Full wwPDB NMR Structure Validation Report (i)

Jun 16, 2024 – 07:40 PM EDT

PDB ID	:	5M9Z
BMRB ID	:	34061
Title	:	Second zinc-binding domain from yeast Pcf11
Authors	:	Mackereth, C.
Deposited on	:	2016-11-02

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:

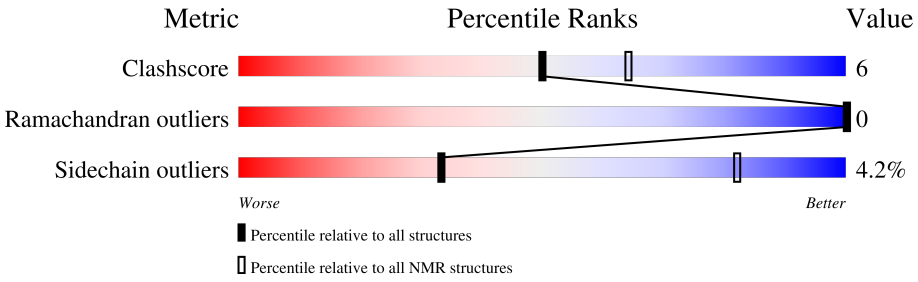
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 90%.

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	97	45% 7% 5% 42%

2 Ensemble composition and analysis

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	A:548-A:554, A:560-A:603 (51)	0.52	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 3 clusters and 4 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 871 atoms, of which 417 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protein PCF11.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	56	870	289	417	70	90	4	0

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	1	1

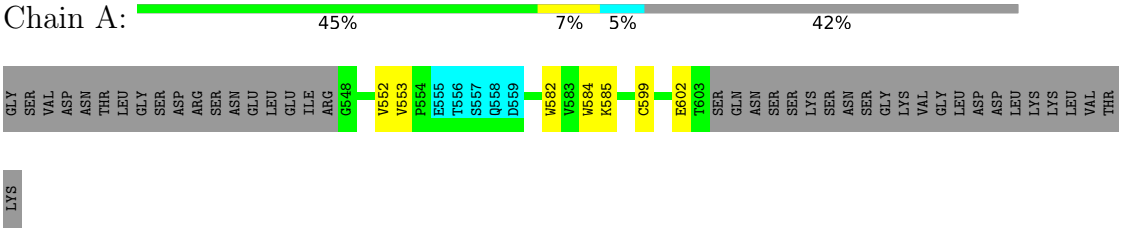
4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: Protein PCF11

Chain A:



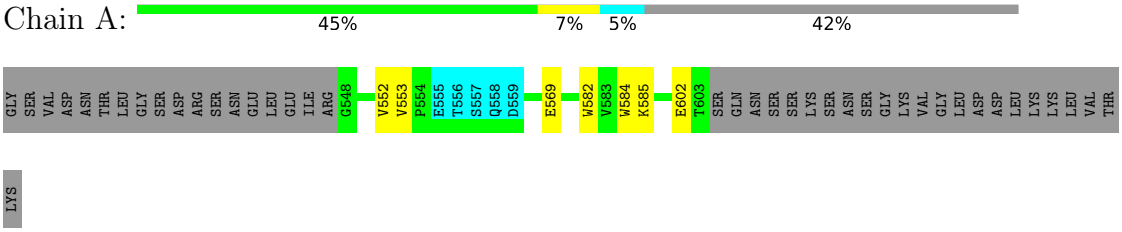
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

• Molecule 1: Protein PCF11

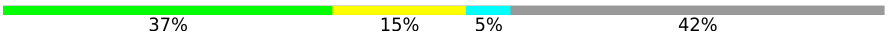
Chain A:

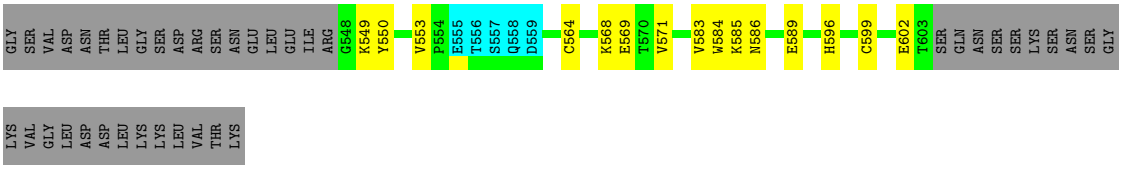


4.2.2 Score per residue for model 2

• Molecule 1: Protein PCF11

Chain A:

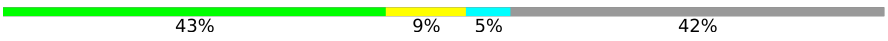


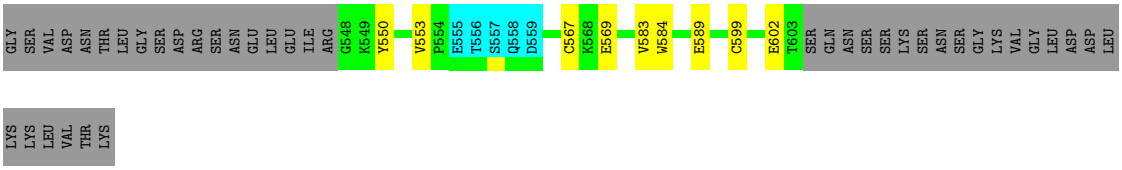


4.2.3 Score per residue for model 3

• Molecule 1: Protein PCF11

Chain A:

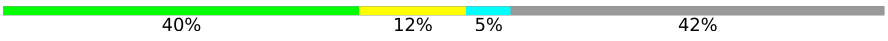


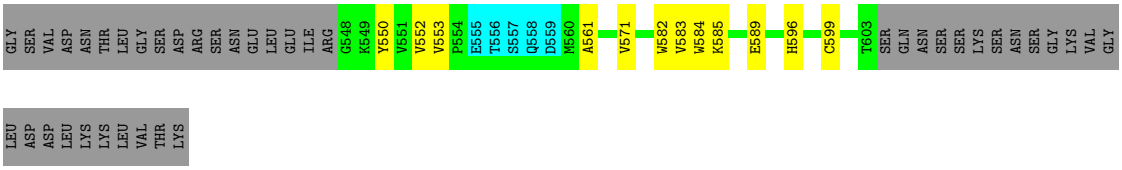


4.2.4 Score per residue for model 4

• Molecule 1: Protein PCF11

Chain A:

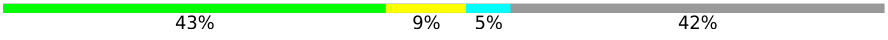


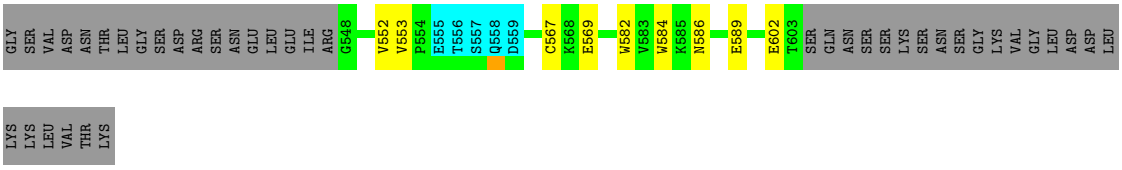


4.2.5 Score per residue for model 5

• Molecule 1: Protein PCF11

Chain A:

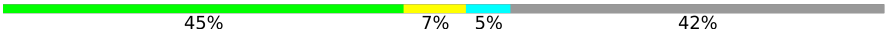


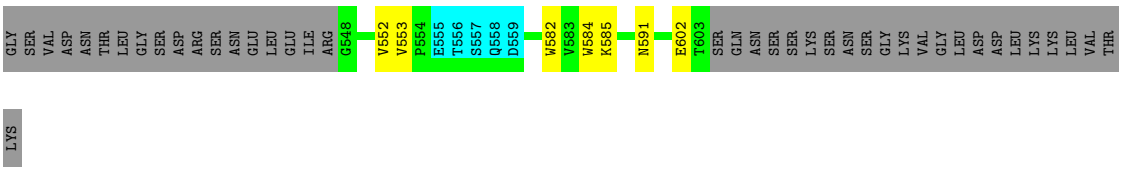


4.2.6 Score per residue for model 6

• Molecule 1: Protein PCF11

Chain A:

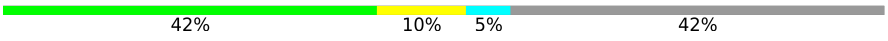


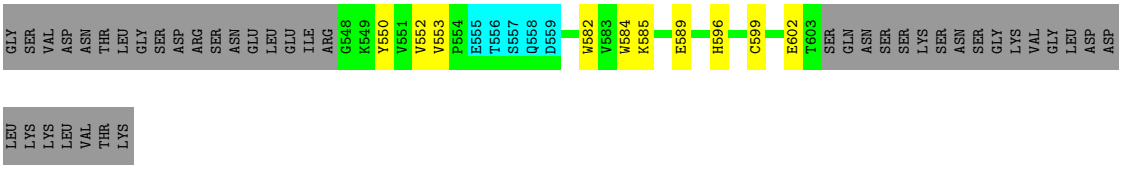


4.2.7 Score per residue for model 7

• Molecule 1: Protein PCF11

Chain A:

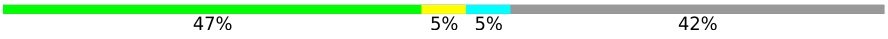


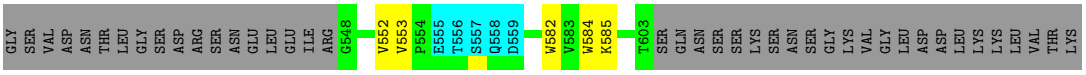


4.2.8 Score per residue for model 8

• Molecule 1: Protein PCF11

Chain A:

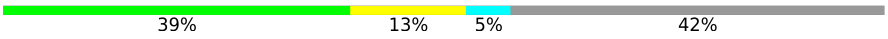


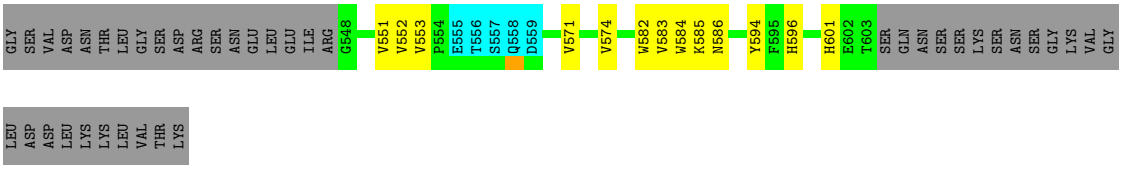


4.2.9 Score per residue for model 9

• Molecule 1: Protein PCF11

Chain A:

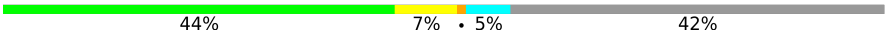


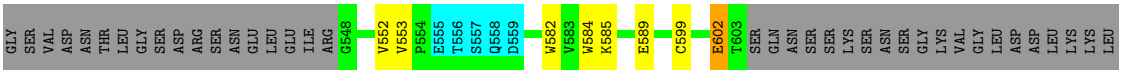


4.2.10 Score per residue for model 10

• Molecule 1: Protein PCF11

Chain A:

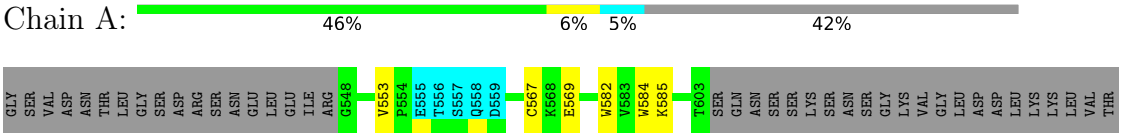




4.2.11 Score per residue for model 11

• Molecule 1: Protein PCF11

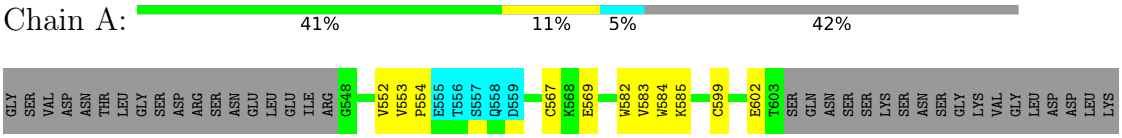
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Protein PCF11

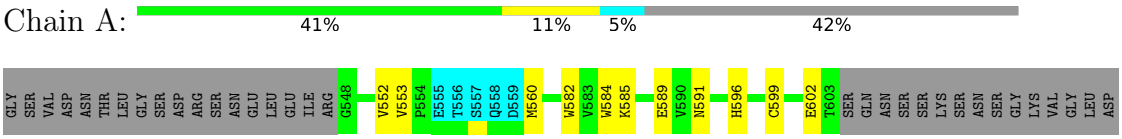
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Protein PCF11

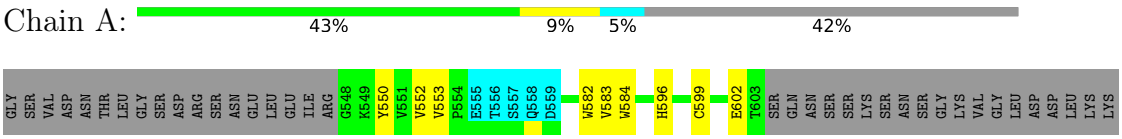
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Protein PCF11

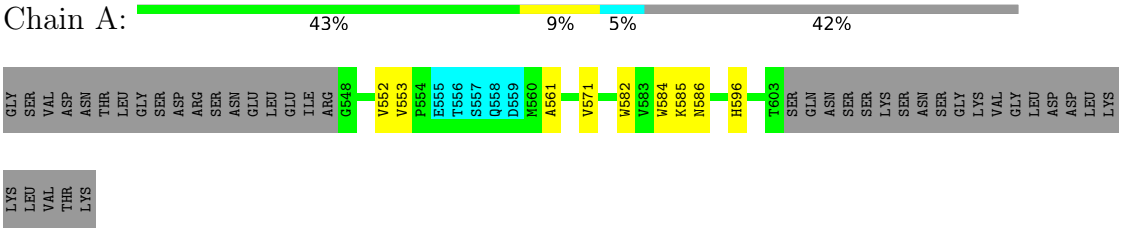
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Protein PCF11

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 40 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
ARIA	structure calculation	

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	633
Number of shifts mapped to atoms	633
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

6 Model quality

6.1 Standard geometry

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

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	414	387	386	5±2
All	All	6225	5805	5790	72

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 6.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:596:HIS:HB2	1:A:599:CYS:SG	0.65	2.31	7	5
1:A:552:VAL:HA	1:A:582:TRP:O	0.61	1.96	12	12
1:A:599:CYS:HA	1:A:602:GLU:HG2	0.58	1.75	7	2
1:A:564:CYS:O	1:A:568:LYS:HA	0.58	1.98	2	1
1:A:553:VAL:HG22	1:A:584:TRP:CD1	0.58	2.33	14	15
1:A:571:VAL:HG22	1:A:596:HIS:CE1	0.56	2.35	9	3
1:A:599:CYS:O	1:A:602:GLU:HG2	0.56	2.00	14	4
1:A:549:LYS:HD2	1:A:549:LYS:N	0.55	2.15	2	1
1:A:567:CYS:SG	1:A:569:GLU:HG3	0.51	2.45	3	4
1:A:551:VAL:HG11	1:A:594:TYR:HE1	0.51	1.64	9	1
1:A:586:ASN:O	1:A:596:HIS:HA	0.49	2.07	2	2
1:A:550:TYR:CE1	1:A:583:VAL:HG12	0.47	2.44	4	4

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:553:VAL:HB	1:A:582:TRP:HB2	0.47	1.85	5	1
1:A:574:VAL:O	1:A:583:VAL:HG22	0.47	2.10	9	1
1:A:553:VAL:HB	1:A:582:TRP:CB	0.46	2.40	14	3
1:A:553:VAL:HB	1:A:582:TRP:HB3	0.46	1.86	7	1
1:A:561:ALA:HA	1:A:571:VAL:O	0.46	2.11	15	2
1:A:599:CYS:HA	1:A:602:GLU:HB3	0.45	1.88	10	1
1:A:599:CYS:O	1:A:602:GLU:HB3	0.43	2.14	12	2
1:A:553:VAL:CG2	1:A:582:TRP:HB3	0.42	2.43	14	4
1:A:583:VAL:HG21	1:A:585:LYS:HE2	0.42	1.91	12	1
1:A:550:TYR:HB3	1:A:585:LYS:HD3	0.42	1.90	7	1
1:A:599:CYS:CA	1:A:602:GLU:HG2	0.42	2.44	7	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	A	49/97 (51%)	49±0 (99±1%)	0±0 (1±1%)	0±0 (0±0%)	100	100
All	All	735/1455 (51%)	730 (99%)	5 (1%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	46/88 (52%)	44±1 (96±3%)	2±1 (4±3%)	33	82
All	All	690/1320 (52%)	661 (96%)	29 (4%)	33	82

All 7 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	585	LYS	10
1	A	589	GLU	7
1	A	602	GLU	5
1	A	569	GLU	2
1	A	586	ASN	2
1	A	591	ASN	2
1	A	601	HIS	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)

Of 1 ligands modelled in this entry, 1 is 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

The completeness of assignment taking into account all chemical shift lists is 90% for the well-defined parts and 86% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *pcf11_ZNF2_FINAL_shifts.str*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	633
Number of shifts mapped to atoms	633
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	53	-0.22 ± 0.45	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	49	0.10 ± 0.36	None needed (< 0.5 ppm)
$^{13}C'$	49	0.02 ± 0.32	None needed (< 0.5 ppm)
^{15}N	47	0.53 ± 0.95	None needed (imprecise)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 90%, i.e. 608 atoms were assigned a chemical shift out of a possible 678. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	238/255 (93%)	97/104 (93%)	96/102 (94%)	45/49 (92%)
Sidechain	294/328 (90%)	201/213 (94%)	92/108 (85%)	1/7 (14%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	76/95 (80%)	42/46 (91%)	30/44 (68%)	4/5 (80%)
Overall	608/678 (90%)	340/363 (94%)	218/254 (86%)	50/61 (82%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 86%, i.e. 633 atoms were assigned a chemical shift out of a possible 733. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	251/280 (90%)	102/114 (89%)	102/112 (91%)	47/54 (87%)
Sidechain	306/358 (85%)	209/231 (90%)	96/119 (81%)	1/8 (12%)
Aromatic	76/95 (80%)	42/46 (91%)	30/44 (68%)	4/5 (80%)
Overall	633/733 (86%)	353/391 (90%)	228/275 (83%)	52/67 (78%)

7.1.4 Statistically unusual chemical shifts (i)

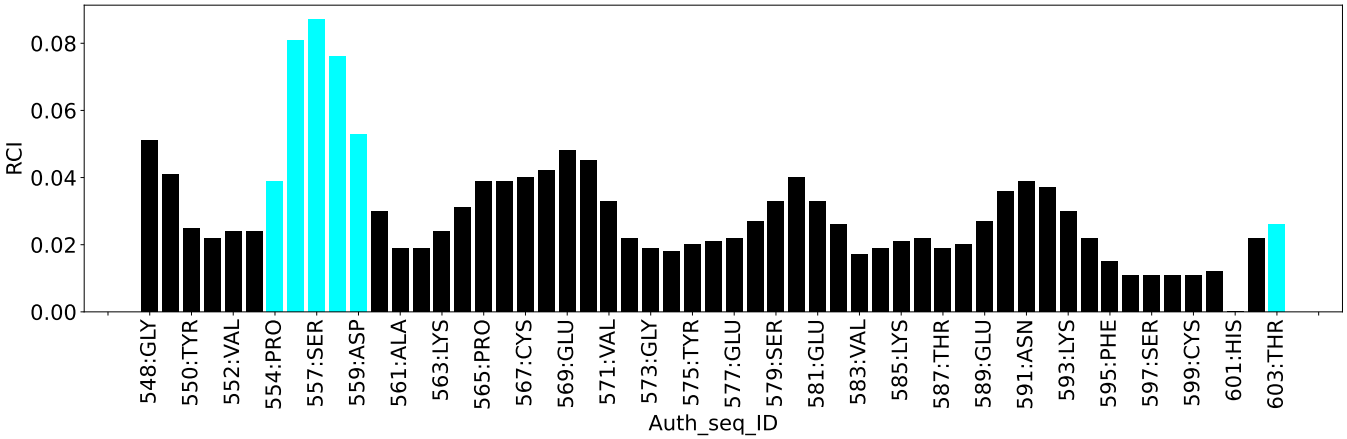
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	571	VAL	HB	0.12	0.43 – 3.54	-6.0
1	A	575	TYR	HB3	0.80	0.93 – 4.76	-5.3
1	A	594	TYR	HB2	1.00	1.09 – 4.72	-5.2

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
0.00
0.02
0.04
0.06
0.08
548:GLY
550:TYR
552:VAL
554:PRO
557:SER
559:ASP
561:ALA
563:LYS
565:PRO
567:CYS
569:GLU
571:VAL
573:GLY
575:TYR
577:GLU
579:SER
581:GLU
583:VAL
585:LYS
587:THR
589:GLU
591:ASN
593:LYS
595:PHE
597:SER
599:CYS
601:HIS
603:THR
Auth_seq_ID