



# Full wwPDB NMR Structure Validation Report ⓘ

Aug 4, 2021 – 10:06 am BST

PDB ID : 7BEV  
Title : NMR structure of an optimized version of the first TPR domain of the human SPAG1 protein  
Authors : Dermouche, S.; Chagot, M.E.; Quinternet, M.  
Deposited on : 2020-12-29

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:

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 : 2.23.1  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.23.1

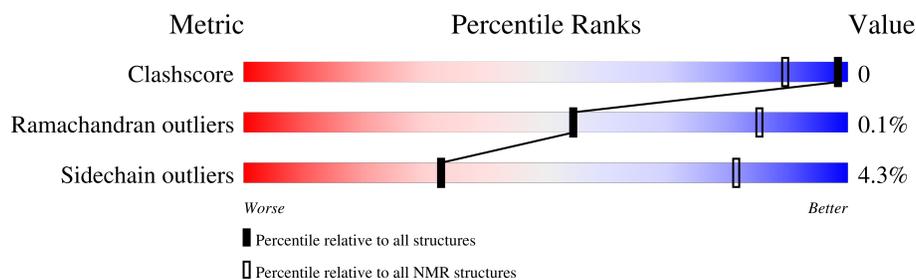
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 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	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	85% 8% 7%

## 2 Ensemble composition and analysis

This entry contains 20 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:209-A:325 (117)	0.26	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, 12, 13, 14, 15, 16, 17, 18, 19, 20
2	10, 11

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2032 atoms, of which 1012 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Sperm-associated antigen 1.

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		S
1	A	126	2032	632	1012	182	203	3	0

There are 11 discrepancies between the modelled and reference sequences:

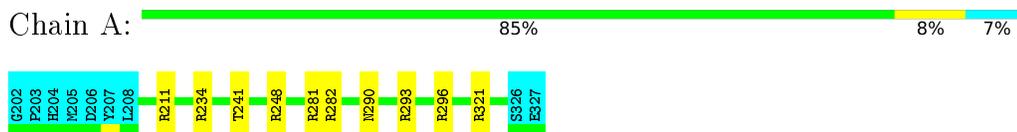
Chain	Residue	Modelled	Actual	Comment	Reference
A	202	GLY	-	expression tag	UNP Q07617
A	203	PRO	-	expression tag	UNP Q07617
A	204	HIS	-	expression tag	UNP Q07617
A	205	MET	-	expression tag	UNP Q07617
A	207	TYR	PHE	conflict	UNP Q07617
A	252	TYR	GLU	conflict	UNP Q07617
A	262	GLU	PHE	conflict	UNP Q07617
A	285	ALA	THR	conflict	UNP Q07617
A	296	ARG	THR	conflict	UNP Q07617
A	300	LYS	SER	conflict	UNP Q07617
A	304	LYS	ASP	conflict	UNP Q07617

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

### 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: Sperm-associated antigen 1

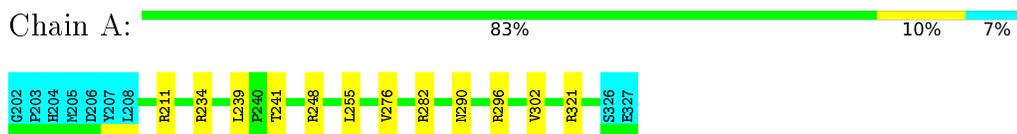


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

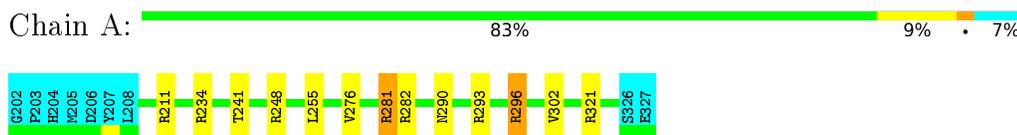
#### 4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Sperm-associated antigen 1



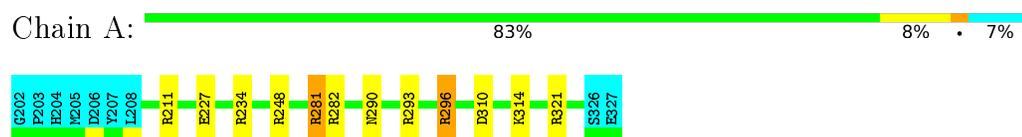
#### 4.2.2 Score per residue for model 2

- Molecule 1: Sperm-associated antigen 1



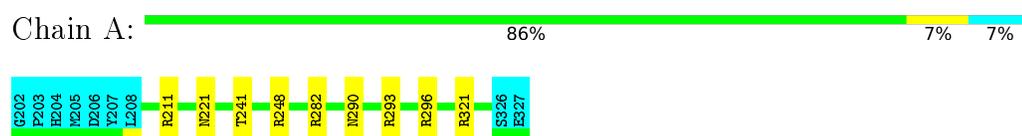
### 4.2.3 Score per residue for model 3

- Molecule 1: Sperm-associated antigen 1



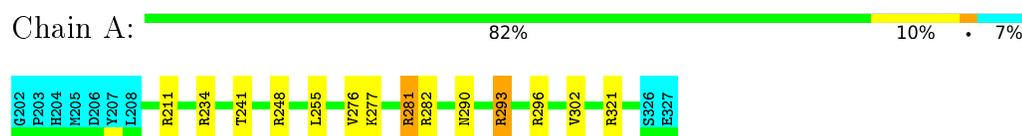
### 4.2.4 Score per residue for model 4

- Molecule 1: Sperm-associated antigen 1



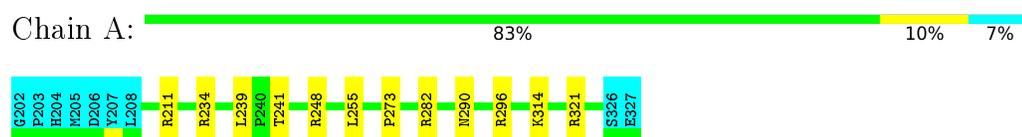
### 4.2.5 Score per residue for model 5

- Molecule 1: Sperm-associated antigen 1



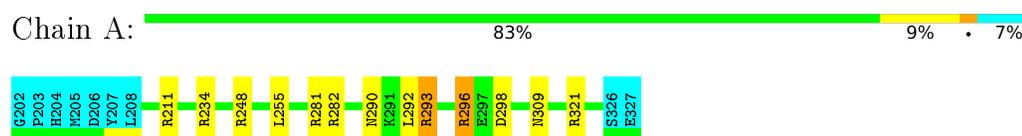
### 4.2.6 Score per residue for model 6

- Molecule 1: Sperm-associated antigen 1



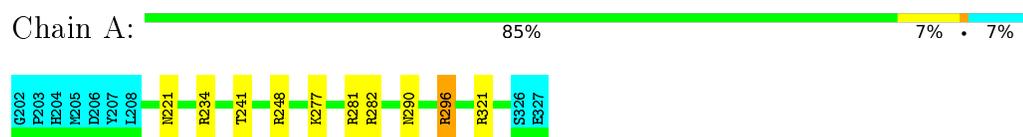
### 4.2.7 Score per residue for model 7

- Molecule 1: Sperm-associated antigen 1



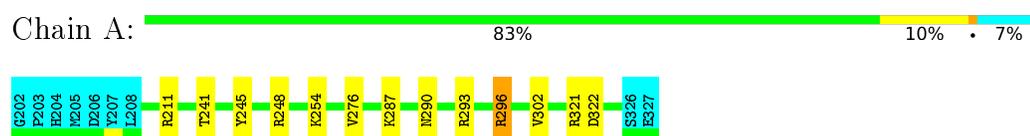
### 4.2.8 Score per residue for model 8

- Molecule 1: Sperm-associated antigen 1



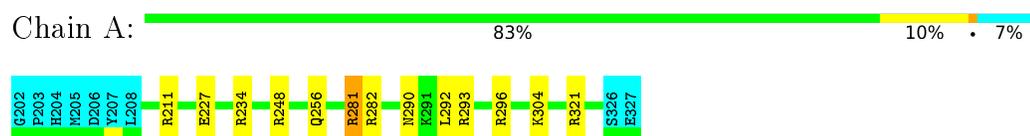
### 4.2.9 Score per residue for model 9

- Molecule 1: Sperm-associated antigen 1



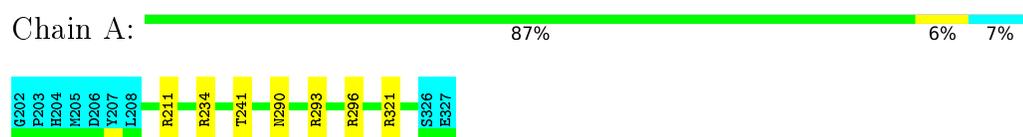
### 4.2.10 Score per residue for model 10

- Molecule 1: Sperm-associated antigen 1



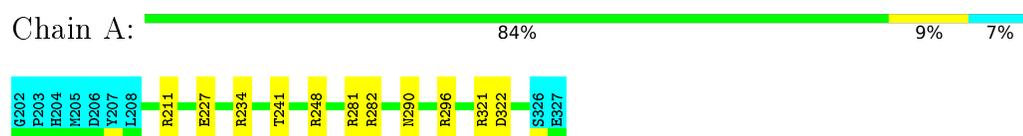
### 4.2.11 Score per residue for model 11

- Molecule 1: Sperm-associated antigen 1



### 4.2.12 Score per residue for model 12

- Molecule 1: Sperm-associated antigen 1



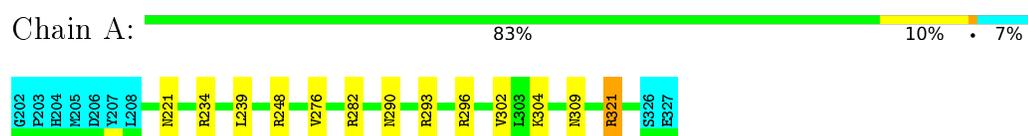
### 4.2.13 Score per residue for model 13

- Molecule 1: Sperm-associated antigen 1



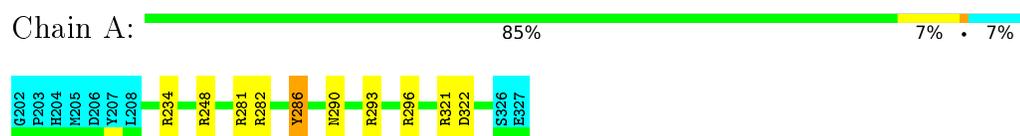
### 4.2.14 Score per residue for model 14

- Molecule 1: Sperm-associated antigen 1



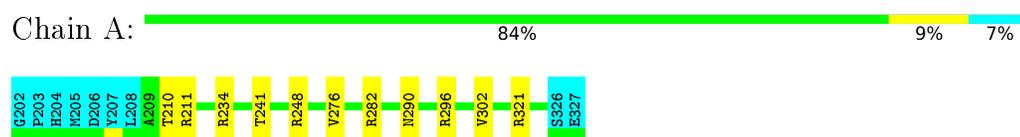
### 4.2.15 Score per residue for model 15

- Molecule 1: Sperm-associated antigen 1



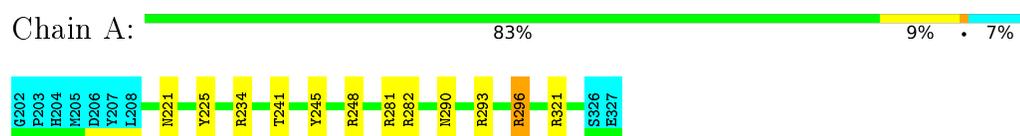
### 4.2.16 Score per residue for model 16

- Molecule 1: Sperm-associated antigen 1



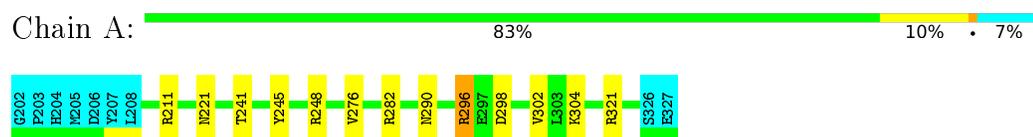
### 4.2.17 Score per residue for model 17

- Molecule 1: Sperm-associated antigen 1



#### 4.2.18 Score per residue for model 18

- Molecule 1: Sperm-associated antigen 1



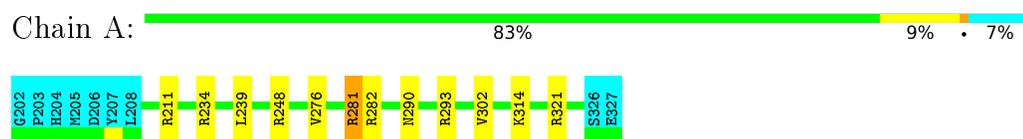
#### 4.2.19 Score per residue for model 19

- Molecule 1: Sperm-associated antigen 1



#### 4.2.20 Score per residue for model 20

- Molecule 1: Sperm-associated antigen 1



## 5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest restraint energy*.

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 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1608
Number of shifts mapped to atoms	1595
Number of unparsed shifts	13
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

## 6 Model quality i

### 6.1 Standard geometry i

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.68±0.01	0±0/960 ( 0.0± 0.0%)	1.08±0.04	7±1/1293 ( 0.5± 0.1%)
All	All	0.68	0/19200 ( 0.0%)	1.08	139/25860 ( 0.5%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.3±0.5
All	All	0	7

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	282	ARG	NE-CZ-NH1	11.27	125.93	120.30	6	18
1	A	281	ARG	NE-CZ-NH1	10.74	125.67	120.30	17	11
1	A	296	ARG	NE-CZ-NH1	9.46	125.03	120.30	6	16
1	A	293	ARG	NE-CZ-NH1	9.32	124.96	120.30	7	14
1	A	211	ARG	NE-CZ-NH1	9.15	124.88	120.30	2	16
1	A	234	ARG	NE-CZ-NH1	8.84	124.72	120.30	12	17
1	A	321	ARG	NE-CZ-NH1	8.45	124.53	120.30	14	20
1	A	248	ARG	NE-CZ-NH1	7.99	124.29	120.30	1	19
1	A	211	ARG	NE-CZ-NH2	-5.78	117.41	120.30	4	1
1	A	281	ARG	NE-CZ-NH2	-5.52	117.54	120.30	3	1
1	A	225	TYR	CB-CG-CD2	-5.42	117.75	121.00	17	1
1	A	282	ARG	NE-CZ-NH2	-5.36	117.62	120.30	17	1
1	A	225	TYR	CB-CG-CD1	-5.17	117.90	121.00	19	1
1	A	234	ARG	NE-CZ-NH2	-5.10	117.75	120.30	20	1
1	A	296	ARG	CD-NE-CZ	5.09	130.73	123.60	6	1

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	286	TYR	CB-CG-CD2	-5.03	117.98	121.00	15	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	245	TYR	Sidechain	3
1	A	286	TYR	Sidechain	2
1	A	296	ARG	Sidechain	1
1	A	321	ARG	Sidechain	1

## 6.2 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	947	948	948	0±0
All	All	18940	18960	18960	9

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:276:VAL:HG13	1:A:302:VAL:HG13	0.44	1.87	1	9

## 6.3 Torsion angles [i](#)

### 6.3.1 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	117/126 (93%)	112±1 (96±1%)	4±1 (4±1%)	0±0 (0±0%)	54 85
All	All	2340/2520 (93%)	2249 (96%)	88 (4%)	3 (0%)	54 85

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	273	PRO	2
1	A	309	ASN	1

### 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	101/109 (93%)	97±1 (96±1%)	4±1 (4±1%)	33 81
All	All	2020/2180 (93%)	1934 (96%)	86 (4%)	33 81

All 21 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	290	ASN	20
1	A	241	THR	12
1	A	296	ARG	7
1	A	255	LEU	5
1	A	281	ARG	5
1	A	221	ASN	5
1	A	239	LEU	4
1	A	322	ASP	4
1	A	304	LYS	4
1	A	227	GLU	3
1	A	314	LYS	3
1	A	277	LYS	2
1	A	293	ARG	2
1	A	292	LEU	2
1	A	298	ASP	2
1	A	310	ASP	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	254	LYS	1
1	A	287	LYS	1
1	A	256	GLN	1
1	A	309	ASN	1
1	A	210	THR	1

### 6.3.3 RNA [i](#)

There are no RNA molecules in this entry.

### 6.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 6.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

### 6.6 Ligand geometry [i](#)

There are no ligands in this entry.

### 6.7 Other polymers [i](#)

There are no such molecules in this entry.

### 6.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 7 Chemical shift validation

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

### 7.1 Chemical shift list 1

File name: working\_cs.cif

Chemical shift list name: 206327opt\_dss\_293K.str

#### 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	1608
Number of shifts mapped to atoms	1595
Number of unparsed shifts	13
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

The following errors were found when reading this chemical shift list.

- Chemical shift has been reported more than once. All 13 occurrences are reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
3	A	202	GLY	HA2	004.034	0.020	1
10	A	203	PRO	HB2	002.330	0.020	2
12	A	203	PRO	HG2	002.002	0.020	1
14	A	203	PRO	HD2	003.609	0.020	2
455	A	240	PRO	HB2	002.472	0.020	2
457	A	240	PRO	HG2	002.260	0.020	2
459	A	240	PRO	HD2	003.878	0.020	2
887	A	273	PRO	HB2	002.379	0.020	2
889	A	273	PRO	HG2	002.051	0.020	2
891	A	273	PRO	HD2	003.632	0.020	2
1354	A	307	PRO	HB2	002.406	0.020	2
1356	A	307	PRO	HG2	001.998	0.020	2
1358	A	307	PRO	HD2	003.219	0.020	2

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

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	126	$-0.59 \pm 0.11$	Should be applied
$^{13}\text{C}_\beta$	122	$0.46 \pm 0.12$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	118	$-0.44 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	121	$-0.06 \pm 0.20$	None needed ( $< 0.5$ ppm)

### 7.1.3 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 91%, i.e. 1364 atoms were assigned a chemical shift out of a possible 1500. 20 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	574/579 (99%)	231/231 (100%)	229/234 (98%)	114/114 (100%)
Sidechain	728/845 (86%)	455/494 (92%)	253/300 (84%)	20/51 (39%)
Aromatic	62/76 (82%)	34/39 (87%)	27/35 (77%)	1/2 (50%)
Overall	1364/1500 (91%)	720/764 (94%)	509/569 (89%)	135/167 (81%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 91%, i.e. 1456 atoms were assigned a chemical shift out of a possible 1605. 21 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	611/622 (98%)	246/248 (99%)	244/252 (97%)	121/122 (99%)
Sidechain	771/891 (87%)	482/522 (92%)	269/318 (85%)	20/51 (39%)
Aromatic	74/92 (80%)	40/47 (85%)	33/41 (80%)	1/4 (25%)
Overall	1456/1605 (91%)	768/817 (94%)	546/611 (89%)	142/177 (80%)

### 7.1.4 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	314	LYS	CE	36.04	46.00 – 37.80	-7.1
1	A	260	SER	HB2	2.13	5.18 – 2.58	-6.7
1	A	256	GLN	HA	1.66	6.41 – 2.11	-6.0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	248	ARG	HE	11.08	10.48 – 4.28	6.0
1	A	215	LYS	HB3	0.29	3.10 – 0.40	-5.4
1	A	260	SER	HB3	2.42	5.25 – 2.45	-5.1

### 7.1.5 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:

