



# Full wwPDB NMR Structure Validation Report ⓘ

May 28, 2020 – 11:39 pm BST

PDB ID : 2MAX  
Title : NMR structure of the RNA polymerase alpha subunit C-terminal domain from *Helicobacter pylori*  
Authors : Borin, B.N.; Krezel, A.M.  
Deposited on : 2013-07-21

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:

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

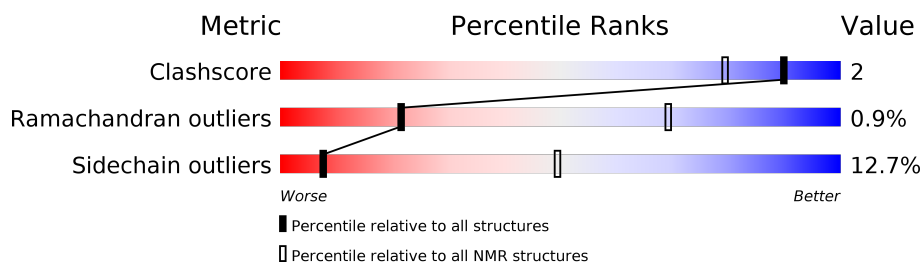
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 83%.

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	126	

## 2 Ensemble composition and analysis i

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:255-A:318, A:324-A:337 (78)	0.23	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. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14
2	11, 15

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1778 atoms, of which 887 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		S
1	A	114	1778	552	887	150	184	5	0

There are 12 discrepancies between the modelled and reference sequences:

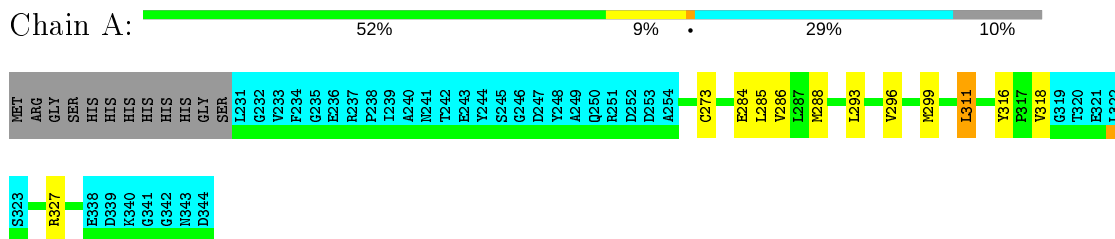
Chain	Residue	Modelled	Actual	Comment	Reference
A	219	MET	-	EXPRESSION TAG	UNP Q9ZJT5
A	220	ARG	-	EXPRESSION TAG	UNP Q9ZJT5
A	221	GLY	-	EXPRESSION TAG	UNP Q9ZJT5
A	222	SER	-	EXPRESSION TAG	UNP Q9ZJT5
A	223	HIS	-	EXPRESSION TAG	UNP Q9ZJT5
A	224	HIS	-	EXPRESSION TAG	UNP Q9ZJT5
A	225	HIS	-	EXPRESSION TAG	UNP Q9ZJT5
A	226	HIS	-	EXPRESSION TAG	UNP Q9ZJT5
A	227	HIS	-	EXPRESSION TAG	UNP Q9ZJT5
A	228	HIS	-	EXPRESSION TAG	UNP Q9ZJT5
A	229	GLY	-	EXPRESSION TAG	UNP Q9ZJT5
A	230	SER	-	EXPRESSION TAG	UNP Q9ZJT5

## 4 Residue-property plots [\(i\)](#)

### 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: DNA-directed RNA polymerase subunit alpha

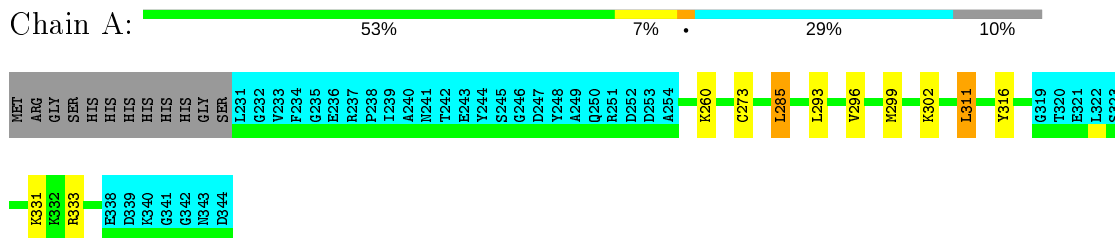


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

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

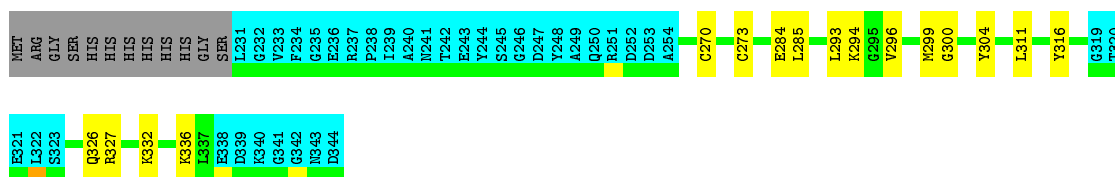
- Molecule 1: DNA-directed RNA polymerase subunit alpha



#### 4.2.2 Score per residue for model 2

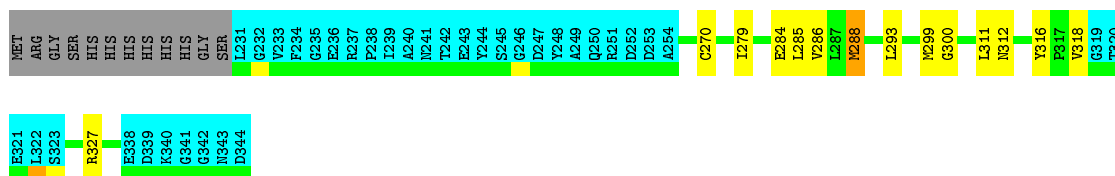
- Molecule 1: DNA-directed RNA polymerase subunit alpha





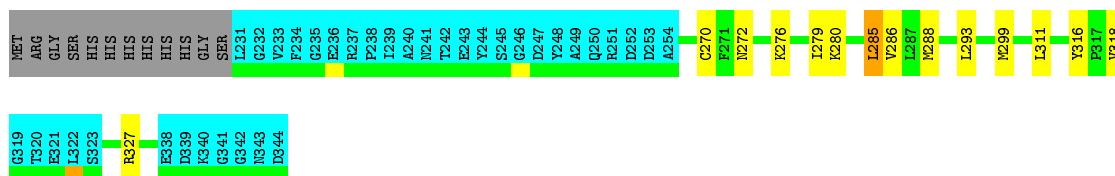
### 4.2.3 Score per residue for model 3

- Molecule 1: DNA-directed RNA polymerase subunit alpha



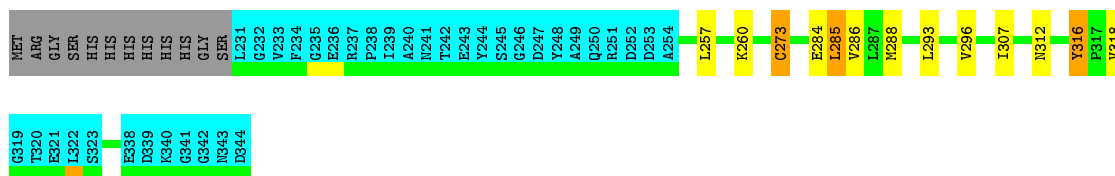
### 4.2.4 Score per residue for model 4

- Molecule 1: DNA-directed RNA polymerase subunit alpha



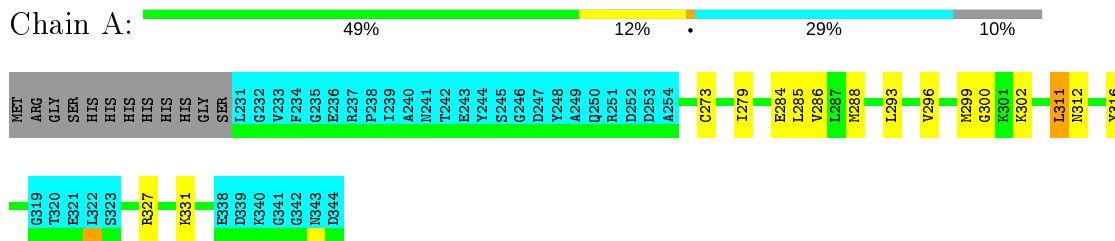
### 4.2.5 Score per residue for model 5

- Molecule 1: DNA-directed RNA polymerase subunit alpha



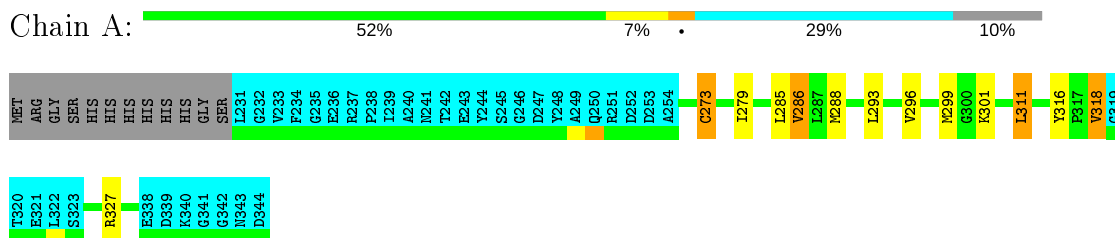
### 4.2.6 Score per residue for model 6

- Molecule 1: DNA-directed RNA polymerase subunit alpha



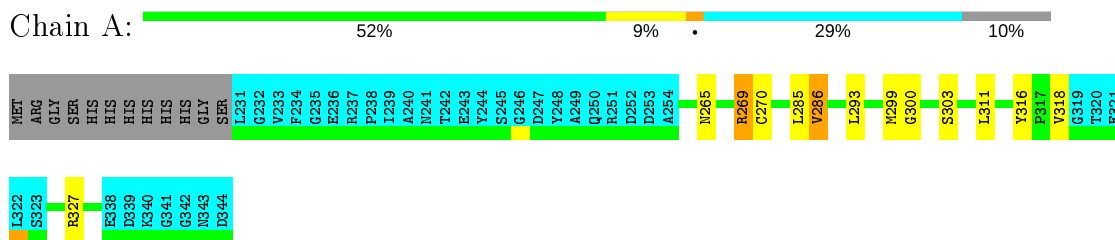
#### 4.2.7 Score per residue for model 7

- Molecule 1: DNA-directed RNA polymerase subunit alpha



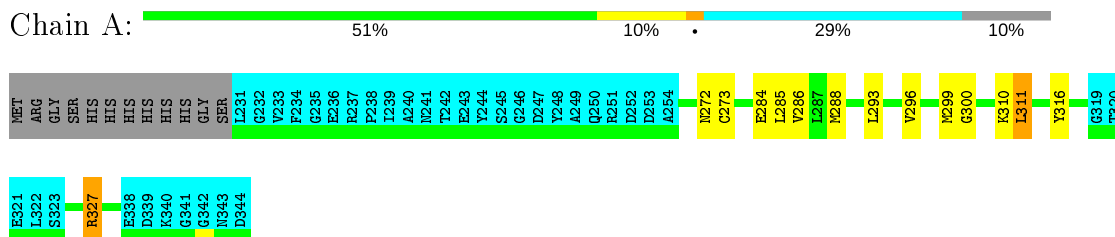
#### 4.2.8 Score per residue for model 8

- Molecule 1: DNA-directed RNA polymerase subunit alpha



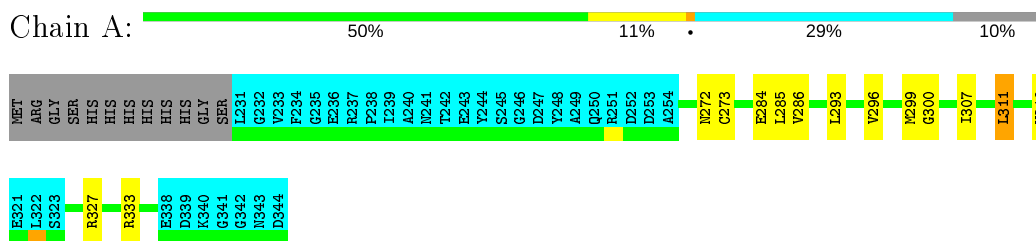
#### 4.2.9 Score per residue for model 9

- Molecule 1: DNA-directed RNA polymerase subunit alpha



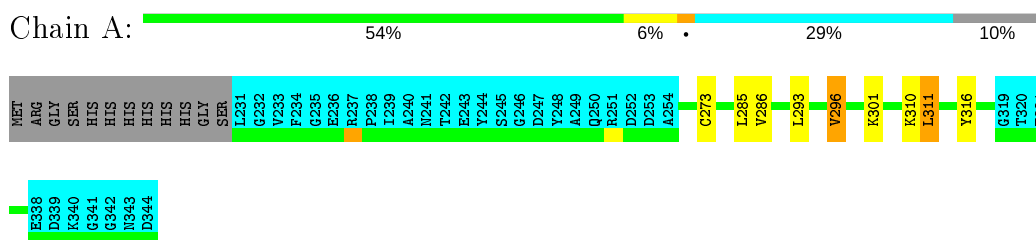
#### 4.2.10 Score per residue for model 10

- Molecule 1: DNA-directed RNA polymerase subunit alpha



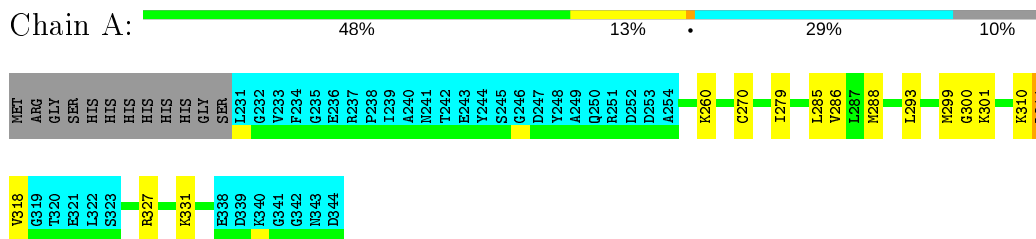
#### 4.2.11 Score per residue for model 11

- Molecule 1: DNA-directed RNA polymerase subunit alpha



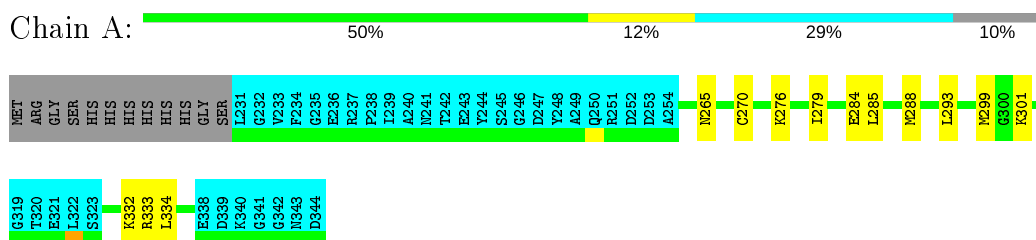
#### 4.2.12 Score per residue for model 12

- Molecule 1: DNA-directed RNA polymerase subunit alpha



#### 4.2.13 Score per residue for model 13

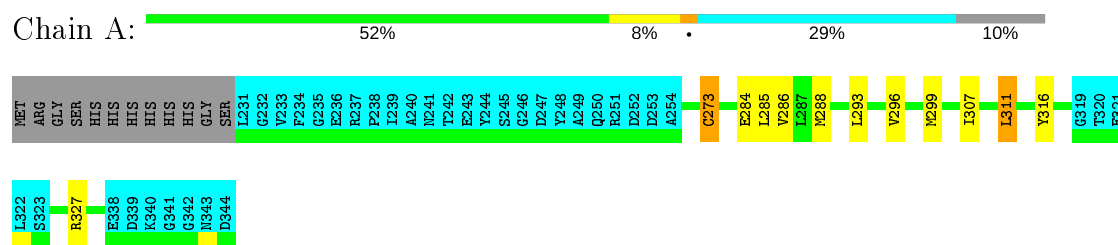
- Molecule 1: DNA-directed RNA polymerase subunit alpha





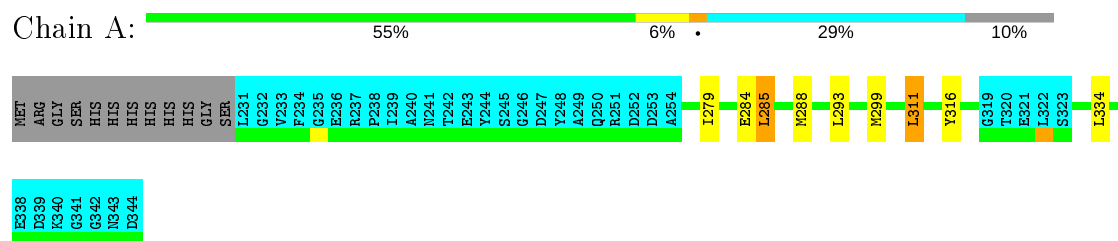
#### 4.2.14 Score per residue for model 14

- Molecule 1: DNA-directed RNA polymerase subunit alpha



#### 4.2.15 Score per residue for model 15

- Molecule 1: DNA-directed RNA polymerase subunit alpha



## 5 Refinement protocol and experimental data overview

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

Of the 500 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
CYANA	structure solution	
TALOS	geometry optimization	
AMBER	refinement	
ProcheckNMR	geometry optimization	
CYANA	refinement	

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	1188
Number of shifts mapped to atoms	1188
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	83%

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

COVALENT-GEOMETRY INFOmissingINFO

### 5.1 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	620	651	651	2±1
All	All	9300	9765	9765	33

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:286:VAL:HG23	1:A:316:TYR:CE2	0.68	2.24	5	1
1:A:257:LEU:HD13	1:A:286:VAL:HG11	0.67	1.65	5	1
1:A:285:LEU:CD2	1:A:318:VAL:HG21	0.63	2.24	5	2
1:A:285:LEU:HD13	1:A:311:LEU:HD23	0.60	1.71	15	2
1:A:286:VAL:HG23	1:A:311:LEU:HD21	0.51	1.81	7	10
1:A:284:GLU:CG	1:A:334:LEU:HD21	0.48	2.38	15	2
1:A:273:CYS:SG	1:A:296:VAL:HG21	0.48	2.48	6	8
1:A:279:ILE:HD13	1:A:288:MET:SD	0.46	2.51	12	7

## 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	78/126 (62%)	71±1 (91±2%)	6±1 (8±2%)	1±1 (1±1%)	21	69
All	All	1170/1890 (62%)	1070 (91%)	89 (8%)	11 (1%)	21	69

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	300	GLY	7
1	A	288	MET	4

### 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	70/107 (65%)	61±2 (87±3%)	9±2 (13±3%)	8	49
All	All	1050/1605 (65%)	917 (87%)	133 (13%)	8	49

All 30 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	293	LEU	15
1	A	316	TYR	15
1	A	285	LEU	15
1	A	299	MET	13
1	A	311	LEU	11
1	A	284	GLU	7
1	A	270	CYS	6
1	A	327	ARG	5
1	A	301	LYS	4
1	A	312	ASN	4
1	A	318	VAL	3
1	A	331	LYS	3
1	A	260	LYS	3
1	A	307	ILE	3
1	A	272	ASN	3
1	A	310	LYS	3
1	A	333	ARG	2
1	A	265	ASN	2
1	A	302	LYS	2
1	A	276	LYS	2
1	A	332	LYS	2
1	A	286	VAL	2
1	A	269	ARG	1
1	A	280	LYS	1
1	A	303	SER	1
1	A	326	GLN	1
1	A	294	LYS	1
1	A	336	LYS	1
1	A	273	CYS	1
1	A	296	VAL	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](#)

There are no ligands in this entry.

### 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 83% for the well-defined parts and 77% 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	1188
Number of shifts mapped to atoms	1188
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

#### 6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	109	$-0.47 \pm 0.11$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	98	$0.08 \pm 0.09$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	0	—	None (insufficient data)
$^{15}\text{N}$	107	$-1.08 \pm 0.38$	Should be applied

#### 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 83%, i.e. 822 atoms were assigned a chemical shift out of a possible 996. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	304/386 (79%)	153/154 (99%)	76/156 (49%)	75/76 (99%)
Sidechain	502/577 (87%)	314/340 (92%)	183/211 (87%)	5/26 (19%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	16/33 (48%)	16/17 (94%)	0/16 (0%)	0/0 (—%)
Overall	822/996 (83%)	483/511 (95%)	259/383 (68%)	80/102 (78%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 77%, i.e. 1069 atoms were assigned a chemical shift out of a possible 1391. 0 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	431/564 (76%)	215/225 (96%)	109/228 (48%)	107/111 (96%)
Sidechain	616/769 (80%)	380/451 (84%)	231/282 (82%)	5/36 (14%)
Aromatic	22/58 (38%)	22/30 (73%)	0/28 (0%)	0/0 (—%)
Overall	1069/1391 (77%)	617/706 (87%)	340/538 (63%)	112/147 (76%)

#### 6.1.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

#### 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:

