



# Full wwPDB NMR Structure Validation Report ⓘ

Aug 30, 2021 – 02:21 PM EDT

PDB ID : 6NZ2  
Title : NMR solution structure of Bcd1p120-303 from *Saccharomyces cerevisiae*  
Authors : Bragantini, B.; Quinternet, M.; Charpentier, B.; Manival, X.  
Deposited on : 2019-02-12

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : 2.23.1  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.23.1

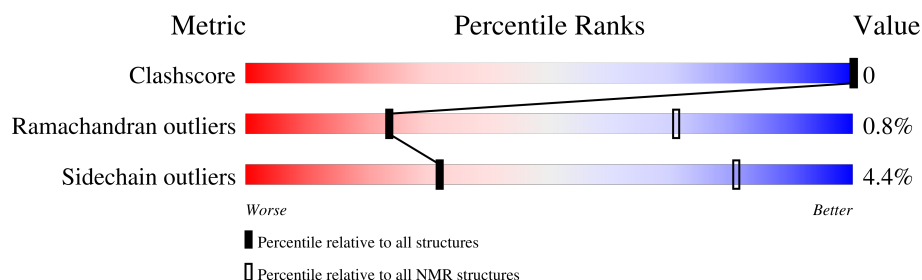
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	188	<div> <div></div> <div>76%</div> <div>7%</div> <div>17%</div> </div>

## 2 Ensemble composition and analysis

This entry contains 20 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:127-A:171, A:177-A:212, A:226-A:300 (156)	0.31	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 and 2 single-model clusters were found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 3052 atoms, of which 1526 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Box C/D snoRNA protein 1.

Mol	Chain	Residues	Atoms						Trace
1	A	188	Total	C	H	N	O	S	0
			3052	963	1526	262	286	15	

There are 4 discrepancies between the modelled and reference sequences:

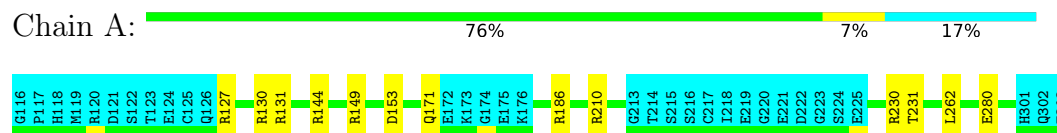
Chain	Residue	Modelled	Actual	Comment	Reference
A	116	GLY	-	expression tag	UNP P38772
A	117	PRO	-	expression tag	UNP P38772
A	118	HIS	-	expression tag	UNP P38772
A	119	MET	-	expression tag	UNP P38772

## 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: Box C/D snoRNA protein 1

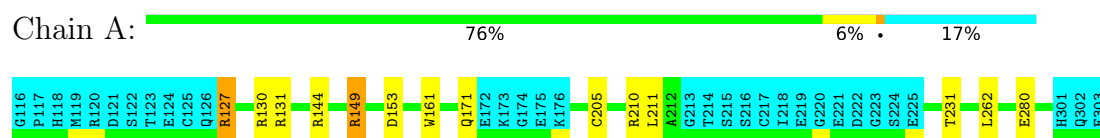


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

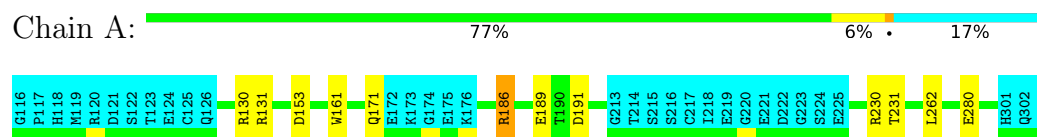
#### 4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Box C/D snoRNA protein 1



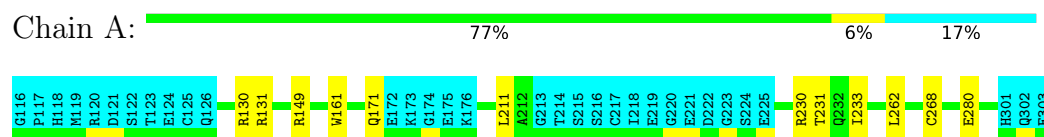
#### 4.2.2 Score per residue for model 2

- Molecule 1: Box C/D snoRNA protein 1



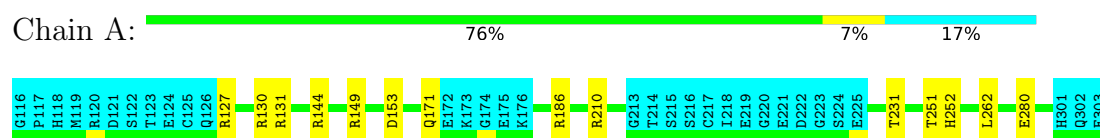
### 4.2.3 Score per residue for model 3

- Molecule 1: Box C/D snoRNA protein 1



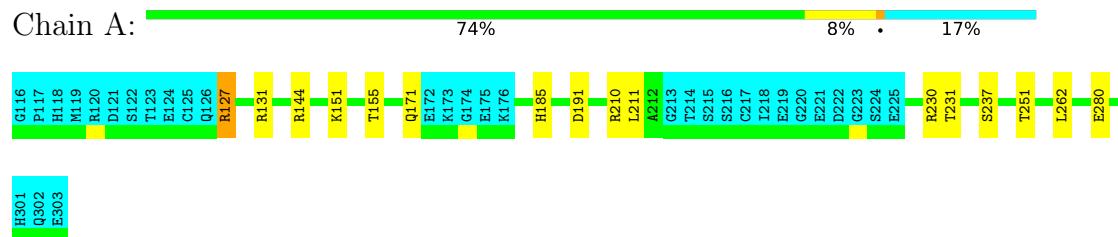
### 4.2.4 Score per residue for model 4

- Molecule 1: Box C/D snoRNA protein 1



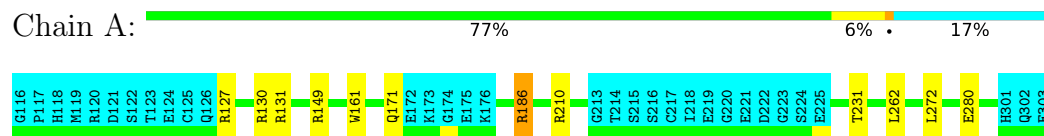
### 4.2.5 Score per residue for model 5

- Molecule 1: Box C/D snoRNA protein 1



### 4.2.6 Score per residue for model 6

- Molecule 1: Box C/D snoRNA protein 1



### 4.2.7 Score per residue for model 7

- Molecule 1: Box C/D snoRNA protein 1

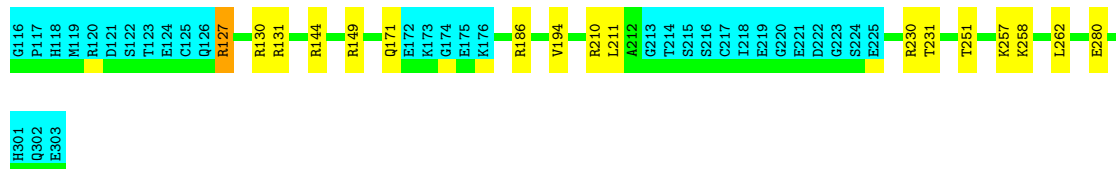




#### 4.2.8 Score per residue for model 8

- Molecule 1: Box C/D snoRNA protein 1

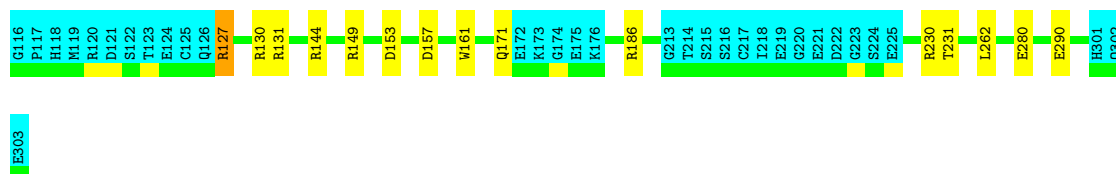
Chain A: 74% 9% 17%



#### 4.2.9 Score per residue for model 9

- Molecule 1: Box C/D snoRNA protein 1

Chain A: 75% 7% 17%



#### 4.2.10 Score per residue for model 10

- Molecule 1: Box C/D snoRNA protein 1

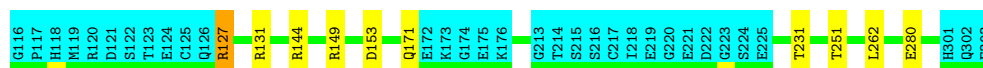
Chain A: 77% 6% 17%



#### 4.2.11 Score per residue for model 11

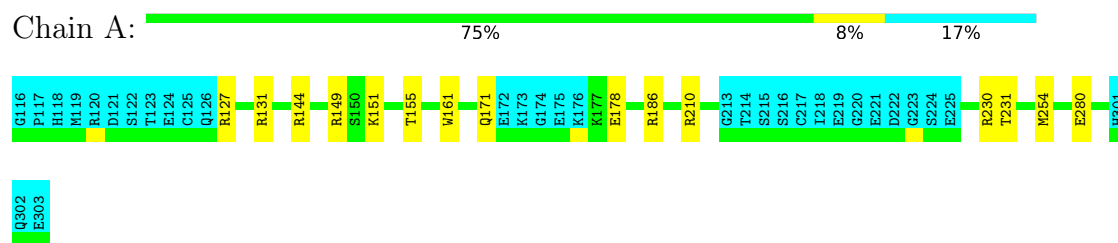
- Molecule 1: Box C/D snoRNA protein 1

Chain A: 78% 5% 17%



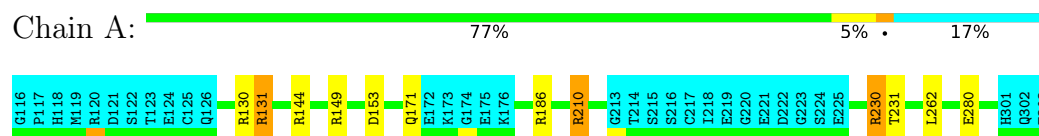
### 4.2.12 Score per residue for model 12

- Molecule 1: Box C/D snoRNA protein 1



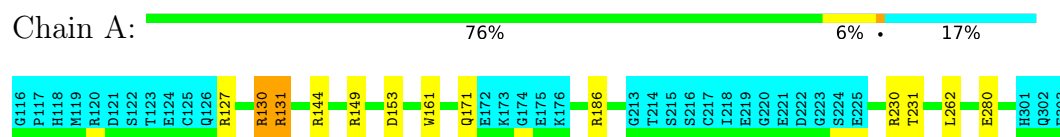
### 4.2.13 Score per residue for model 13

- Molecule 1: Box C/D snoRNA protein 1



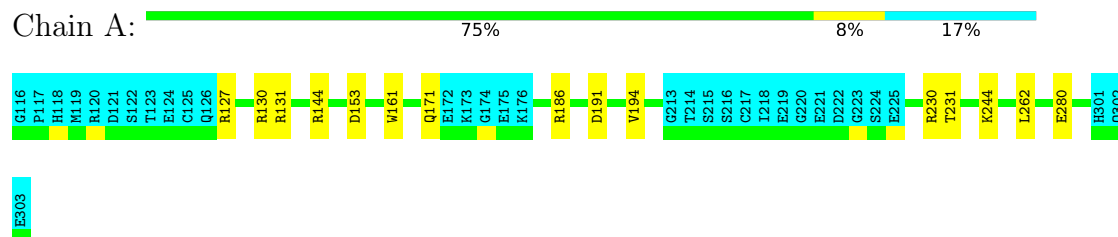
### 4.2.14 Score per residue for model 14

- Molecule 1: Box C/D snoRNA protein 1



### 4.2.15 Score per residue for model 15

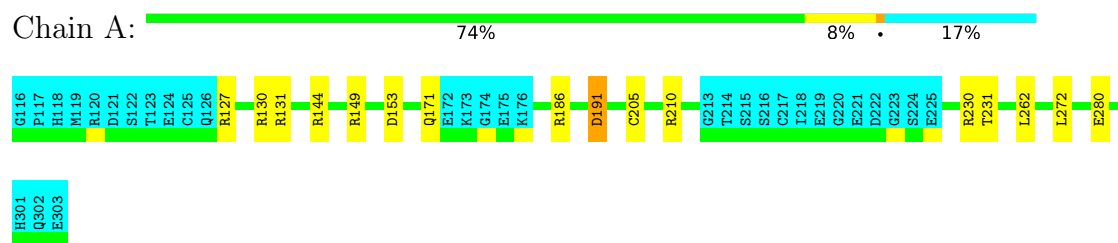
- Molecule 1: Box C/D snoRNA protein 1





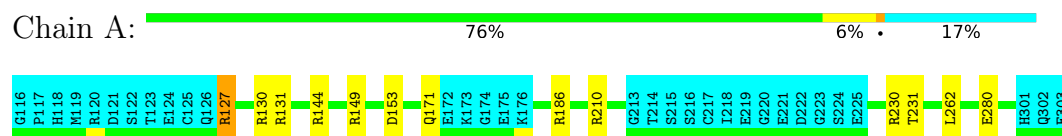
#### 4.2.16 Score per residue for model 16

- Molecule 1: Box C/D snoRNA protein 1



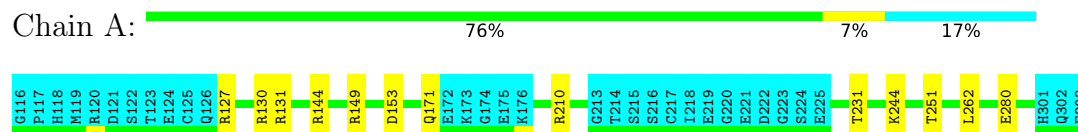
#### 4.2.17 Score per residue for model 17

- Molecule 1: Box C/D snoRNA protein 1



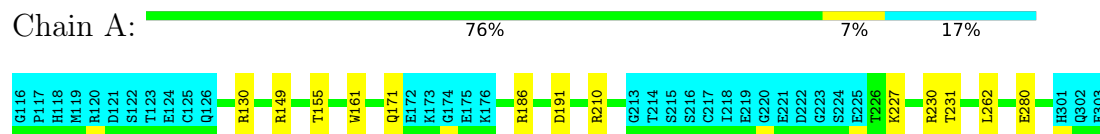
#### 4.2.18 Score per residue for model 18

- Molecule 1: Box C/D snoRNA protein 1



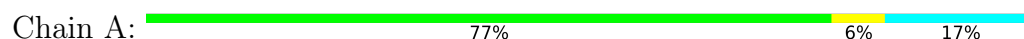
#### 4.2.19 Score per residue for model 19

- Molecule 1: Box C/D snoRNA protein 1



#### 4.2.20 Score per residue for model 20

- Molecule 1: Box C/D snoRNA protein 1



G116	P117	H118	M119	R120	D121	S122	T123	E124	C125	Q126	R127	R131	R144	R149	Q171	E172	K173	G174	E175	K176	R186	D191	C205	R210	G213	T214	S215	S216	C217	I218	E219	G220	E221	D222	G223	S224	E225	T231	E280	H301	Q302	E303
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

## 5 Refinement protocol and experimental data overview

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
CNS	refinement	
CYANA	structure calculation	
TALOS	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	2381
Number of shifts mapped to atoms	2381
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 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	0.68±0.01	0±0/1314 ( 0.0± 0.0%)	1.05±0.02	6±2/1767 ( 0.3± 0.1%)
All	All	0.68	0/26280 ( 0.0%)	1.05	123/35340 ( 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	0.8±0.9
All	All	0	15

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	149	ARG	NE-CZ-NH1	10.63	125.61	120.30	16	14
1	A	210	ARG	NE-CZ-NH1	9.11	124.86	120.30	13	11
1	A	130	ARG	NE-CZ-NH1	9.05	124.82	120.30	13	13
1	A	230	ARG	NE-CZ-NH1	8.49	124.54	120.30	8	14
1	A	144	ARG	NE-CZ-NH1	8.38	124.49	120.30	14	15
1	A	186	ARG	NE-CZ-NH1	8.27	124.43	120.30	15	14
1	A	131	ARG	NE-CZ-NH1	8.17	124.39	120.30	3	19
1	A	127	ARG	NE-CZ-NH1	7.13	123.87	120.30	20	11
1	A	131	ARG	NE-CZ-NH2	-7.13	116.74	120.30	3	4
1	A	149	ARG	NE-CZ-NH2	-6.50	117.05	120.30	14	2
1	A	130	ARG	NE-CZ-NH2	-6.45	117.08	120.30	15	1
1	A	127	ARG	CD-NE-CZ	5.80	131.73	123.60	18	1
1	A	186	ARG	NE-CZ-NH2	-5.76	117.42	120.30	19	1
1	A	149	ARG	CD-NE-CZ	5.73	131.62	123.60	14	1
1	A	191	ASP	CB-CG-OD1	-5.72	113.15	118.30	16	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	131	ARG	CD-NE-CZ	5.04	130.66	123.60	11	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	161	TRP	Peptide	7
1	A	131	ARG	Sidechain	2
1	A	130	ARG	Sidechain	2
1	A	149	ARG	Sidechain	1
1	A	242	TYR	Sidechain	1
1	A	127	ARG	Sidechain	1
1	A	254	MET	Peptide	1

## 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	1286	1319	1319	0±0
All	All	25720	26380	26380	2

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:211:LEU:HD22	1:A:237:SER:HB3	0.40	1.93	5	1
1:A:211:LEU:HD13	1:A:233:ILE:HG22	0.40	1.92	3	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	156/188 (83%)	154±1 (99±1%)	1±1 (0±0%)	1±0 (1±0%)	24	71
All	All	3120/3760 (83%)	3081 (99%)	15 (0%)	24 (1%)	24	71

All 2 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	280	GLU	20
1	A	251	THR	4

### 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	145/172 (84%)	139±1 (96±1%)	6±1 (4±1%)	32	81
All	All	2900/3440 (84%)	2771 (96%)	129 (4%)	32	81

All 31 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	171	GLN	20
1	A	231	THR	20
1	A	262	LEU	18
1	A	153	ASP	13
1	A	127	ARG	9
1	A	191	ASP	8
1	A	205	CYS	3
1	A	186	ARG	3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	155	THR	3
1	A	210	ARG	3
1	A	149	ARG	2
1	A	211	LEU	2
1	A	151	LYS	2
1	A	272	LEU	2
1	A	131	ARG	2
1	A	194	VAL	2
1	A	161	TRP	2
1	A	244	LYS	2
1	A	189	GLU	1
1	A	268	CYS	1
1	A	251	THR	1
1	A	252	HIS	1
1	A	185	HIS	1
1	A	257	LYS	1
1	A	258	LYS	1
1	A	157	ASP	1
1	A	290	GLU	1
1	A	178	GLU	1
1	A	230	ARG	1
1	A	130	ARG	1
1	A	227	LYS	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

The completeness of assignment taking into account all chemical shift lists is 90% for the well-defined parts and 89% 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

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	2381
Number of shifts mapped to atoms	2381
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	7

#### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	187	$0.61 \pm 0.17$	Should be applied
$^{13}\text{C}_\beta$	176	$0.90 \pm 0.09$	Should be applied
$^{13}\text{C}'$	186	$0.02 \pm 0.08$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	181	$-0.44 \pm 0.23$	None needed ( $< 0.5$ ppm)

#### 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. 1839 atoms were assigned a chemical shift out of a possible 2053. 25 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	769/770 (100%)	306/307 (100%)	312/312 (100%)	151/151 (100%)
Sidechain	952/1106 (86%)	591/652 (91%)	345/402 (86%)	16/52 (31%)

*Continued on next page...*

*Continued from previous page...*

	<b>Total</b>	<b><sup>1</sup>H</b>	<b><sup>13</sup>C</b>	<b><sup>15</sup>N</b>
Aromatic	118/177 (67%)	76/95 (80%)	39/75 (52%)	3/7 (43%)
Overall	1839/2053 (90%)	973/1054 (92%)	696/789 (88%)	170/210 (81%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 89%, i.e. 2144 atoms were assigned a chemical shift out of a possible 2404. 25 out of 25 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	<b>Total</b>	<b><sup>1</sup>H</b>	<b><sup>13</sup>C</b>	<b><sup>15</sup>N</b>
Backbone	921/928 (99%)	367/370 (99%)	373/376 (99%)	181/182 (99%)
Sidechain	1103/1285 (86%)	690/759 (91%)	395/467 (85%)	18/59 (31%)
Aromatic	120/191 (63%)	77/103 (75%)	40/79 (51%)	3/9 (33%)
Overall	2144/2404 (89%)	1134/1232 (92%)	808/922 (88%)	202/250 (81%)

#### 7.1.4 Statistically unusual chemical shifts ⓘ

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	131	ARG	NE	111.45	92.63 – 76.73	16.8
1	A	127	ARG	NE	110.86	92.63 – 76.73	16.5
1	A	230	ARG	NE	110.64	92.63 – 76.73	16.3
1	A	144	ARG	NE	108.52	92.63 – 76.73	15.0
1	A	196	GLY	N	131.68	129.07 – 90.27	5.7
1	A	132	GLY	N	129.34	129.07 – 90.27	5.1
1	A	230	ARG	HG2	0.20	2.92 – 0.22	-5.1

#### 7.1.5 Random Coil Index (RCI) plots ⓘ

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:

