W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report ⓘ

May 29, 2020 – 12:21 am BST

PDB ID	:	2N6J
Title	:	Solution structure of Zmp1, a zinc-dependent metalloprotease secreted by Clostridium difficile
Authors	:	Banci, L.; Cantini, F.; Scarselli, M.; Rubino, J.T.; Martinelli, M.
Deposited on	:	2015-08-24

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

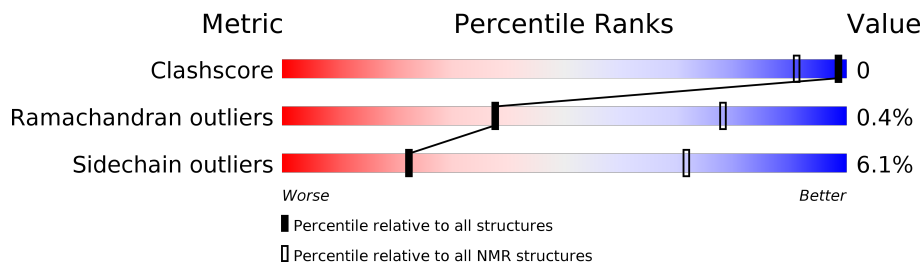|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 91%.

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		0.4%
Sidechain outliers		6.1%

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	195	91% · 5%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 13 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:28-A:100, A:109-A:220 (185)	0.55	13

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 4 clusters and 6 single-model clusters were found.

Cluster number	Models
1	1, 7, 18, 30
2	5, 8, 14, 17
3	6, 13, 29
4	11, 34, 40
Single-model clusters	2; 23; 24; 36; 38; 28

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3040 atoms, of which 1511 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Zinc metalloprotease Zmp1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	195	3036	969	1509	259	297	2	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	GLY	-	EXPRESSION TAG	UNP Q183R7

- 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 water.

Mol	Chain	Residues	Atoms		
			Total	H	O
3	A	1	3	2	1

4 Residue-property plots ⓘ

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.

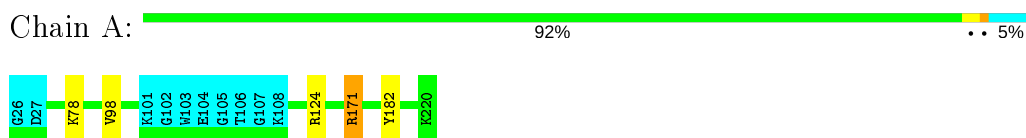• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

91%   • 5%

G26 D27 | R60 | K78 | V98 | K101 G102 W103 E104 G105 T106 G107 K108 | R124 | R171 | L204 | K220
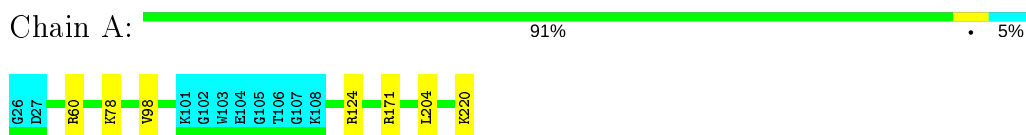
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Zinc metalloprotease Zmp1
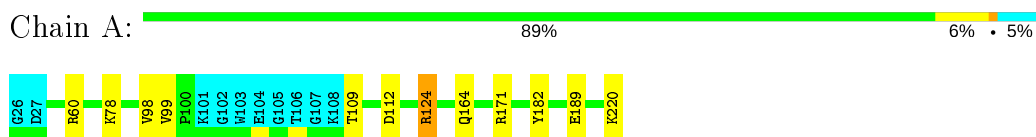
Chain A:

92%   •• 5%

G26 D27 | K78 | V98 | K101 G102 W103 E104 G105 T106 G107 K108 | R124 | R171 | Y182 | K220

4.2.2 Score per residue for model 2

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

89%   6%   5%

G26 D27 | R60 | S83 | T87 | Y92 | K96 | K101 G102 W103 E104 G105 T106 G107 K108 | T109 | R124 | F127 | E140 | R171 | E185 | N197 | K220
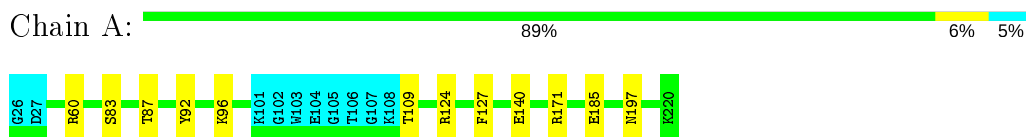
4.2.3 Score per residue for model 5

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

89%   6%   · 5%

G26 D27 R60 K78 V98 V99 P100 K101 G102 W103 E104 G105 T106 G107 K108 T109 D112 R124 Q164 R171 Y182 E189 K220

4.2.4 Score per residue for model 6

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

89%   6%   · 5%

G26 D27 R60 N76 L77 K78 Y92 V98 K101 G102 W103 E104 G105 T106 G107 K108 R124 K168 R171 Y178 L204 K220
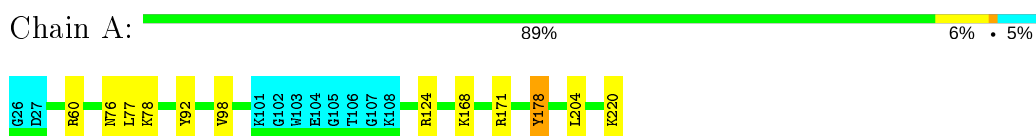
4.2.5 Score per residue for model 7

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

88%   7%   5%

G26 D27 G47 Y92 V98 K101 G102 W103 E104 G105 T106 G107 K108 R124 E140 H150 K158 R171 Y182 N197 N201 L204 K220

4.2.6 Score per residue for model 8

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

87%   7%   · 5%

G26 D27 Q40 T46 G47 N48 I64 V98 K101 G102 W103 E104 G105 T106 G107 K108 L116 G117 G118 R124 K130 K163 L173 N177 Y182 L204 Y212 K220

4.2.7 Score per residue for model 11

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

89%   5%   · 5%

G26 D27 D50 R60 V98 K101 G102 W103 E104 G105 T106 G107 K108 R171 E184 E189 N197 N201 S202 K203 L204 K220

4.2.8 Score per residue for model 13 (medoid)

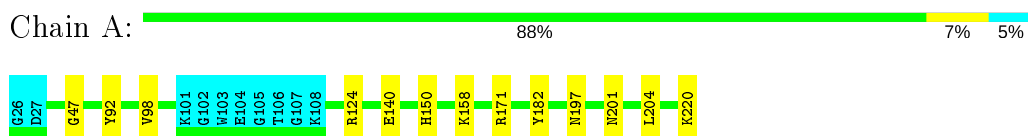• Molecule 1: Zinc metalloprotease Zmp1

4.2.9 Score per residue for model 14

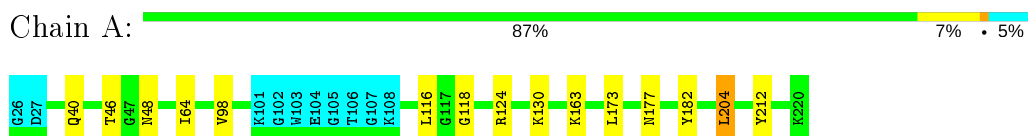• Molecule 1: Zinc metalloprotease Zmp1

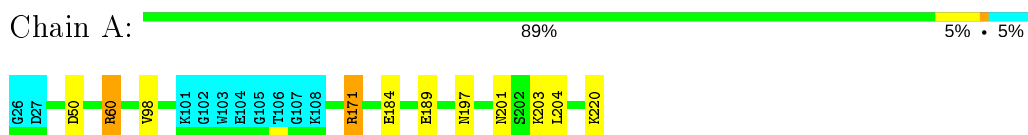#### 4.2.10 Score per residue for model 17

• Molecule 1: Zinc metalloprotease Zmp1

4.2.11 Score per residue for model 18
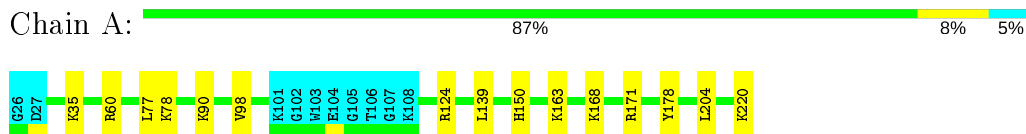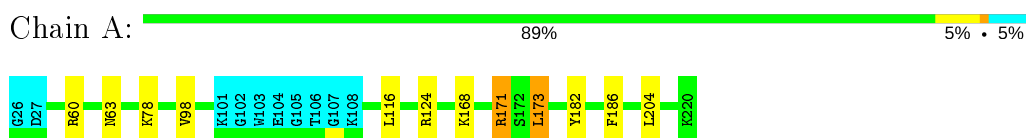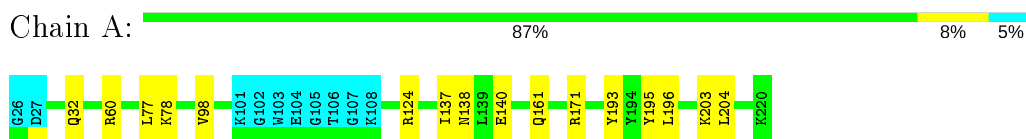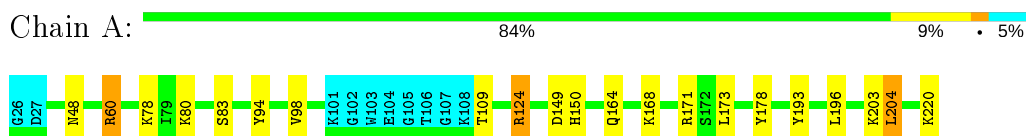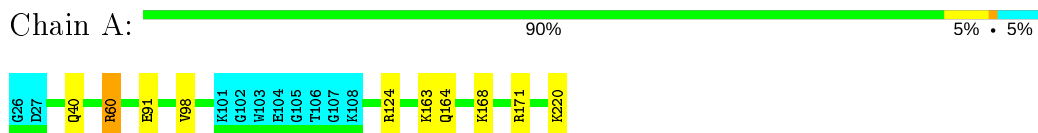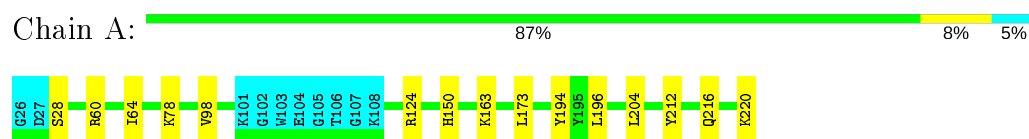
• Molecule 1: Zinc metalloprotease Zmp1

4.2.12 Score per residue for model 23

• Molecule 1: Zinc metalloprotease Zmp1

4.2.13 Score per residue for model 24

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

4.2.14 Score per residue for model 29

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

4.2.15 Score per residue for model 30

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

4.2.16 Score per residue for model 34

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:
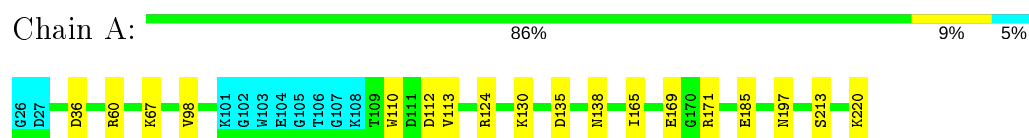
4.2.17 Score per residue for model 36

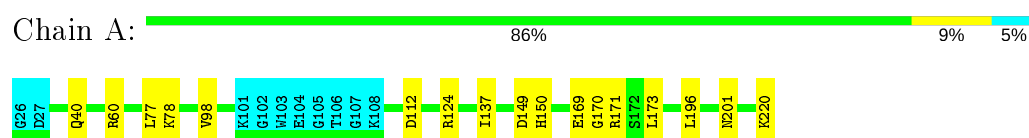• Molecule 1: Zinc metalloprotease Zmp1

Chain A:
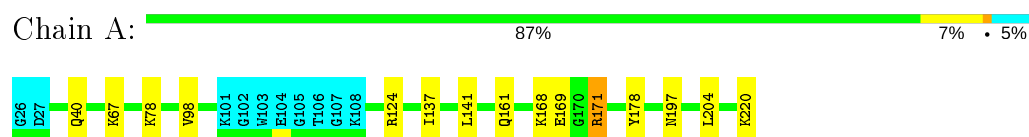
4.2.18 Score per residue for model 38

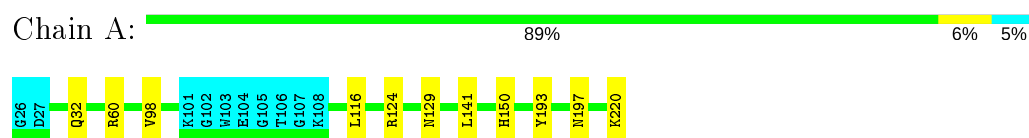• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

88% 6% • 5%

G26 D27 S28 Q40 R60 D65 K73 V98 K101 G102 W103 E104 G105 T106 G107 K108 S119 R124 F127 R171 S172 L173 Y178 K220
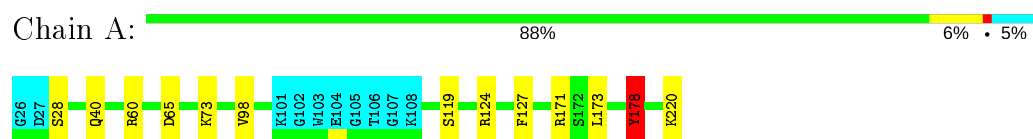
4.2.19 Score per residue for model 40

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

87% 8% 5%

G26 D27 Q40 V43 R60 N63 K90 V98 K101 G102 W103 E104 G105 T106 G107 K108 L116 S119 R124 D149 H150 Q161 L173 L196 K220
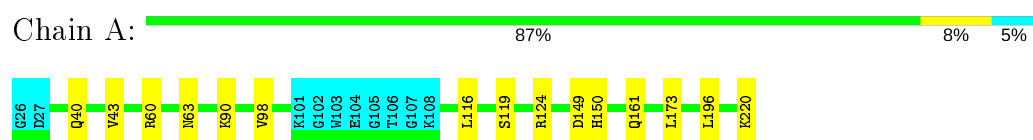
4.2.20 Score per residue for model 28

• Molecule 1: Zinc metalloprotease Zmp1

Chain A:

88% 7% 5%

G26 D27 D36 N48 R60 N63 K78 V98 K101 G102 W103 E104 G105 T106 G107 K108 R124 G170 R171 S172 L173 Y178 L204 K220
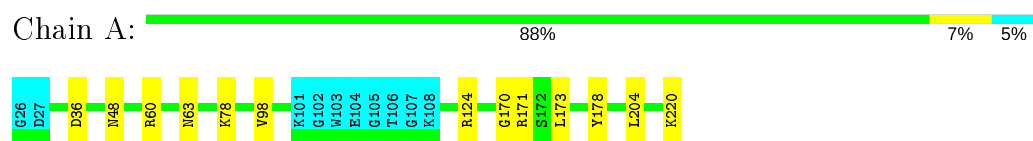
5 Refinement protocol and experimental data overview (i)

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

Of the 400 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CYANA	structure solution	2.1
CYANA	refinement	2.1
PSVS	structure solution	
PSVS	refinement	
AMBER	refinement	10
CARA	refinement	2

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	2373
Number of shifts mapped to atoms	2373
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 Too-close contacts (i)

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	1455	1442	1441	0±0
All	All	29140	28880	28820	5

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	Models Total
1:A:173:LEU:HD23	1:A:186:PHE:CD1	0.51	2.40	14	1
1:A:193:TYR:CG	1:A:204:LEU:HD23	0.45	2.46	18	1
1:A:110:TRP:HA	1:A:113:VAL:HG13	0.44	1.88	34	1
1:A:173:LEU:HD13	1:A:204:LEU:HD21	0.43	1.90	8	1
1:A:194:TYR:CZ	1:A:212:TYR:CD2	0.41	3.09	24	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	184/195 (94%)	174±3 (94±1%)	9±3 (5±2%)	1±1 (0±1%)	38	78
All	All	3680/3900 (94%)	3476 (94%)	189 (5%)	15 (0%)	38	78

All 10 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	130	LYS	2
1	A	64	ILE	2
1	A	127	PHE	2
1	A	170	GLY	2
1	A	28	SER	2
1	A	47	GLY	1
1	A	129	ASN	1
1	A	65	ASP	1
1	A	118	GLY	1
1	A	48	ASN	1

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	157/163 (96%)	147±3 (94±2%)	10±3 (6±2%)	22	71
All	All	3140/3260 (96%)	2949 (94%)	191 (6%)	22	71

All 59 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	98	VAL	19
1	A	220	LYS	15
1	A	78	LYS	11
1	A	204	LEU	11
1	A	150	HIS	7
1	A	40	GLN	6
1	A	197	ASN	6
1	A	173	LEU	5
1	A	168	LYS	5
1	A	196	LEU	5
1	A	182	TYR	4
1	A	124	ARG	4
1	A	116	LEU	4
1	A	77	LEU	4
1	A	163	LYS	4
1	A	109	THR	3
1	A	112	ASP	3
1	A	203	LYS	3
1	A	60	ARG	3
1	A	161	GLN	3
1	A	140	GLU	3
1	A	169	GLU	3
1	A	171	ARG	3
1	A	63	ASN	3
1	A	178	TYR	3
1	A	149	ASP	3
1	A	201	ASN	3
1	A	164	GLN	3
1	A	137	ILE	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	138	ASN	2
1	A	36	ASP	2
1	A	67	LYS	2
1	A	90	LYS	2
1	A	83	SER	2
1	A	185	GLU	2
1	A	189	GLU	2
1	A	32	GLN	2
1	A	48	ASN	2
1	A	141	LEU	2
1	A	73	LYS	1
1	A	96	LYS	1
1	A	216	GLN	1
1	A	165	ILE	1
1	A	94	TYR	1
1	A	139	LEU	1
1	A	35	LYS	1
1	A	135	ASP	1
1	A	46	THR	1
1	A	80	LYS	1
1	A	184	GLU	1
1	A	158	LYS	1
1	A	43	VAL	1
1	A	91	GLU	1
1	A	87	THR	1
1	A	99	VAL	1
1	A	50	ASP	1
1	A	177	ASN	1
1	A	213	SER	1
1	A	76	ASN	1

5.2.3 RNA (i)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains (i)

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

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

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

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 91% for the well-defined parts and 90% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

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	2373
Number of shifts mapped to atoms	2373
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	12

6.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	192	-0.65 ± 0.17	Should be applied
$^{13}C_\beta$	176	0.38 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	154	-0.28 ± 0.17	None needed (< 0.5 ppm)
^{15}N	184	0.59 ± 0.34	None needed (imprecise)

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 91%, i.e. 2045 atoms were assigned a chemical shift out of a possible 2259. 29 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	869/915 (95%)	361/365 (99%)	331/370 (89%)	177/180 (98%)
Sidechain	1022/1148 (89%)	628/669 (94%)	367/427 (86%)	27/52 (52%)

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	154/196 (79%)	81/102 (79%)	64/85 (75%)	9/9 (100%)
Overall	2045/2259 (91%)	1070/1136 (94%)	762/882 (86%)	213/241 (88%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 90%, i.e. 2124 atoms were assigned a chemical shift out of a possible 2365. 29 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	905/965 (94%)	375/385 (97%)	346/390 (89%)	184/190 (97%)
Sidechain	1055/1192 (89%)	649/695 (93%)	379/443 (86%)	27/54 (50%)
Aromatic	164/208 (79%)	86/108 (80%)	68/90 (76%)	10/10 (100%)
Overall	2124/2365 (90%)	1110/1188 (93%)	793/923 (86%)	221/254 (87%)

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	124	ARG	NE	137.43	92.63 − 76.73	33.2
1	A	60	ARG	NE	135.95	92.63 − 76.73	32.2
1	A	100	PRO	HG3	-0.63	3.56 − 0.26	-7.7
1	A	95	LEU	HG	-0.56	3.16 − -0.14	-6.3
1	A	124	ARG	HD3	1.64	4.36 − 1.86	-5.9
1	A	194	TYR	HD2	5.18	8.44 − 5.44	-5.9
1	A	194	TYR	HD1	5.18	8.44 − 5.44	-5.9
1	A	100	PRO	HG2	0.12	3.48 − 0.38	-5.8
1	A	65	ASP	H	5.02	11.17 − 5.47	-5.8
1	A	170	GLY	HA2	1.88	5.87 − 2.07	-5.5
1	A	136	ALA	H	11.40	11.19 − 5.19	5.4
1	A	210	GLN	H	11.29	11.17 − 5.27	5.2

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: