



# Full wwPDB NMR Structure Validation Report ⓘ

May 29, 2020 – 07:56 am BST

PDB ID : 5NMY  
Title : NMR solution structure of lysostaphin  
Authors : Tossavainen, H.; Raulinaitis, V.; Permi, P.  
Deposited on : 2017-04-07

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

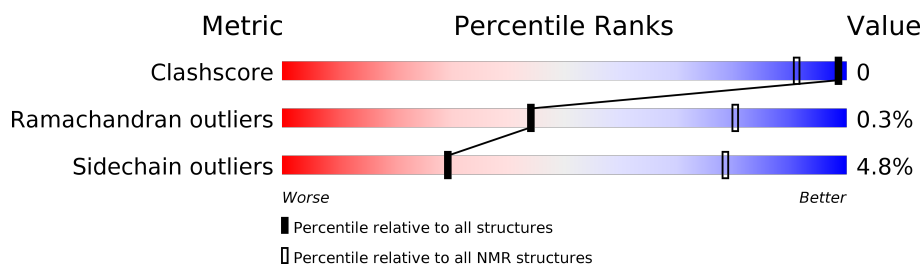
# 1 Overall quality at a glance

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	245	

## 2 Ensemble composition and analysis

This entry contains 15 models. Model 10 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:253-A:268, A:278-A:304, A:313-A:350, A:360-A:388 (110)	0.30	10
2	A:402-A:465, A:470-A:493 (88)	0.25	6

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 5 clusters and 1 single-model cluster was found.

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

### 3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3722 atoms, of which 1823 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Lysostaphin.

Mol	Chain	Residues	Atoms						Trace
1	A	245	Total	C	H	N	O	S	0
			3721	1208	1823	331	352	7	

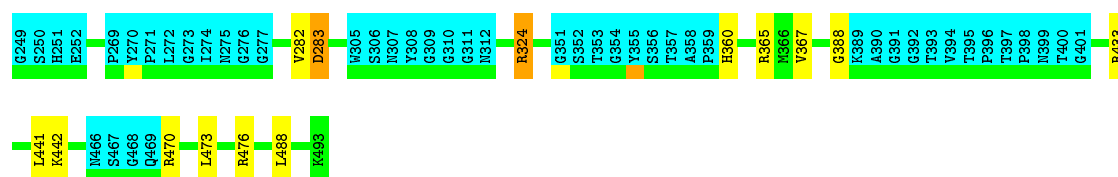
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	249	GLY	-	expression tag	UNP P10547
A	250	SER	-	expression tag	UNP P10547

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

Mol	Chain	Residues	Atoms	
2	A	1	Total	Zn
			1	1

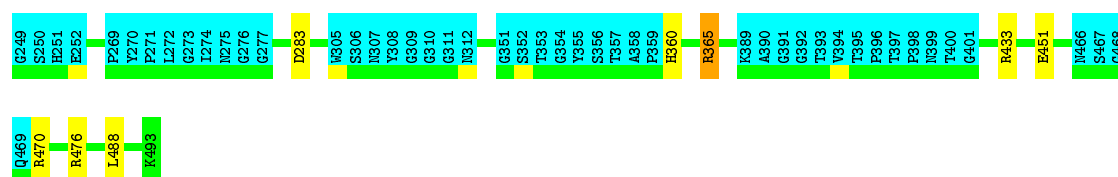




### 4.2.3 Score per residue for model 3

- Molecule 1: Lysostaphin

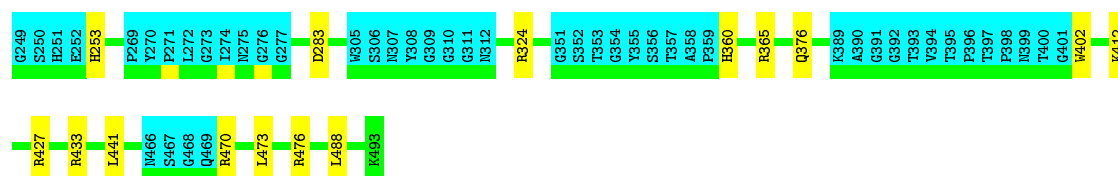
Chain A: 78% 19%



### 4.2.4 Score per residue for model 4

- Molecule 1: Lysostaphin

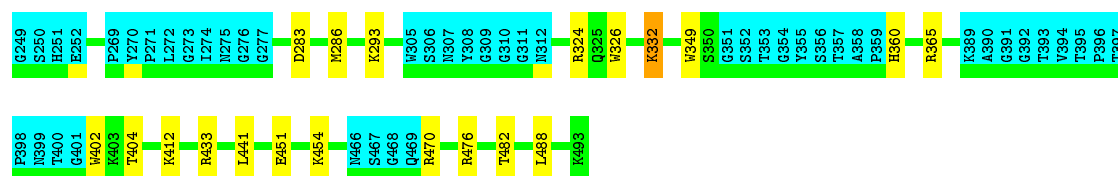
Chain A: 75% 6% 19%



### 4.2.5 Score per residue for model 5

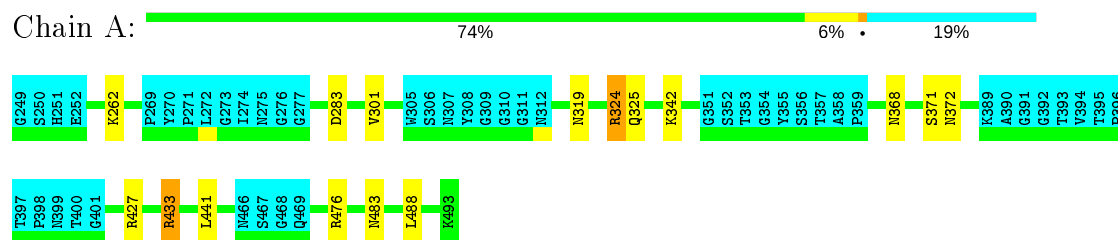
- Molecule 1: Lysostaphin

Chain A: 73% 8% 19%



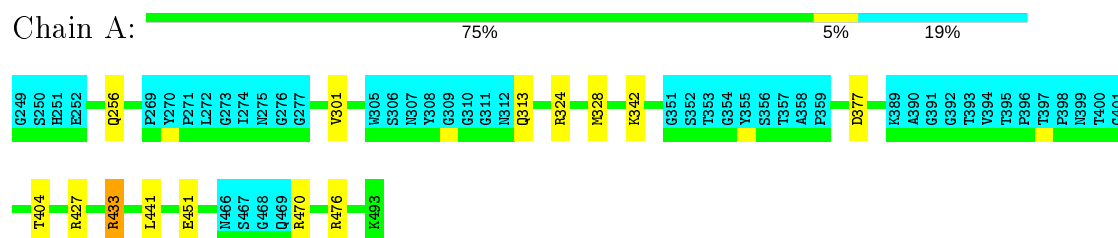
### 4.2.6 Score per residue for model 6

- Molecule 1: Lysostaphin



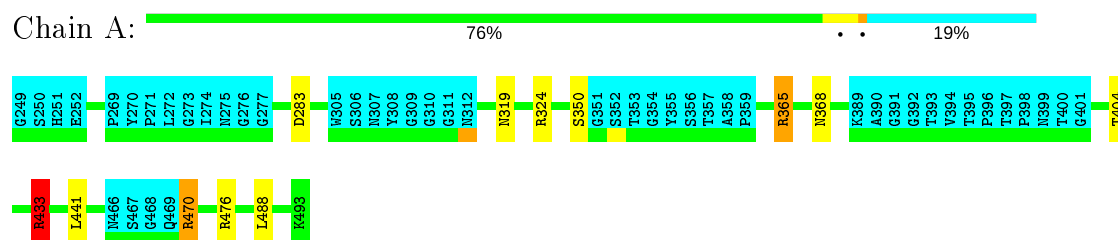
#### 4.2.7 Score per residue for model 7

- Molecule 1: Lysostaphin



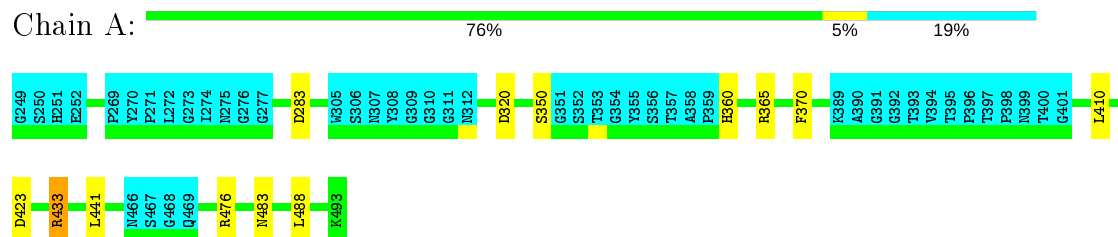
#### 4.2.8 Score per residue for model 8

- Molecule 1: Lysostaphin



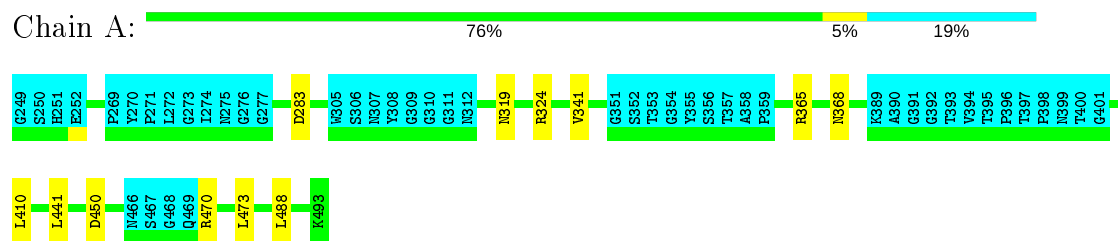
#### 4.2.9 Score per residue for model 9

- Molecule 1: Lysostaphin



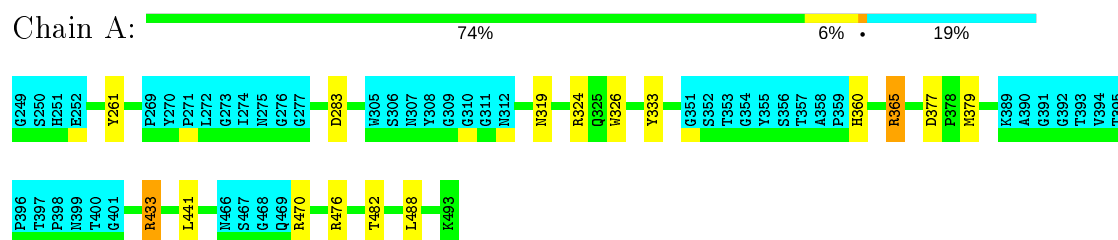
### 4.2.10 Score per residue for model 10 (medoid)

- Molecule 1: Lysostaphin



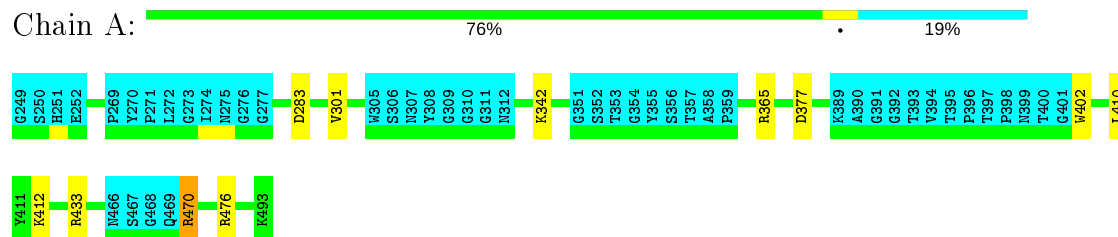
### 4.2.11 Score per residue for model 11

- Molecule 1: Lysostaphin



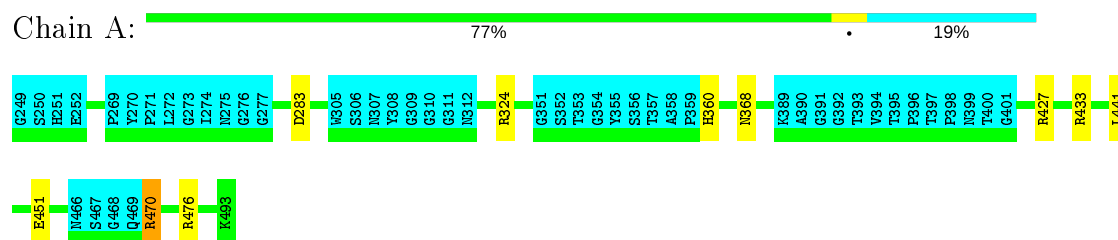
### 4.2.12 Score per residue for model 12

- Molecule 1: Lysostaphin



### 4.2.13 Score per residue for model 13

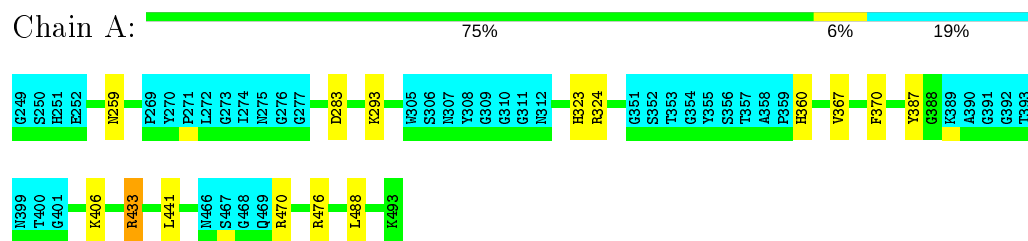
- Molecule 1: Lysostaphin





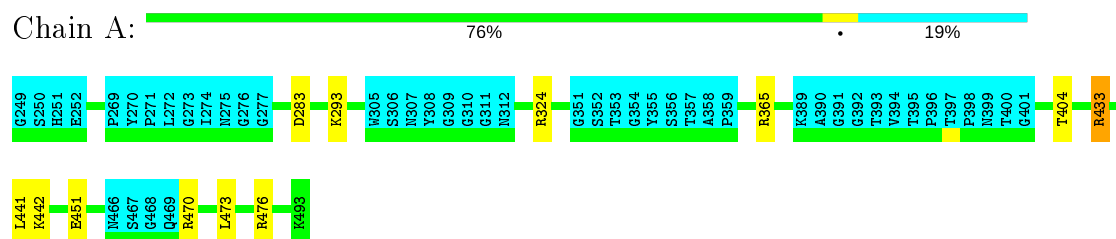
#### 4.2.14 Score per residue for model 14

- Molecule 1: Lysostaphin



#### 4.2.15 Score per residue for model 15

- Molecule 1: Lysostaphin



## 5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 15 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 calculation	
Amber	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	2622
Number of shifts mapped to atoms	2622
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	1579	1538	1538	0±1
All	All	23700	23070	23070	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	Total
1:A:402:TRP:CH2	1:A:454:LYS:HE3	0.49	2.43	5	1
1:A:402:TRP:CZ2	1:A:412:LYS:HE2	0.49	2.43	12	1
1:A:402:TRP:CE2	1:A:412:LYS:HE3	0.44	2.48	4	1
1:A:332:LYS:HE3	1:A:349:TRP:CZ3	0.43	2.47	5	1
1:A:402:TRP:CZ2	1:A:412:LYS:HE3	0.43	2.48	5	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	197/245 (80%)	190±1 (96±1%)	7±1 (3±1%)	1±1 (0±0%)	44	80
All	All	2955/3675 (80%)	2848 (96%)	99 (3%)	8 (0%)	44	80

All 6 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	368	ASN	3
1	A	319	ASN	1
1	A	259	ASN	1
1	A	253	HIS	1
1	A	350	SER	1
1	A	388	GLY	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	167/199 (84%)	159±2 (95±1%)	8±2 (5±1%)	29	78
All	All	2505/2985 (84%)	2385 (95%)	120 (5%)	29	78

All 43 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	433	ARG	14
1	A	441	LEU	13
1	A	488	LEU	11
1	A	360	HIS	8
1	A	451	GLU	5
1	A	473	LEU	5
1	A	319	ASN	4
1	A	342	LYS	4
1	A	404	THR	4
1	A	324	ARG	3
1	A	377	ASP	3
1	A	301	VAL	3
1	A	293	LYS	3
1	A	410	LEU	3
1	A	450	ASP	2
1	A	367	VAL	2
1	A	482	THR	2
1	A	365	ARG	2
1	A	370	PHE	2
1	A	326	TRP	2
1	A	442	LYS	2
1	A	483	ASN	2
1	A	286	MET	1
1	A	282	VAL	1
1	A	470	ARG	1
1	A	368	ASN	1
1	A	341	VAL	1
1	A	323	HIS	1
1	A	406	LYS	1
1	A	313	GLN	1
1	A	371	SER	1
1	A	350	SER	1
1	A	379	MET	1
1	A	332	LYS	1
1	A	333	TYR	1
1	A	320	ASP	1
1	A	372	ASN	1
1	A	262	LYS	1
1	A	256	GLN	1
1	A	325	GLN	1
1	A	283	ASP	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	328	MET	1
1	A	376	GLN	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 [i](#)

The completeness of assignment taking into account all chemical shift lists is 83% for the well-defined parts and 80% for the entire structure.

### 6.1 Chemical shift list 1

File name: input\_cs.cif

Chemical shift list name: *LSSv3.1.bmrB*

#### 6.1.1 Bookkeeping [i](#)

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	2622
Number of shifts mapped to atoms	2622
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	25

#### 6.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	235	$-1.16 \pm 0.18$	Should be applied
$^{13}\text{C}_\beta$	206	$-1.53 \pm 0.07$	Should be applied
$^{13}\text{C}'$	0	—	None (insufficient data)
$^{15}\text{N}$	207	$0.04 \pm 0.34$	None needed ( $< 0.5$ ppm)

#### 6.1.3 Completeness of resonance assignments [i](#)

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. 2012 atoms were assigned a chemical shift out of a possible 2436. 23 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	768/976 (79%)	384/389 (99%)	198/396 (50%)	186/191 (97%)
Sidechain	1010/1153 (88%)	629/681 (92%)	358/419 (85%)	23/53 (43%)

*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	234/307 (76%)	143/161 (89%)	81/131 (62%)	10/15 (67%)
Overall	2012/2436 (83%)	1156/1231 (94%)	637/946 (67%)	219/259 (85%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 80%, i.e. 2315 atoms were assigned a chemical shift out of a possible 2896. 25 out of 26 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	886/1201 (74%)	444/478 (93%)	235/490 (48%)	207/233 (89%)
Sidechain	1165/1345 (87%)	727/796 (91%)	411/489 (84%)	27/60 (45%)
Aromatic	264/350 (75%)	162/183 (89%)	91/150 (61%)	11/17 (65%)
Overall	2315/2896 (80%)	1333/1457 (91%)	737/1129 (65%)	245/310 (79%)

#### 6.1.4 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	454	LYS	HB3	-1.02	3.10 – 0.40	-10.3
1	A	454	LYS	HE3	0.98	3.86 – 1.96	-10.2
1	A	449	TYR	CD1	120.26	139.11 – 126.41	-9.8
1	A	293	LYS	HB3	-0.64	3.10 – 0.40	-8.8
1	A	454	LYS	HD3	-0.36	2.75 – 0.45	-8.5
1	A	324	ARG	HD3	1.30	4.36 – 1.86	-7.2
1	A	454	LYS	HE2	1.59	3.87 – 1.97	-7.0
1	A	258	LEU	HG	-0.71	3.16 – -0.14	-6.7
1	A	474	PRO	HB3	-0.39	3.81 – 0.21	-6.7
1	A	372	ASN	HB3	0.59	4.41 – 1.11	-6.6
1	A	454	LYS	HD2	0.10	2.76 – 0.46	-6.6
1	A	431	PRO	HA	2.39	6.05 – 2.75	-6.1
1	A	258	LEU	HD21	-0.96	2.14 – -0.66	-6.1
1	A	258	LEU	HD23	-0.96	2.14 – -0.66	-6.1
1	A	258	LEU	HD22	-0.96	2.14 – -0.66	-6.1
1	A	412	LYS	HD3	0.25	2.75 – 0.45	-5.9
1	A	324	ARG	HB3	0.16	3.17 – 0.37	-5.8
1	A	315	GLY	HA2	6.13	5.87 – 2.07	5.7
1	A	324	ARG	HB2	0.29	3.15 – 0.45	-5.6
1	A	431	PRO	HB3	0.09	3.81 – 0.21	-5.3
1	A	293	LYS	H	4.94	11.24 – 5.14	-5.3
1	A	454	LYS	HB2	0.45	3.03 – 0.53	-5.3

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	412	LYS	HD2	0.40	2.76 – 0.46	-5.2
1	A	431	PRO	HB2	0.26	3.82 – 0.32	-5.2
1	A	324	ARG	HD2	1.97	4.27 – 1.97	-5.0

### 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:

