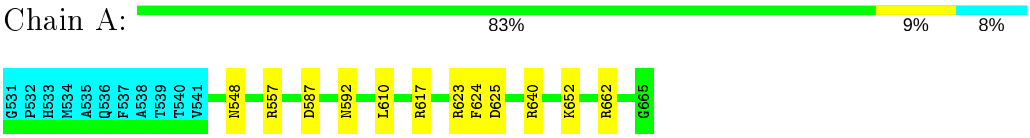
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: RNA polymerase II-associated protein 3

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest restraint energies*.

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

Software name	Classification	Version
CYANA	structure calculation	
TALOS	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 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1798
Number of shifts mapped to atoms	1798
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	92%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	1033	1059	1059	1±1
All	All	20660	21180	21180	15

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:599:HIS:CG	1:A:638:ILE:HG23	0.49	2.43	4	10
1:A:599:HIS:CD2	1:A:638:ILE:HG23	0.44	2.48	12	3
1:A:608:PRO:HA	1:A:611:ILE:HD12	0.41	1.92	5	1
1:A:611:ILE:HG22	1:A:646:ILE:HD11	0.40	1.92	5	1

5.2 Torsion angles (i)

5.2.1 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	123/135 (91%)	120±1 (98±1%)	2±1 (2±1%)	1±0 (1±0%)	26	73
All	All	2460/2700 (91%)	2406 (98%)	37 (2%)	17 (1%)	26	73

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	664	GLY	17

5.2.2 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	117/125 (94%)	111±2 (95±1%)	6±2 (5±1%)	27	76
All	All	2340/2500 (94%)	2219 (95%)	121 (5%)	27	76

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	624	PHE	20
1	A	610	LEU	19

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	548	ASN	18
1	A	652	LYS	13
1	A	587	ASP	8
1	A	609	LEU	4
1	A	576	SER	4
1	A	648	LYS	4
1	A	557	ARG	4
1	A	571	LYS	4
1	A	616	GLN	4
1	A	562	SER	3
1	A	585	ASN	3
1	A	630	PHE	3
1	A	570	LEU	2
1	A	592	ASN	2
1	A	625	ASP	1
1	A	561	SER	1
1	A	626	MET	1
1	A	633	GLU	1
1	A	551	GLN	1
1	A	556	PHE	1

5.2.3 RNA

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains

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

5.4 Carbohydrates

There are no carbohydrates in this entry.

5.5 Ligand geometry

There are no ligands in this entry.

5.6 Other polymers

There are no such molecules in this entry.

5.7 Polymer linkage issues

There are no chain breaks in this entry.

6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 92% for the well-defined parts and 91% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *535665_ref_dss_298K.str*

6.1.1 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	1798
Number of shifts mapped to atoms	1798
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	134	-0.50 ± 0.17	Should be applied
$^{13}C_{\beta}$	131	0.25 ± 0.07	None needed (< 0.5 ppm)
$^{13}C'$	133	-0.27 ± 0.11	None needed (< 0.5 ppm)
^{15}N	125	0.53 ± 0.16	Should be applied

6.1.3 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 92%, i.e. 1533 atoms were assigned a chemical shift out of a possible 1663. 21 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	603/604 (100%)	240/240 (100%)	247/248 (100%)	116/116 (100%)
Sidechain	840/924 (91%)	526/549 (96%)	298/335 (89%)	16/40 (40%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	90/135 (67%)	63/73 (86%)	27/60 (45%)	0/2 (0%)
Overall	1533/1663 (92%)	829/862 (96%)	572/643 (89%)	132/158 (84%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 91%, i.e. 1629 atoms were assigned a chemical shift out of a possible 1783. 22 out of 22 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	651/657 (99%)	259/261 (99%)	267/270 (99%)	125/126 (99%)
Sidechain	886/975 (91%)	554/579 (96%)	315/355 (89%)	17/41 (41%)
Aromatic	92/151 (61%)	65/82 (79%)	27/66 (41%)	0/3 (0%)
Overall	1629/1783 (91%)	878/922 (95%)	609/691 (88%)	142/170 (84%)

6.1.4 Statistically unusual chemical shifts (i)

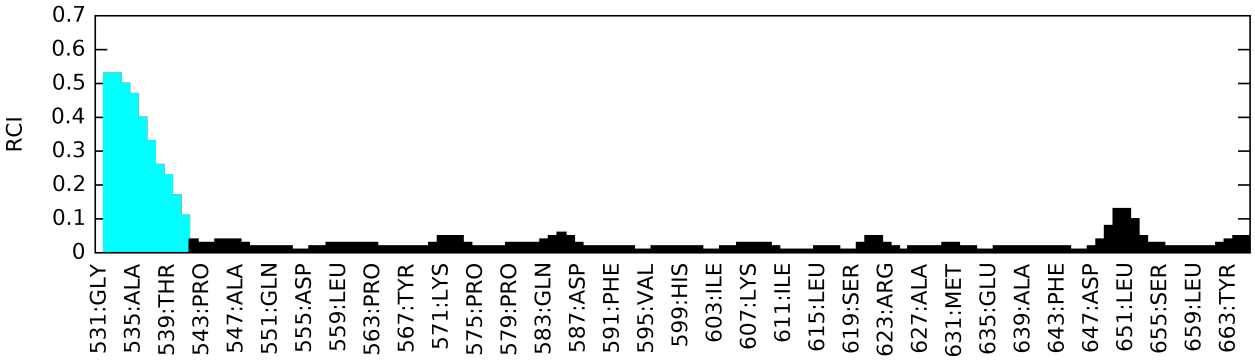
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	619	SER	HA	1.47	6.53 – 2.43	-7.3
1	A	569	TYR	HH	1.01	16.26 – 2.26	-5.9

6.1.5 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.

Random coil index (RCI) for chain A:


RCI
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
531:GLY
535:ALA
539:THR
543:PRO
547:ALA
551:GLN
555:ASP
559:LEU
563:PRO
567:TYR
571:LYS
575:PRO
579:PRO
583:GLN
587:ASP
591:PHE
595:VAL
599:HIS
603:ILE
607:LYS
611:ILE
615:LEU
619:SER
623:ARG
627:ALA
631:MET
635:GLU
639:ALA
643:PHE
647:ASP
651:LEU
655:SER
659:LEU
663:TYR