



# Full wwPDB NMR Structure Validation Report ⓘ

Jun 12, 2024 – 12:58 AM EDT

PDB ID : 2AI6  
BMRB ID : 6625  
Title : Solution structure of human phosphohistidine phosphatase 1  
Authors : Gong, W.; Cui, G.; Jin, C.; Xia, B.  
Deposited on : 2005-07-29

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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

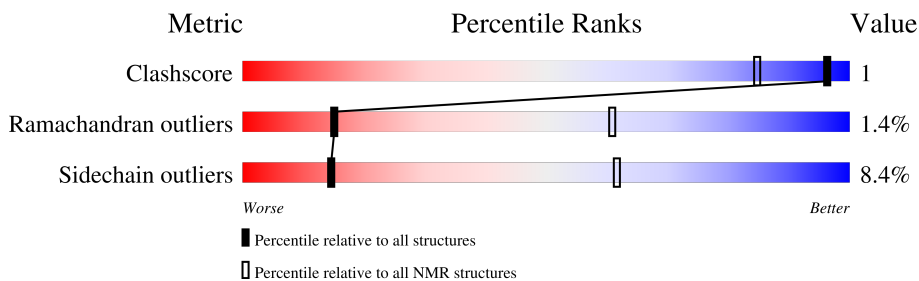
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*


The overall completeness of chemical shifts assignment is 85%.

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	125	

## 2 Ensemble composition and analysis

This entry contains 21 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:5-A:14, A:20-A:34, A:38-A:80, A:89-A:113, A:117-A:121 (98)	0.34	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, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21
2	6, 16

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1904 atoms, of which 931 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called 14 kDa phosphohistidine phosphatase.

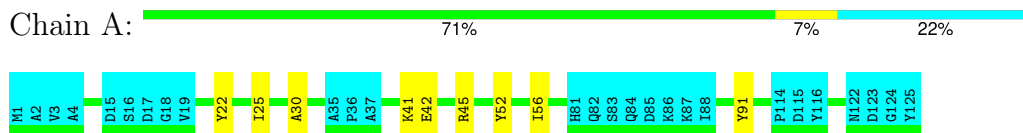
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	125	1904	613	931	165	189	6	0

## 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: 14 kDa phosphohistidine phosphatase

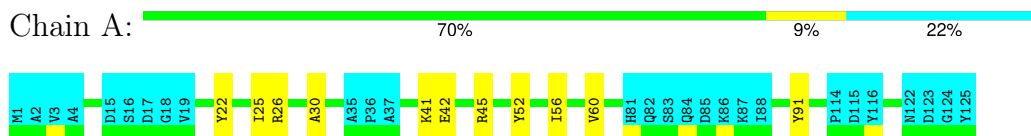


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

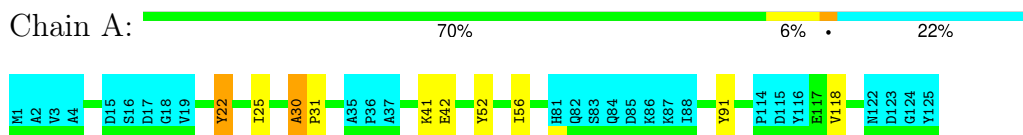
#### 4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: 14 kDa phosphohistidine phosphatase



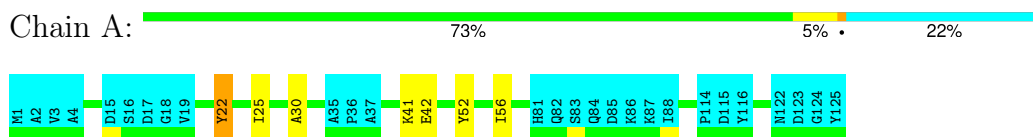
#### 4.2.2 Score per residue for model 2

- Molecule 1: 14 kDa phosphohistidine phosphatase



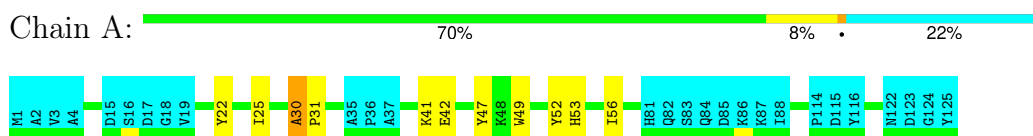
### 4.2.3 Score per residue for model 3

- Molecule 1: 14 kDa phosphohistidine phosphatase



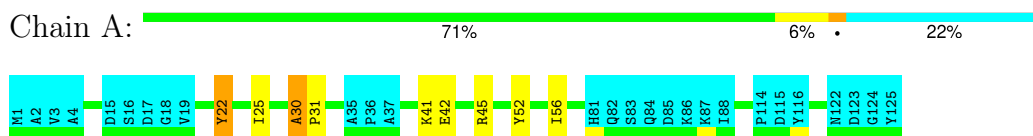
### 4.2.4 Score per residue for model 4

- Molecule 1: 14 kDa phosphohistidine phosphatase



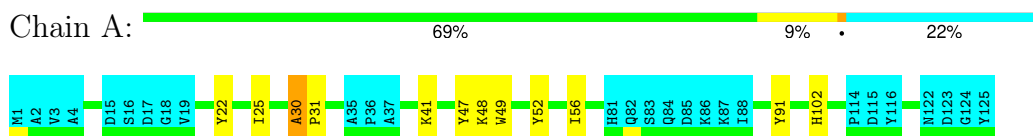
### 4.2.5 Score per residue for model 5

- Molecule 1: 14 kDa phosphohistidine phosphatase



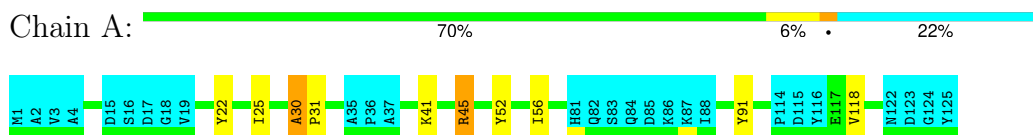
### 4.2.6 Score per residue for model 6

- Molecule 1: 14 kDa phosphohistidine phosphatase



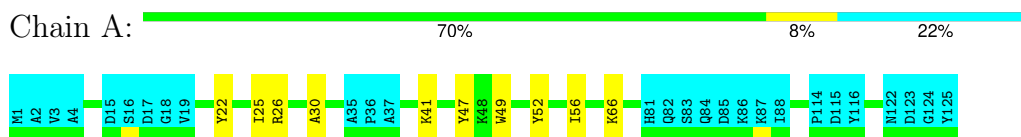
### 4.2.7 Score per residue for model 7

- Molecule 1: 14 kDa phosphohistidine phosphatase



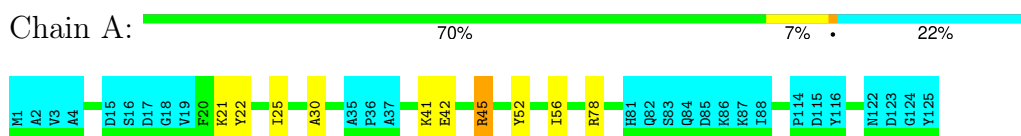
### 4.2.8 Score per residue for model 8

- Molecule 1: 14 kDa phosphohistidine phosphatase



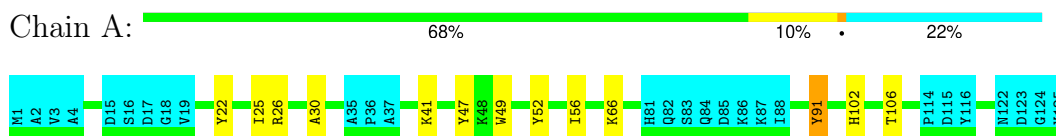
### 4.2.9 Score per residue for model 9

- Molecule 1: 14 kDa phosphohistidine phosphatase



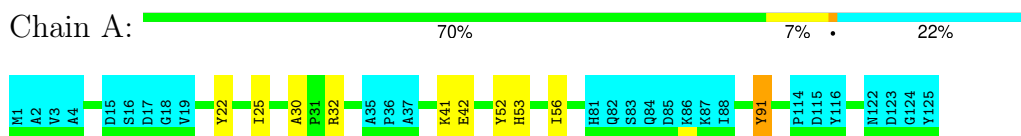
### 4.2.10 Score per residue for model 10

- Molecule 1: 14 kDa phosphohistidine phosphatase



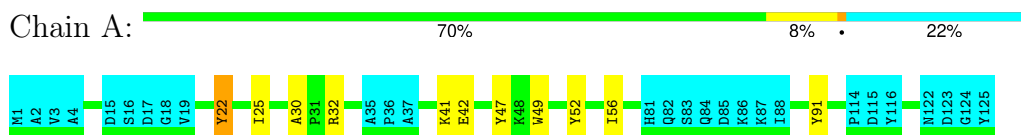
### 4.2.11 Score per residue for model 11

- Molecule 1: 14 kDa phosphohistidine phosphatase



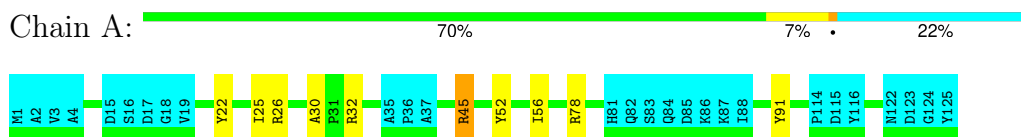
### 4.2.12 Score per residue for model 12

- Molecule 1: 14 kDa phosphohistidine phosphatase



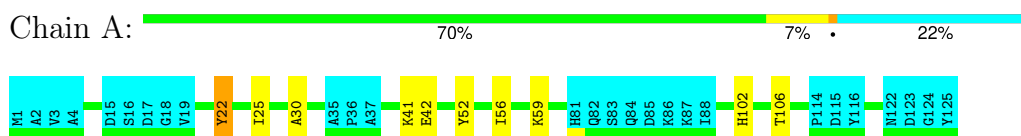
### 4.2.13 Score per residue for model 13

- Molecule 1: 14 kDa phosphohistidine phosphatase



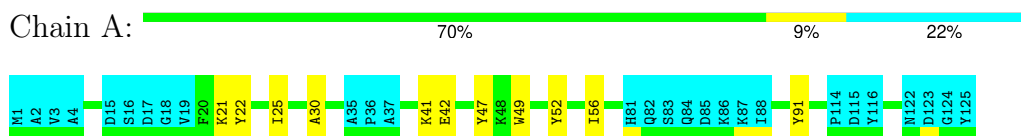
### 4.2.14 Score per residue for model 14

- Molecule 1: 14 kDa phosphohistidine phosphatase



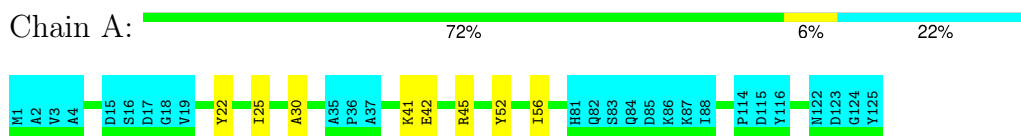
### 4.2.15 Score per residue for model 15

- Molecule 1: 14 kDa phosphohistidine phosphatase



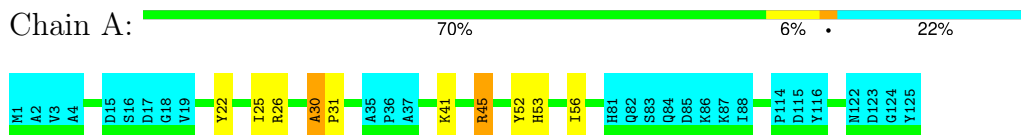
### 4.2.16 Score per residue for model 16

- Molecule 1: 14 kDa phosphohistidine phosphatase



### 4.2.17 Score per residue for model 17

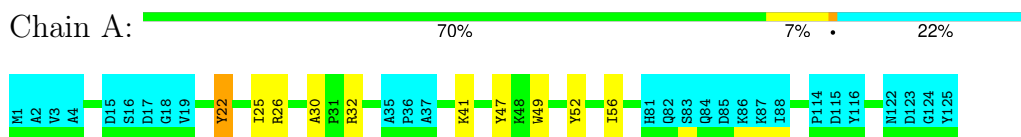
- Molecule 1: 14 kDa phosphohistidine phosphatase





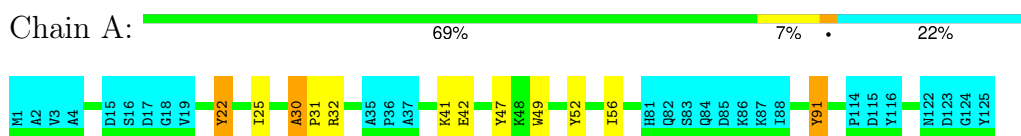
#### 4.2.18 Score per residue for model 18

- Molecule 1: 14 kDa phosphohistidine phosphatase



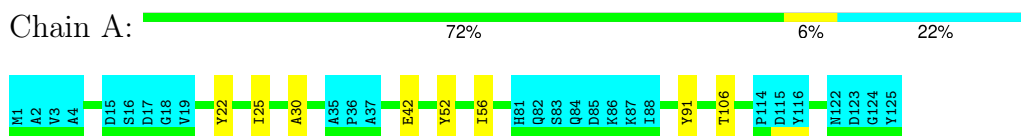
#### 4.2.19 Score per residue for model 19

- Molecule 1: 14 kDa phosphohistidine phosphatase



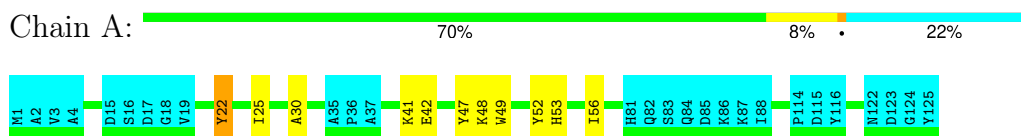
#### 4.2.20 Score per residue for model 20

- Molecule 1: 14 kDa phosphohistidine phosphatase



#### 4.2.21 Score per residue for model 21

- Molecule 1: 14 kDa phosphohistidine phosphatase



## 5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 21 were deposited, based on the following criterion: *structures with favorable non-bond energy*.

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

Software name	Classification	Version
DYANA	structure solution	cyana2.0
Amber	refinement	7.0

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	1293
Number of shifts mapped to atoms	1293
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

## 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	0.67±0.00	0±0/790 ( 0.0± 0.0%)	1.00±0.17	1±1/1069 ( 0.1± 0.1%)
All	All	0.67	0/16590 ( 0.0%)	1.02	18/22449 ( 0.1%)

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	0.7±0.6
All	All	0	14

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	42	GLU	OE1-CD-OE2	-37.59	78.20	123.30	1	1
1	A	42	GLU	CG-CD-OE2	21.35	160.99	118.30	1	1
1	A	42	GLU	CG-CD-OE1	-17.70	82.91	118.30	1	1
1	A	26	ARG	NE-CZ-NH2	-7.66	116.47	120.30	1	6
1	A	78	ARG	NE-CZ-NH2	-6.22	117.19	120.30	9	2
1	A	45	ARG	NE-CZ-NH1	5.45	123.02	120.30	17	4
1	A	91	TYR	CB-CG-CD2	-5.31	117.81	121.00	11	3

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	22	TYR	Sidechain	8

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	45	ARG	Sidechain	6

## 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	770	748	748	1±1
All	All	16170	15708	15708	24

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:30:ALA:H	1:A:31:PRO:HD3	0.45	1.72	6	3
1:A:47:TYR:CD1	1:A:49:TRP:CZ2	0.43	3.06	8	8
1:A:53:HIS:CD2	1:A:53:HIS:H	0.43	2.31	11	4
1:A:30:ALA:N	1:A:31:PRO:CD	0.43	2.82	2	6
1:A:30:ALA:H	1:A:31:PRO:CD	0.42	2.27	4	1
1:A:45:ARG:HG3	1:A:60:VAL:HG21	0.42	1.90	1	1
1:A:47:TYR:CD2	1:A:49:TRP:CZ2	0.40	3.09	4	1

## 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	98/125 (78%)	92±1 (94±1%)	4±1 (4±1%)	1±0 (1±0%)	15	61
All	All	2058/2625 (78%)	1939 (94%)	90 (4%)	29 (1%)	15	61

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	30	ALA	21
1	A	32	ARG	5
1	A	102	HIS	3

### 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	79/100 (79%)	72±1 (92±1%)	7±1 (8±1%)	14	61
All	All	1659/2100 (79%)	1520 (92%)	139 (8%)	14	61

All 13 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	22	TYR	21
1	A	25	ILE	21
1	A	52	TYR	21
1	A	56	ILE	21
1	A	41	LYS	19
1	A	42	GLU	13
1	A	91	TYR	11
1	A	106	THR	3
1	A	118	VAL	2
1	A	48	LYS	2
1	A	66	LYS	2
1	A	21	LYS	2
1	A	59	LYS	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 [i](#)

The completeness of assignment taking into account all chemical shift lists is 85% for the well-defined parts and 80% for the entire structure.

### 7.1 Chemical shift list 1

File name: working\_cs.cif

Chemical shift list name: *assigned\_chem\_shift\_list\_1*

#### 7.1.1 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	1293
Number of shifts mapped to atoms	1293
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	13

#### 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
$^{13}\text{C}_\alpha$	118	-0.05 $\pm$ 0.09	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	108	-0.07 $\pm$ 0.16	None needed (< 0.5 ppm)
$^{13}\text{C}'$	110	-0.20 $\pm$ 0.09	None needed (< 0.5 ppm)
$^{15}\text{N}$	103	-0.08 $\pm$ 0.23	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 85%, i.e. 1110 atoms were assigned a chemical shift out of a possible 1303. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	481/493 (98%)	197/202 (98%)	193/196 (98%)	91/95 (96%)
Sidechain	569/676 (84%)	379/441 (86%)	185/212 (87%)	5/23 (22%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	60/134 (45%)	30/65 (46%)	28/63 (44%)	2/6 (33%)
Overall	1110/1303 (85%)	606/708 (86%)	406/471 (86%)	98/124 (79%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 80%, i.e. 1293 atoms were assigned a chemical shift out of a possible 1626. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	559/626 (89%)	228/256 (89%)	228/250 (91%)	103/120 (86%)
Sidechain	669/841 (80%)	444/547 (81%)	217/266 (82%)	8/28 (29%)
Aromatic	65/159 (41%)	33/77 (43%)	30/75 (40%)	2/7 (29%)
Overall	1293/1626 (80%)	705/880 (80%)	475/591 (80%)	113/155 (73%)

#### 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.

List Id	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	109	ILE	CD1	67.53	5.18 – 21.60	33.0
1	A	26	ARG	NE	129.96	76.53 – 92.65	28.1
1	A	45	ARG	NE	115.07	76.53 – 92.65	18.9
1	A	49	TRP	NE1	109.45	118.53 – 139.98	-9.2
1	A	94	SER	HB2	1.99	2.61 – 5.13	-7.5
1	A	115	ASP	CB	31.34	32.98 – 48.76	-6.0
1	A	68	GLY	N	130.09	91.59 – 127.52	5.7
1	A	85	ASP	CB	31.98	32.98 – 48.76	-5.6
1	A	78	ARG	HB2	0.38	0.52 – 3.08	-5.5
1	A	21	LYS	HB2	0.46	0.58 – 2.97	-5.5
1	A	124	GLY	N	129.39	91.59 – 127.52	5.5
1	A	75	GLY	N	129.14	91.59 – 127.52	5.5
1	A	77	GLY	N	128.23	91.59 – 127.52	5.2

#### 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. 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:

