WORLDWIDE PDB
PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report ⓘ

Oct 30, 2024 – 06:25 AM EDT

PDB ID	:	2M7D
BMRB ID	:	19183
Title	:	Trp-cage 16b P12W: a Hyperstable Miniprotein
Authors	:	Williams, D.V.; Andersen, N.H.; Kier, B.L.
Deposited on	:	2013-04-19

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 ⓘ symbol.

The types of validation reports are described at
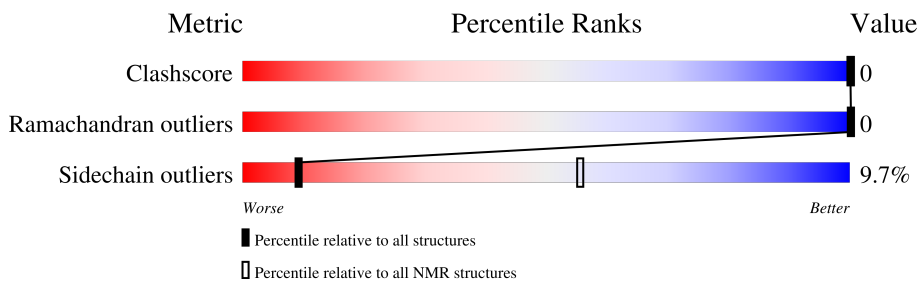http://www.wwpdb.org/validation/2017/FAQs#types.

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 52%.

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	Percentile Ranks	Value
Clashscore		0
Ramachandran outliers		0
Sidechain outliers		9.7%

Worse *Better*

▮ Percentile relative to all structures
▯ Percentile relative to all NMR structures

Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	20	75% · 5% · 20%

2 Ensemble composition and analysis ⓘ

This entry contains 30 models. Model 20 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:3-A:9, A:11-A:14, A:16-A:20 (16)	0.12	20

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 6 clusters and 1 single-model cluster was found.

Cluster number	Models
1	11, 14, 16, 17, 18, 20, 22, 23
2	7, 9, 13, 15, 19, 27, 28, 29
3	3, 12, 25, 26
4	5, 8, 10, 30
5	1, 4, 6
6	21, 24
Single-model clusters	2

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 287 atoms, of which 135 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Trp-Cage mini-protein.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	20	287	97	135	26	29	0

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: Trp-Cage mini-protein
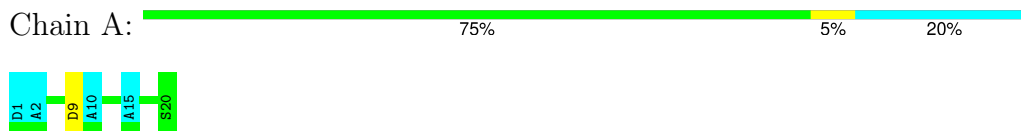
4.2 Scores per residue for each member of the ensemble
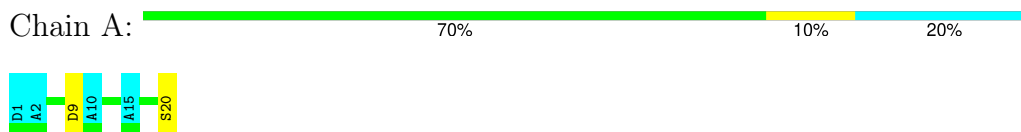
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1
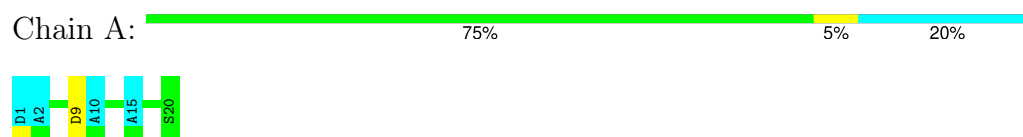
• Molecule 1: Trp-Cage mini-protein
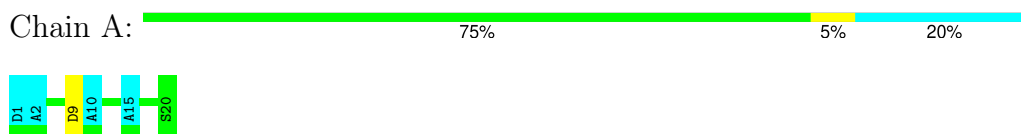
4.2.2 Score per residue for model 2

• Molecule 1: Trp-Cage mini-protein
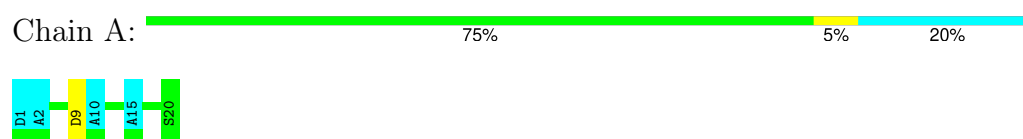
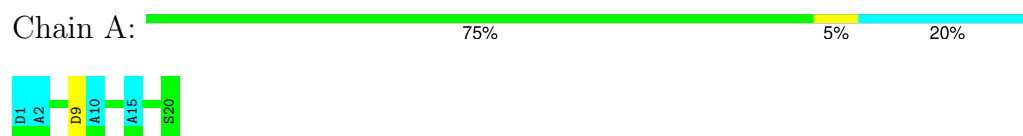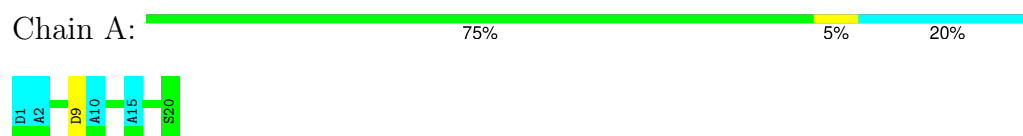### 4.2.3 Score per residue for model 3

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.5 Score per residue for model 5
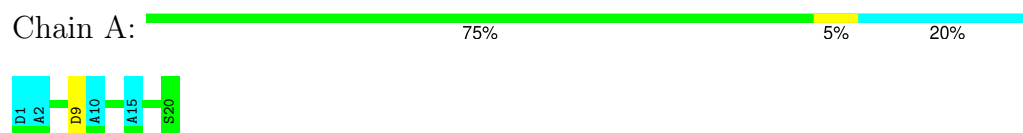
• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.6 Score per residue for model 6
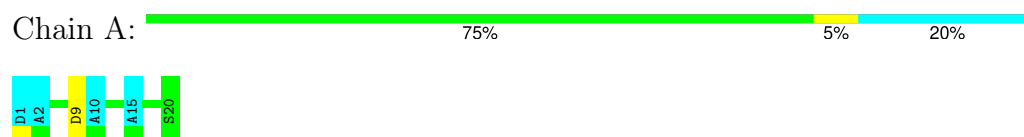
• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.7 Score per residue for model 7

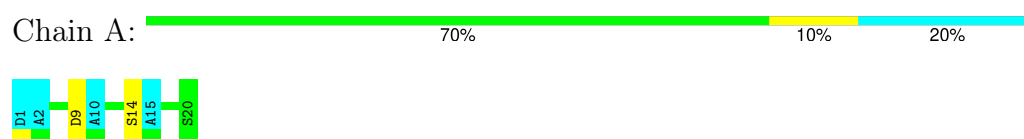• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.8   Score per residue for model 8

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.9   Score per residue for model 9

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.10   Score per residue for model 10

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.11   Score per residue for model 11

• Molecule 1: Trp-Cage mini-protein
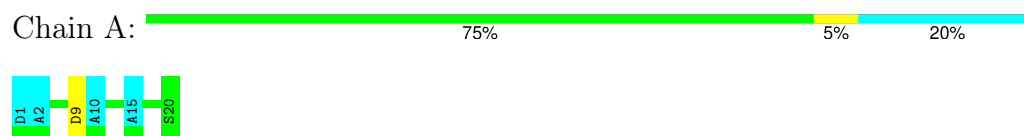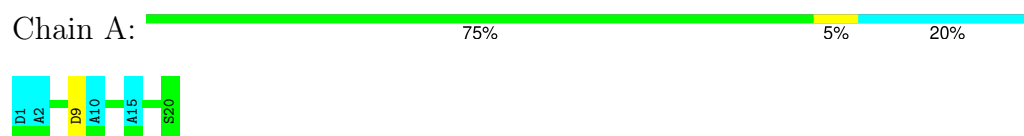
Chain A:

4.2.12   Score per residue for model 12

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.13 Score per residue for model 13

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.14 Score per residue for model 14

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.15 Score per residue for model 15

• Molecule 1: Trp-Cage mini-protein
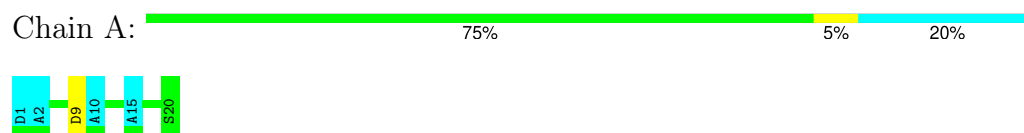
Chain A:

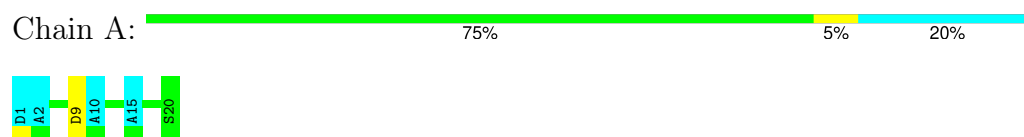### 4.2.16 Score per residue for model 16

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.17 Score per residue for model 17

• Molecule 1: Trp-Cage mini-protein

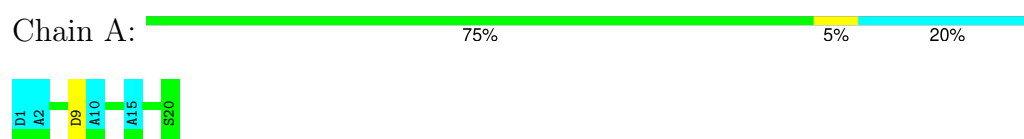Chain A:

4.2.18 Score per residue for model 18

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.19 Score per residue for model 19

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.20 Score per residue for model 20 (medoid)

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.21 Score per residue for model 21

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.22 Score per residue for model 22

• Molecule 1: Trp-Cage mini-protein

Chain A:

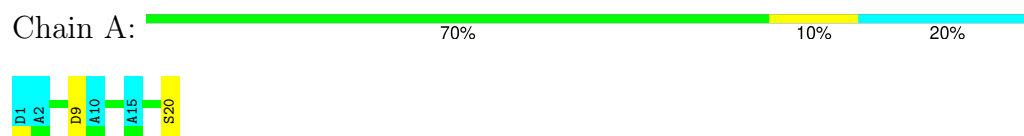#### 4.2.23 Score per residue for model 23

• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.24 Score per residue for model 24

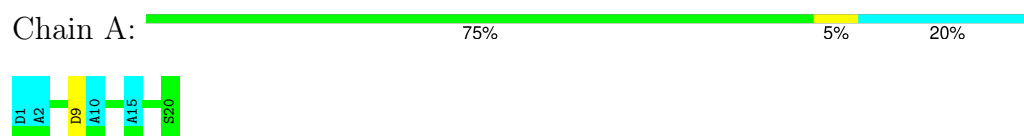• Molecule 1: Trp-Cage mini-protein

Chain A:

4.2.25 Score per residue for model 25

• Molecule 1: Trp-Cage mini-protein
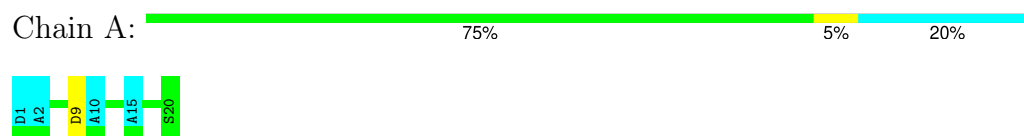
Chain A:
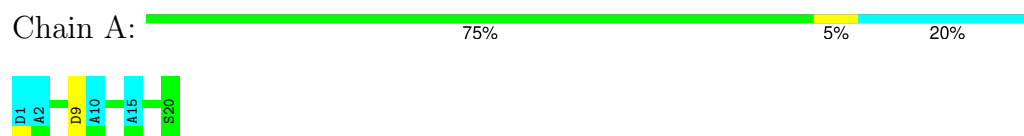
4.2.26 Score per residue for model 26

• Molecule 1: Trp-Cage mini-protein

Chain A:
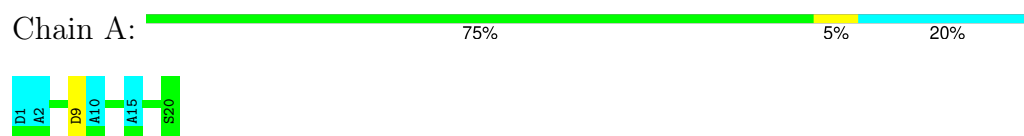
4.2.27 Score per residue for model 27

• Molecule 1: Trp-Cage mini-protein

Chain A:
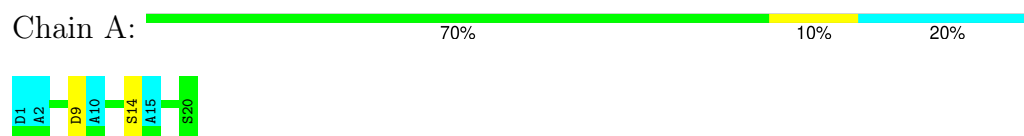
4.2.28 Score per residue for model 28

• Molecule 1: Trp-Cage mini-protein

Chain A:
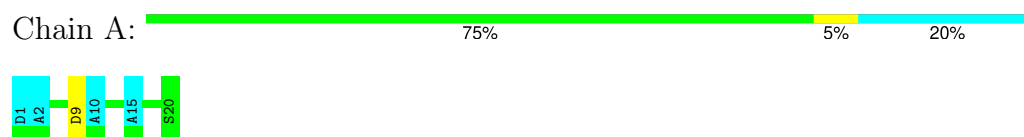
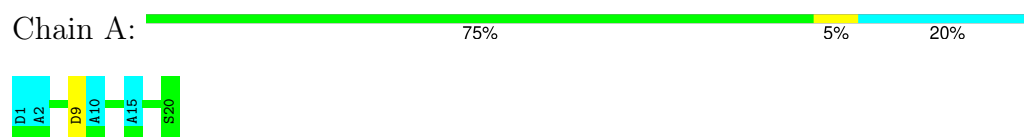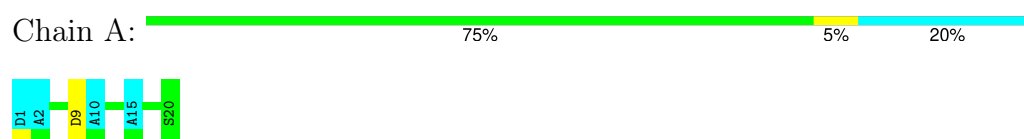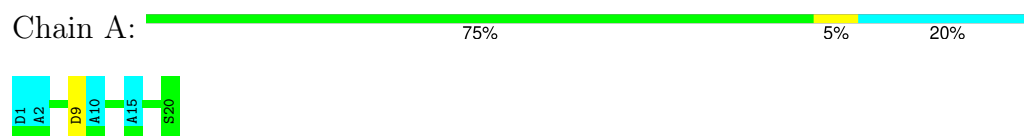### 4.2.29 Score per residue for model 29
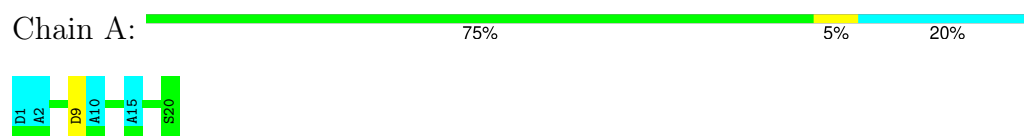
• Molecule 1: Trp-Cage mini-protein

Chain A:
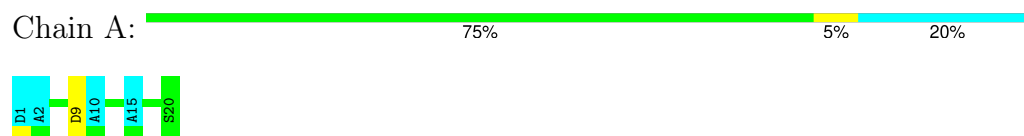
4.2.30 Score per residue for model 30
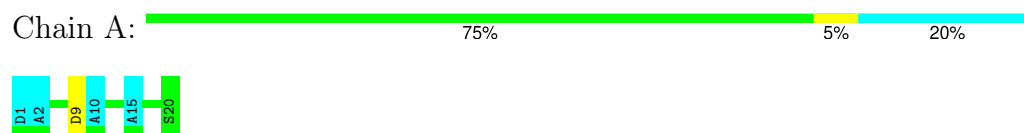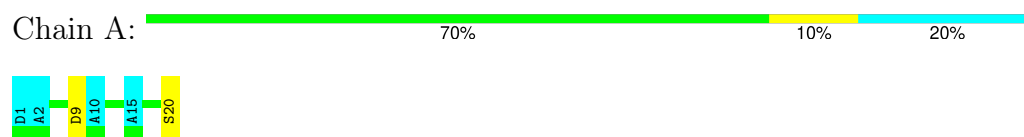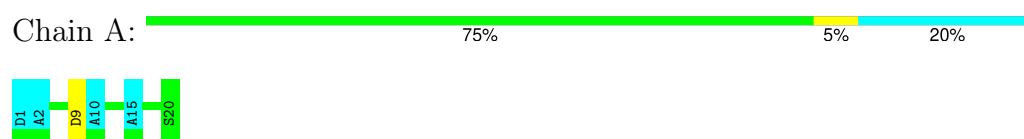
• Molecule 1: Trp-Cage mini-protein

Chain A:
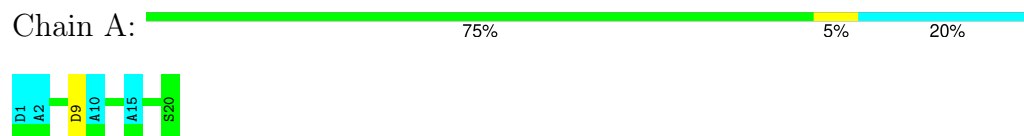
5 Refinement protocol and experimental data overview ⓘ

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

Of the 50 calculated structures, 30 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	structure solution	1.2
CNS	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	121
Number of shifts mapped to atoms	119
Number of unparsed shifts	0
Number of shifts with mapping errors	2
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	52%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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

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
All	All	3870	3420	3420	-

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 -.

There are no clashes.

6.3 Torsion angles ⓘ

6.3.1 Protein backbone ⓘ

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	15/20 (75%)	14±0 (93±0%)	1±0 (7±0%)	0±0 (0±0%)	100	100
All	All	450/600 (75%)	420 (93%)	30 (7%)	0 (0%)	100	100

There are no Ramachandran outliers.

6.3.2 Protein sidechains ⓘ

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	12/13 (92%)	11±0 (90±3%)	1±0 (10±3%)	9	55
All	All	360/390 (92%)	325 (90%)	35 (10%)	9	55

All 3 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	9	ASP	30
1	A	20	SER	3
1	A	14	SER	2

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

2 non-standard protein/DNA/RNA residues are modelled in this entry.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

6.6 Ligand geometry ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

There are no such molecules in this entry.

6.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 52% for the well-defined parts and 51% 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 ⓘ

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	121
Number of shifts mapped to atoms	119
Number of unparsed shifts	0
Number of shifts with mapping errors	2
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 2 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	10	DAL	MB	1.263	0.01	1
1	A	15	DAL	MB	1.613	0.01	1

7.1.2 Chemical shift referencing ⓘ

No chemical shift referencing corrections were calculated (not enough data).

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 52%, i.e. 107 atoms were assigned a chemical shift out of a possible 207. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	30/75 (40%)	30/30 (100%)	0/32 (0%)	0/13 (0%)
Sidechain	61/99 (62%)	61/65 (94%)	0/30 (0%)	0/4 (0%)
Aromatic	16/33 (48%)	16/16 (100%)	0/15 (0%)	0/2 (0%)
Overall	107/207 (52%)	107/111 (96%)	0/77 (0%)	0/19 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 51%, i.e. 115 atoms were assigned a chemical shift out of a possible 225. 0 out of 1 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	33/85 (39%)	33/34 (97%)	0/36 (0%)	0/15 (0%)
Sidechain	66/107 (62%)	66/70 (94%)	0/33 (0%)	0/4 (0%)
Aromatic	16/33 (48%)	16/16 (100%)	0/15 (0%)	0/2 (0%)
Overall	115/225 (51%)	115/120 (96%)	0/84 (0%)	0/21 (0%)

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.

List Id	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	11	GLY	HA2	0.31	2.15 − 5.77	-10.1
1	A	18	PRO	HA	2.03	2.78 − 6.00	-7.3
1	A	18	PRO	HB3	-0.42	0.25 − 3.76	-6.9

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.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: