WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report ⓘ

Apr 20, 2024 – 08:08 AM EDT

PDB ID	:	5KVP
BMRB ID	:	30139
Title	:	Solution structure of the catalytic domain of zoocin A
Authors	:	Timkovich, R.; Xing, M.; Simmonds, R.S.
Deposited on	:	2016-07-15

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

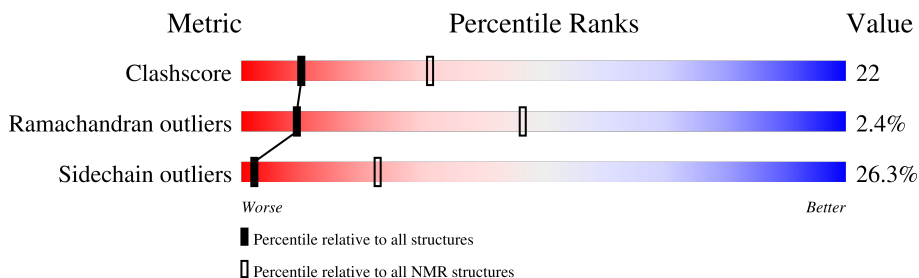|		
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 87%.

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		22
Ramachandran outliers		2.4%
Sidechain outliers		26.3%

Worse *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all NMR structures

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	159	40% 32% 8% 21%

2   Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:18-A:113, A:118-A:147 (126)	0.38	4

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 1 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 9, 10
Single-model clusters	8

3 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 2320 atoms, of which 1128 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Zoocin A endopeptidase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	159	2318	746	1128	219	220	5	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP O54308
A	2	ARG	-	expression tag	UNP O54308
A	3	GLY	-	expression tag	UNP O54308
A	4	SER	-	expression tag	UNP O54308
A	5	HIS	-	expression tag	UNP O54308
A	6	HIS	-	expression tag	UNP O54308
A	7	HIS	-	expression tag	UNP O54308
A	8	HIS	-	expression tag	UNP O54308
A	9	HIS	-	expression tag	UNP O54308
A	10	HIS	-	expression tag	UNP O54308
A	11	GLY	-	expression tag	UNP O54308
A	12	SER	-	expression tag	UNP O54308
A	74	ALA	CYS	engineered mutation	UNP O54308

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

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	1	1

- Molecule 3 is UNKNOWN LIGAND (three-letter code: UNL) (formula:).

Mol	Chain	Residues	Atoms	
			Total	O
3	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.
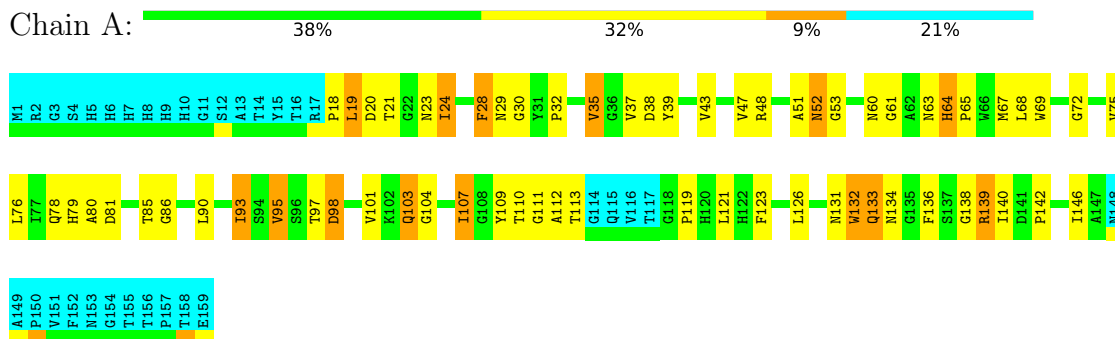
• Molecule 1: Zoocin A endopeptidase

4.2 Scores per residue for each member of the ensemble
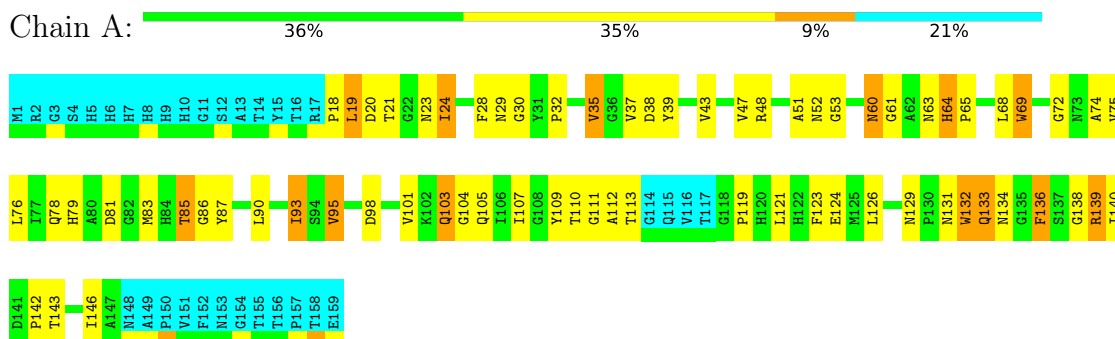
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Zoocin A endopeptidase

4.2.2 Score per residue for model 2

• Molecule 1: Zoocin A endopeptidase

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: Zoocin A endopeptidase

Chain A:

4.2.4 Score per residue for model 4 (medoid)

• Molecule 1: Zoocin A endopeptidase

Chain A:

4.2.5 Score per residue for model 5

• Molecule 1: Zoocin A endopeptidase

4.2.6 Score per residue for model 6

• Molecule 1: Zoocin A endopeptidase
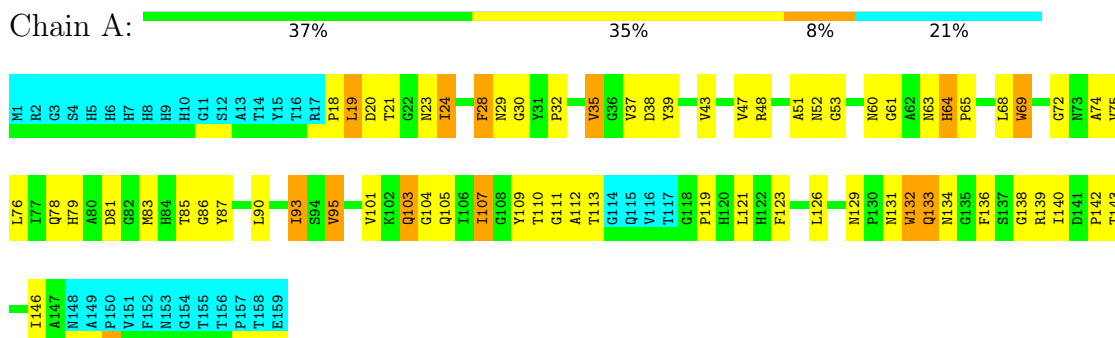

4.2.7 Score per residue for model 7

• Molecule 1: Zoocin A endopeptidase

4.2.8　Score per residue for model 8

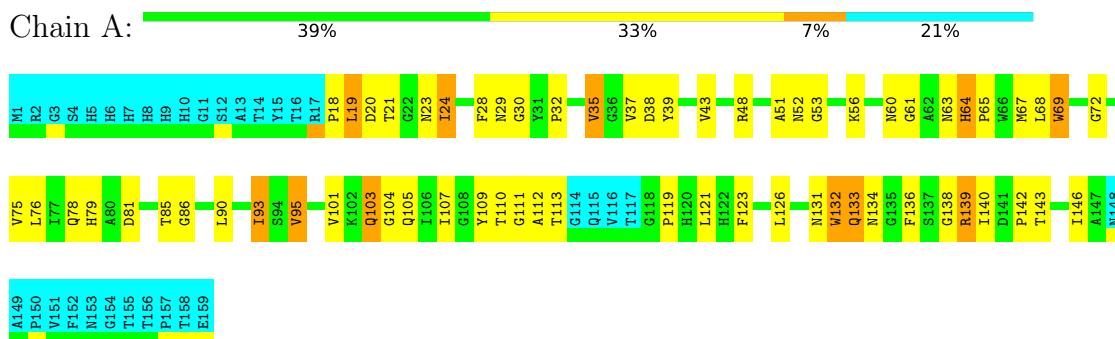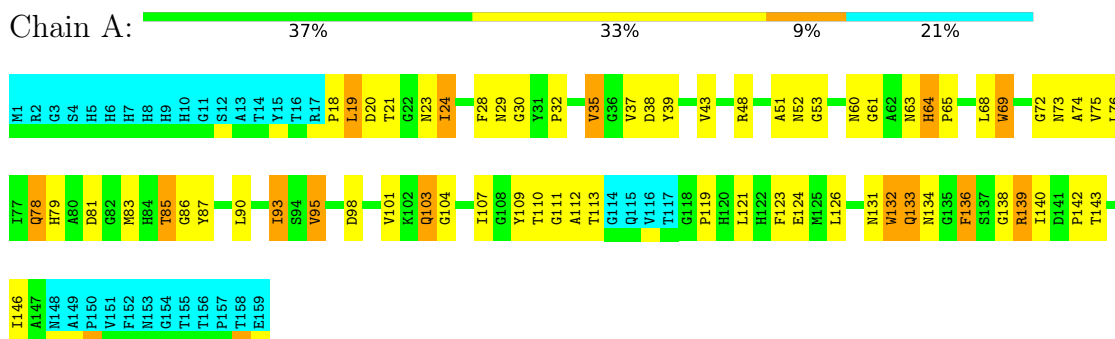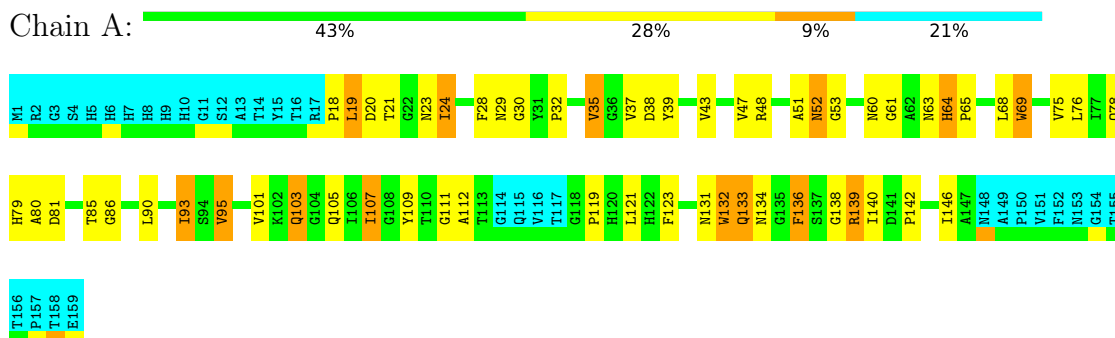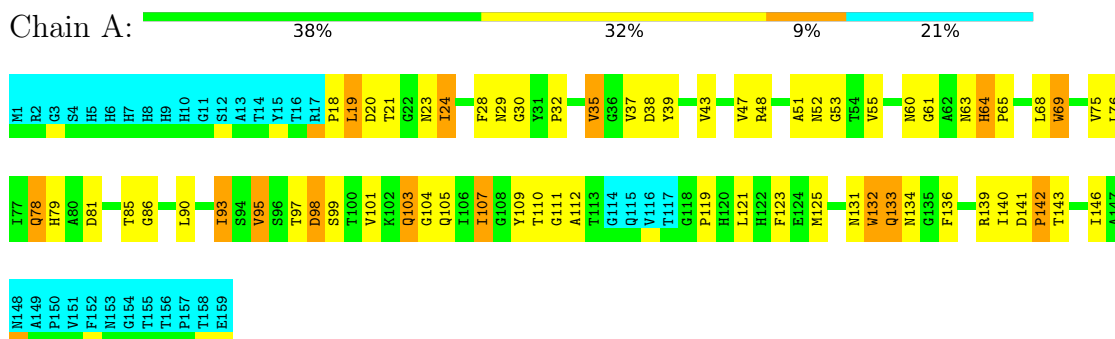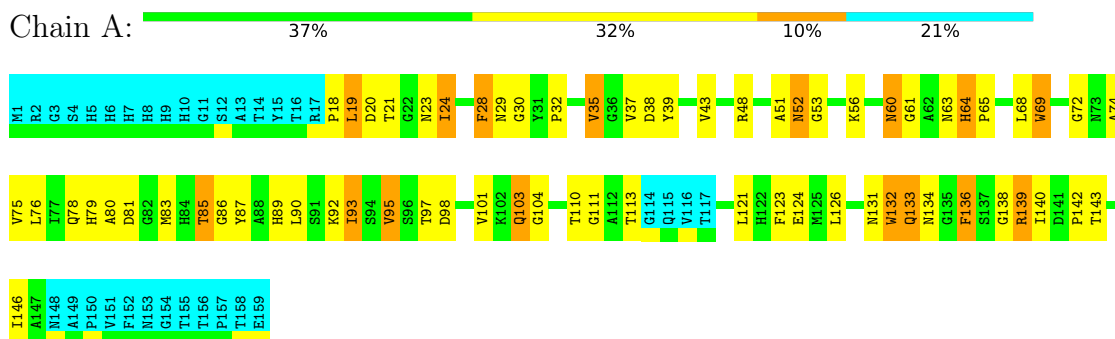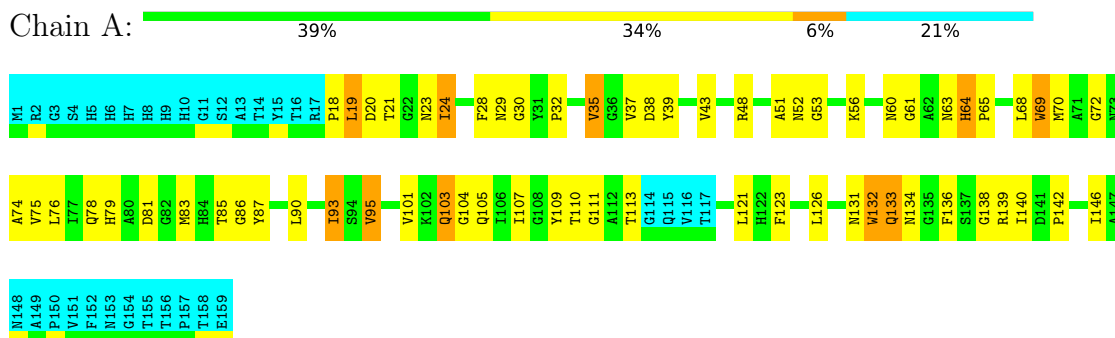• Molecule 1: Zoocin A endopeptidase

Chain A:

4.2.9　Score per residue for model 9

• Molecule 1: Zoocin A endopeptidase

Chain A:

4.2.10　Score per residue for model 10

• Molecule 1: Zoocin A endopeptidase

Chain A:

5 Refinement protocol and experimental data overview ⓘ

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

Of the 100 calculated structures, 10 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	1.3
CNS	structure solution	1.3
NMRView	structure solution	
NMRPipe	structure solution	
NMRDraw	structure solution	
TALOS+	structure solution	
TopSpin	structure solution	1.2
DEEPVIEW/SWISS-PDB VIEWER	structure solution	
PROCHECKNMR	structure solution	

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	1631
Number of shifts mapped to atoms	1631
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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

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	934	893	888	40±3
All	All	9360	8930	8880	400

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 22.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:65:PRO:HD2	1:A:69:TRP:NE1	0.73	1.98	2	10
1:A:19:LEU:HD12	1:A:21:THR:O	0.68	1.89	10	10
1:A:38:ASP:HA	1:A:121:LEU:O	0.68	1.89	9	10
1:A:103:GLN:HE21	1:A:103:GLN:C	0.66	1.94	10	10
1:A:43:VAL:HG23	1:A:111:GLY:HA2	0.62	1.69	7	10
1:A:83:MET:HB2	1:A:126:LEU:O	0.62	1.93	8	2
1:A:24:ILE:HD12	1:A:24:ILE:H	0.61	1.56	10	10
1:A:93:ILE:HG23	1:A:95:VAL:H	0.61	1.56	2	10
1:A:53:GLY:HA3	1:A:78:GLN:O	0.60	1.96	8	10
1:A:65:PRO:O	1:A:69:TRP:CD1	0.59	2.55	2	10
1:A:30:GLY:O	1:A:32:PRO:HD3	0.59	1.98	2	10
1:A:18:PRO:HB2	1:A:39:TYR:HB3	0.58	1.74	7	10

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:125:MET:SD	1:A:142:PRO:HA	0.56	2.41	7	1
1:A:51:ALA:HB3	1:A:79:HIS:ND1	0.56	2.15	8	10
1:A:75:VAL:O	1:A:86:GLY:HA2	0.55	2.02	1	10
1:A:61:GLY:HA2	1:A:64:HIS:CD2	0.55	2.37	3	7
1:A:48:ARG:HG2	1:A:104:GLY:O	0.54	2.03	7	8
1:A:126:LEU:HA	1:A:138:GLY:O	0.54	2.01	8	8
1:A:142:PRO:O	1:A:146:ILE:HG12	0.54	2.03	2	10
1:A:19:LEU:HD12	1:A:21:THR:H	0.53	1.63	8	10
1:A:112:ALA:HB2	1:A:119:PRO:HA	0.53	1.79	2	8
1:A:35:VAL:HG12	1:A:139:ARG:HG2	0.52	1.82	5	10
1:A:61:GLY:HA2	1:A:64:HIS:CG	0.52	2.40	7	9
1:A:72:GLY:HA3	1:A:113:THR:OG1	0.52	2.04	1	7
1:A:51:ALA:HB3	1:A:79:HIS:CE1	0.52	2.39	5	8
1:A:52:ASN:ND2	1:A:80:ALA:HB2	0.52	2.20	10	4
1:A:19:LEU:HB2	1:A:39:TYR:CD2	0.51	2.40	2	8
1:A:19:LEU:HB2	1:A:39:TYR:CD1	0.49	2.42	3	2
1:A:131:ASN:C	1:A:133:GLN:H	0.49	2.11	6	10
1:A:136:PHE:CE1	1:A:138:GLY:HA2	0.49	2.43	5	1
1:A:136:PHE:O	1:A:139:ARG:HG3	0.49	2.08	2	3
1:A:19:LEU:CD1	1:A:21:THR:O	0.49	2.61	10	9
1:A:133:GLN:N	1:A:133:GLN:NE2	0.49	2.61	2	10
1:A:60:ASN:HD22	1:A:60:ASN:N	0.48	2.06	10	10
1:A:37:VAL:HG11	1:A:142:PRO:HG3	0.48	1.86	6	1
1:A:139:ARG:HD2	1:A:139:ARG:O	0.48	2.08	7	5
1:A:97:THR:O	1:A:98:ASP:HB2	0.47	2.09	8	3
1:A:47:VAL:O	1:A:107:ILE:HG13	0.47	2.10	7	4
1:A:74:ALA:HA	1:A:87:TYR:O	0.47	2.10	2	5
1:A:90:LEU:HB3	1:A:109:TYR:O	0.47	2.10	6	9
1:A:68:LEU:HG	1:A:69:TRP:CD1	0.46	2.46	3	10
1:A:85:THR:HA	1:A:124:GLU:O	0.46	2.10	2	3
1:A:60:ASN:O	1:A:64:HIS:HB3	0.46	2.10	1	8
1:A:37:VAL:HG12	1:A:123:PHE:O	0.46	2.11	1	10
1:A:132:TRP:N	1:A:133:GLN:NE2	0.45	2.64	9	10
1:A:24:ILE:HG13	1:A:28:PHE:CD2	0.45	2.47	10	1
1:A:68:LEU:HG	1:A:69:TRP:NE1	0.45	2.27	7	10
1:A:136:PHE:CE2	1:A:138:GLY:HA2	0.45	2.47	6	1
1:A:69:TRP:CD1	1:A:69:TRP:N	0.44	2.85	4	9
1:A:51:ALA:HB3	1:A:79:HIS:CG	0.44	2.46	10	1
1:A:48:ARG:HA	1:A:105:GLN:O	0.44	2.12	7	7
1:A:24:ILE:HD12	1:A:24:ILE:N	0.43	2.26	10	2
1:A:125:MET:SD	1:A:141:ASP:O	0.43	2.77	7	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:90:LEU:HA	1:A:111:GLY:N	0.42	2.29	6	3
1:A:24:ILE:HG13	1:A:28:PHE:CE1	0.42	2.49	8	2
1:A:78:GLN:HA	1:A:83:MET:O	0.42	2.14	3	3
1:A:28:PHE:O	1:A:30:GLY:N	0.41	2.50	1	1
1:A:47:VAL:HG21	1:A:90:LEU:HD22	0.41	1.93	2	2
1:A:24:ILE:HG12	1:A:34:HIS:O	0.41	2.15	10	1
1:A:65:PRO:HD2	1:A:69:TRP:CD1	0.41	2.50	5	1
1:A:61:GLY:HA3	1:A:73:ASN:HD21	0.41	1.75	5	1
1:A:55:VAL:HG23	1:A:99:SER:O	0.41	2.15	7	1
1:A:52:ASN:HD21	1:A:80:ALA:HB2	0.41	1.75	10	2
1:A:65:PRO:O	1:A:69:TRP:HB2	0.41	2.15	3	2
1:A:37:VAL:HG13	1:A:39:TYR:CE1	0.41	2.51	6	2
1:A:24:ILE:HG13	1:A:28:PHE:CE2	0.41	2.50	10	1
1:A:61:GLY:HA2	1:A:64:HIS:ND1	0.40	2.31	2	1
1:A:103:GLN:C	1:A:103:GLN:NE2	0.40	2.72	8	1
1:A:24:ILE:HG13	1:A:28:PHE:CD1	0.40	2.52	8	1

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	126/159 (79%)	100±0 (79±0%)	23±0 (19±0%)	3±0 (2±0%)	9	46
All	All	1260/1590 (79%)	996 (79%)	234 (19%)	30 (2%)	9	46

All 3 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	29	ASN	10
1	A	132	TRP	10
1	A	134	ASN	10

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	95/122 (78%)	70±1 (74±1%)	25±1 (26±1%)	2	22
All	All	950/1220 (78%)	700 (74%)	250 (26%)	2	22

All 33 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	19	LEU	10
1	A	20	ASP	10
1	A	23	ASN	10
1	A	24	ILE	10
1	A	28	PHE	10
1	A	35	VAL	10
1	A	52	ASN	10
1	A	63	ASN	10
1	A	64	HIS	10
1	A	76	LEU	10
1	A	81	ASP	10
1	A	85	THR	10
1	A	93	ILE	10
1	A	95	VAL	10
1	A	101	VAL	10
1	A	103	GLN	10
1	A	133	GLN	10
1	A	136	PHE	10
1	A	140	ILE	10
1	A	107	ILE	9
1	A	110	THR	9
1	A	69	TRP	9
1	A	139	ARG	7
1	A	143	THR	6
1	A	98	ASP	5
1	A	60	ASN	3
1	A	56	LYS	3
1	A	67	MET	2
1	A	129	ASN	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	78	GLN	2
1	A	142	PRO	1
1	A	92	LYS	1
1	A	70	MET	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 1 is monoatomic and 1 is unknown - leaving 0 for Mogul analysis.

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 87% for the well-defined parts and 81% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *c74achemshifts_v2.txt*

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	1631
Number of shifts mapped to atoms	1631
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	151	0.14 ± 0.15	None needed (< 0.5 ppm)
$^{13}C_\beta$	127	-0.10 ± 0.07	None needed (< 0.5 ppm)
$^{13}C'$	150	0.18 ± 0.24	None needed (< 0.5 ppm)
^{15}N	137	0.34 ± 0.44	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 87%, i.e. 1380 atoms were assigned a chemical shift out of a possible 1587. 0 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	628/631 (100%)	261/262 (100%)	251/252 (100%)	116/117 (99%)
Sidechain	712/781 (91%)	466/517 (90%)	231/242 (95%)	15/22 (68%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	40/175 (23%)	36/86 (42%)	0/74 (0%)	4/15 (27%)
Overall	1380/1587 (87%)	763/865 (88%)	482/568 (85%)	135/154 (88%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 81%, i.e. 1631 atoms were assigned a chemical shift out of a possible 2008. 0 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	747/796 (94%)	309/330 (94%)	301/318 (95%)	137/148 (93%)
Sidechain	842/970 (87%)	549/640 (86%)	274/299 (92%)	19/31 (61%)
Aromatic	42/242 (17%)	38/119 (32%)	0/96 (0%)	4/27 (15%)
Overall	1631/2008 (81%)	896/1089 (82%)	575/713 (81%)	160/206 (78%)

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	130	PRO	HB2	-0.75	0.37 − 3.78	-8.3
1	A	64	HIS	HB2	0.71	1.36 − 4.85	-6.9
1	A	38	ASP	HA	6.67	3.04 − 6.12	6.8
1	A	65	PRO	HD2	1.85	1.93 − 5.38	-5.2
1	A	104	GLY	HA3	2.05	2.08 − 5.71	-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.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: