



Full wwPDB NMR Structure Validation Report ⓘ

Oct 24, 2016 – 12:31 PM EDT

PDB ID : 5TBN
Title : Solution NMR structure of PHF20 PHD domain in complex with a histone H3K4me2 peptide
Authors : Cui, G.; Botuyan, M.V.; Mer, G.
Deposited on : 2016-09-12

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

<http://wwpdb.org/validation/2016/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
Mogul : 1.7.1 (RC1), CSD as537be (2016)
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : rb-20027939
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027939

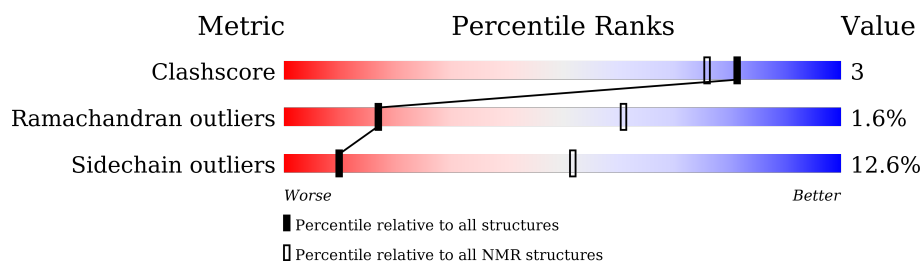
1 Overall quality at a glance

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	57	
2	C	12	

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:4-A:53, C:1-C:3, C:5-C:5 (54)	0.10	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 4 clusters and 2 single-model clusters were found.

Cluster number	Models
1	1, 4, 11, 14, 17, 18, 19
2	6, 8, 10, 13, 15, 16
3	2, 9, 20
4	3, 5
Single-model clusters	7; 12

3 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 1078 atoms, of which 517 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PHD finger protein 20.

Mol	Chain	Residues	Atoms						Trace
1	A	57	Total	C	H	N	O	S	0
			883	288	413	76	95	11	

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP Q9BVI0
A	-2	HIS	-	expression tag	UNP Q9BVI0
A	-1	MET	-	expression tag	UNP Q9BVI0

- Molecule 2 is a protein called Histone H3.1.

Mol	Chain	Residues	Atoms					Trace
2	C	12	Total	C	H	N	O	1
			193	52	104	21	16	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	12	NH2	-	amidation	UNP P68431

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

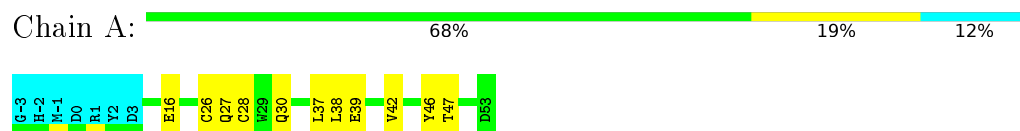
Mol	Chain	Residues	Atoms	
3	A	2	Total	Zn
			2	2

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: PHD finger protein 20



- Molecule 2: Histone H3.1

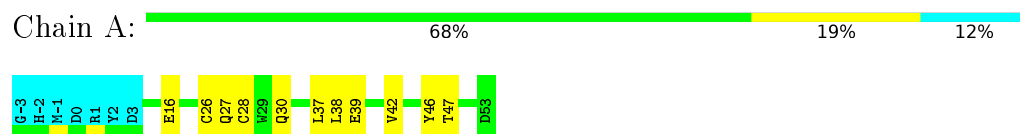


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: PHD finger protein 20

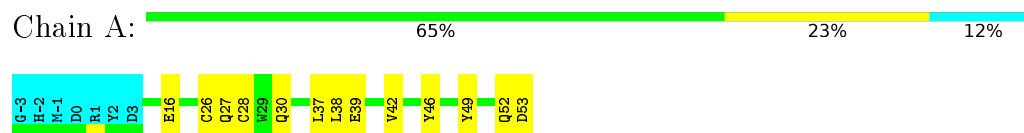


- Molecule 2: Histone H3.1



4.2.2 Score per residue for model 2

- Molecule 1: PHD finger protein 20

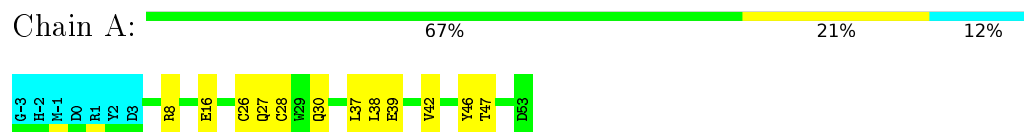


- Molecule 2: Histone H3.1



4.2.3 Score per residue for model 3

- Molecule 1: PHD finger protein 20

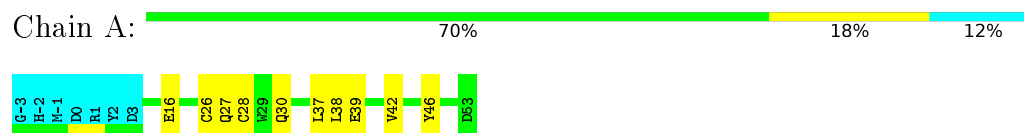


- Molecule 2: Histone H3.1



4.2.4 Score per residue for model 4

- Molecule 1: PHD finger protein 20



- Molecule 2: Histone H3.1



4.2.5 Score per residue for model 5

- Molecule 1: PHD finger protein 20

Chain A:  68% 19% 12%



- Molecule 2: Histone H3.1

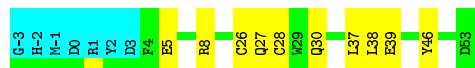
Chain C:  17% 17% 67%



4.2.6 Score per residue for model 6

- Molecule 1: PHD finger protein 20

Chain A:  70% 18% 12%



- Molecule 2: Histone H3.1

Chain C:  8% 17% 8% 67%



4.2.7 Score per residue for model 7

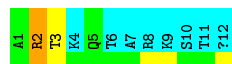
- Molecule 1: PHD finger protein 20

Chain A:  67% 21% 12%



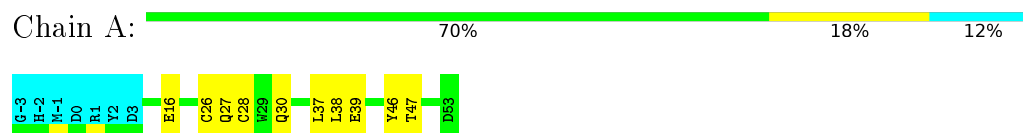
- Molecule 2: Histone H3.1

Chain C:  17% 8% 8% 67%



4.2.8 Score per residue for model 8

- Molecule 1: PHD finger protein 20

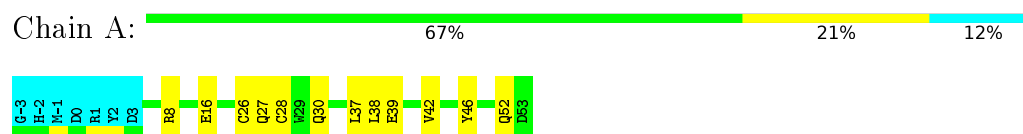


- Molecule 2: Histone H3.1



4.2.9 Score per residue for model 9

- Molecule 1: PHD finger protein 20

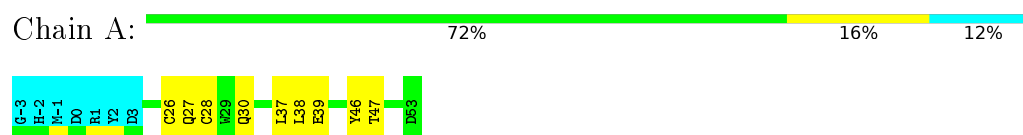


- Molecule 2: Histone H3.1



4.2.10 Score per residue for model 10

- Molecule 1: PHD finger protein 20

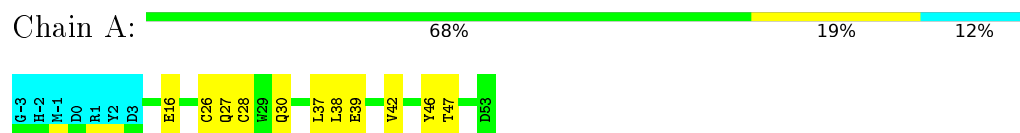


- Molecule 2: Histone H3.1



4.2.11 Score per residue for model 11

- Molecule 1: PHD finger protein 20

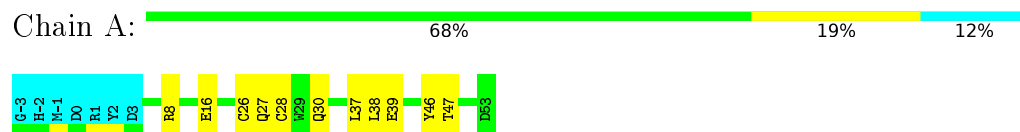


- Molecule 2: Histone H3.1



4.2.12 Score per residue for model 12

- Molecule 1: PHD finger protein 20

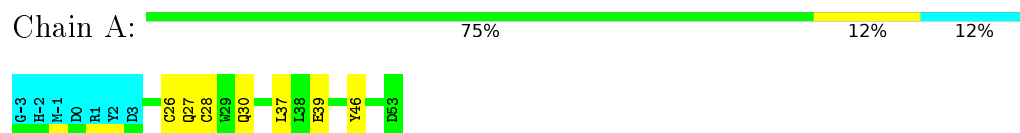


- Molecule 2: Histone H3.1



4.2.13 Score per residue for model 13

- Molecule 1: PHD finger protein 20



- Molecule 2: Histone H3.1



4.2.14 Score per residue for model 14

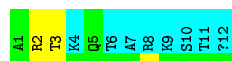
- Molecule 1: PHD finger protein 20

Chain A:  70% 18% 12%



- Molecule 2: Histone H3.1

Chain C:  17% 17% 67%



4.2.15 Score per residue for model 15

- Molecule 1: PHD finger protein 20

Chain A:  70% 18% 12%



- Molecule 2: Histone H3.1

Chain C:  17% 17% 67%



4.2.16 Score per residue for model 16

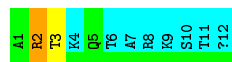
- Molecule 1: PHD finger protein 20

Chain A:  72% 16% 12%



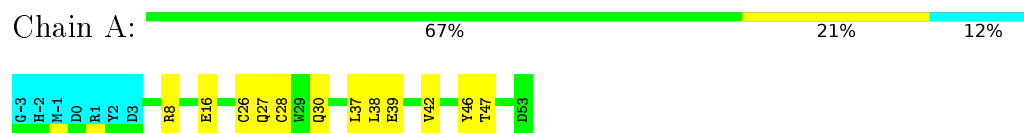
- Molecule 2: Histone H3.1

Chain C:  17% 8% 8% 67%



4.2.17 Score per residue for model 17

- Molecule 1: PHD finger protein 20

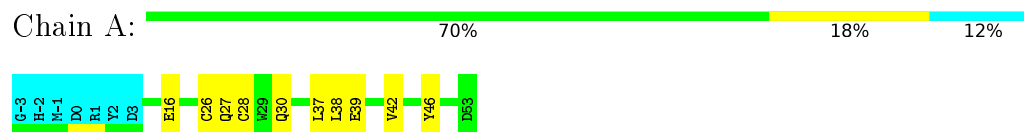


- Molecule 2: Histone H3.1



4.2.18 Score per residue for model 18

- Molecule 1: PHD finger protein 20

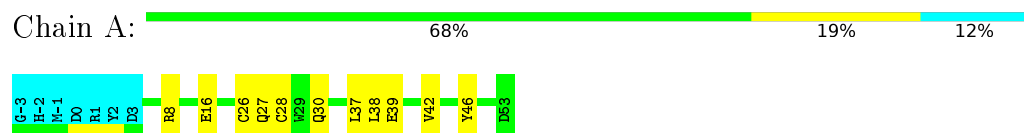


- Molecule 2: Histone H3.1



4.2.19 Score per residue for model 19

- Molecule 1: PHD finger protein 20



- Molecule 2: Histone H3.1



4.2.20 Score per residue for model 20

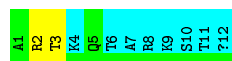
- Molecule 1: PHD finger protein 20

Chain A:  68% 19% 12%



- Molecule 2: Histone H3.1

Chain C:  17% 17% 67%



5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 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
AMBER	refinement	14

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)	5tbn_cs.cif
Number of chemical shift lists	1
Total number of shifts	805
Number of shifts mapped to atoms	805
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.

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MLY, NH2

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	0.74±0.00	0±0/416 (0.0±0.0%)	1.20±0.01	0±0/560 (0.1±0.1%)
2	C	0.96±0.07	0±0/31 (0.0±0.0%)	1.63±0.25	1±0/40 (1.5±1.2%)
All	All	0.76	0/8940 (0.0%)	1.24	21/12000 (0.2%)

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
2	C	2	ARG	NE-CZ-NH1	7.12	123.86	120.30	4	11
2	C	2	ARG	NE-CZ-NH2	6.24	123.42	120.30	6	1
1	A	8	ARG	NE-CZ-NH1	5.43	123.01	120.30	3	8
1	A	21	ILE	CA-CB-CG2	5.04	120.97	110.90	7	1

There are no chirality outliers.

There are no planarity outliers.

6.2 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	409	362	362	3±1
2	C	32	35	35	1±0
All	All	8860	7940	7940	54

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 3.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:42:VAL:HG21	2:C:3:THR:HG21	0.47	1.86	18	12
1:A:26:CYS:SG	1:A:28:CYS:SG	0.45	3.06	18	20
1:A:28:CYS:SG	1:A:30:GLN:NE2	0.42	2.92	15	20
1:A:49:TYR:HA	1:A:53:ASP:H	0.41	1.76	7	2

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	49/57 (86%)	44±1 (89±1%)	5±1 (11±1%)	0±0 (0±0%)	100	100
2	C	3/12 (25%)	2±1 (70±35%)	0±0 (2±7%)	1±1 (28±35%)	0	1
All	All	1040/1380 (75%)	913 (88%)	110 (11%)	17 (2%)	17	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)
2	C	3	THR	8
2	C	2	ARG	8
2	C	5	GLN	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	48/54 (89%)	42±1 (87±2%)	6±1 (13±2%)	9	50

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	C	3/8 (38%)	3±0 (97±10%)	0±0 (3±10%)	49 89
All	All	1020/1240 (82%)	891 (87%)	129 (13%)	10 52

All 10 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	46	TYR	20
1	A	37	LEU	20
1	A	27	GLN	20
1	A	39	GLU	20
1	A	38	LEU	18
1	A	16	GLU	14
1	A	47	THR	11
1	A	52	GLN	3
2	C	5	GLN	2
1	A	5	GLU	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	MLY	C	4	2	8,10,11	0.92±0.04	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types,

if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	MLY	C	4	2	9,11,13	0.81±0.17	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	MLY	C	4	2	-	0±0,7,9,11	0±0,0,0,0

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 86% for the entire structure.

7.1 Chemical shift list 1

File name: 5tbn_cs.cif

Chemical shift list name: *assigned_chem_shift_list_0*

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	805
Number of shifts mapped to atoms	805
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	68	-0.49 ± 0.45	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	65	0.18 ± 0.27	None needed (< 0.5 ppm)
$^{13}\text{C}'$	55	0.41 ± 0.19	None needed (< 0.5 ppm)
^{15}N	52	-0.53 ± 0.62	None needed (imprecise)

7.1.3 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. 606 atoms were assigned a chemical shift out of a possible 664. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	254/268 (95%)	102/107 (95%)	104/108 (96%)	48/53 (91%)
Sidechain	300/343 (87%)	188/201 (94%)	104/127 (82%)	8/15 (53%)

Continued on next page...

Continued from previous page...

	Total	¹H	¹³C	¹⁵N
Aromatic	52/53 (98%)	27/28 (96%)	23/23 (100%)	2/2 (100%)
Overall	606/664 (91%)	317/336 (94%)	231/258 (90%)	58/70 (83%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 86%, i.e. 716 atoms were assigned a chemical shift out of a possible 828. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹H	¹³C	¹⁵N
Backbone	293/333 (88%)	119/133 (89%)	122/134 (91%)	52/66 (79%)
Sidechain	359/427 (84%)	226/251 (90%)	125/154 (81%)	8/22 (36%)
Aromatic	64/68 (94%)	33/36 (92%)	29/29 (100%)	2/3 (67%)
Overall	716/828 (86%)	378/420 (90%)	276/317 (87%)	62/91 (68%)

7.1.4 Statistically unusual chemical shifts ⓘ

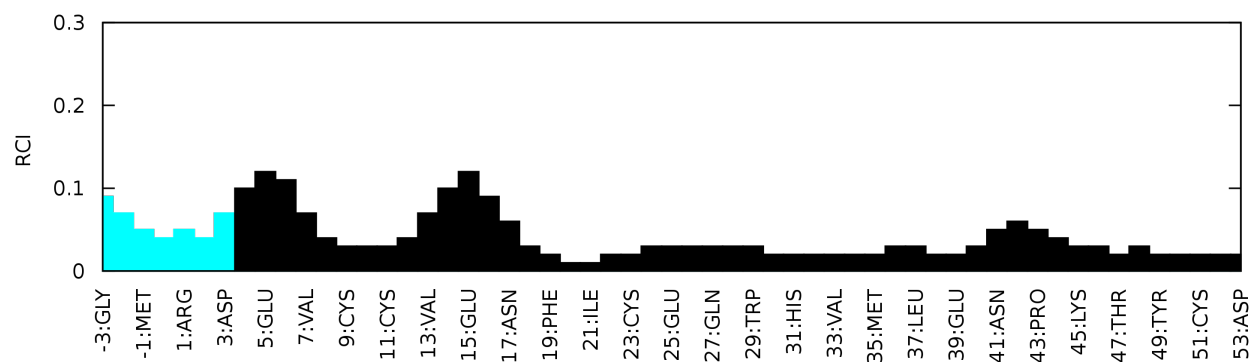
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	22	GLN	HB3	-1.50	3.37 – 0.67	-13.0
1	A	22	GLN	HG3	0.37	3.75 – 0.85	-6.7

7.1.5 Random Coil Index (RCI) plots ⓘ

The images below report *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:



Random coil index (RCI) for chain C:

