



# Full wwPDB NMR Structure Validation Report ⓘ

Feb 13, 2017 – 01:19 am GMT

PDB ID : 2MSR  
Title : Solution structure of LEDGF/p75 IBD in complex with MLL1 peptide (140-160)  
Authors : Cermakova, K.; Tesina, P.; Demeulemeester, J.; El Ashkar, S.; Mereau, H.; Schwaller, J.; Rezacova, P.; Veverka, V.; De Rijck, J.  
Deposited on : 2014-08-05

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

<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  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : trunk28760  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

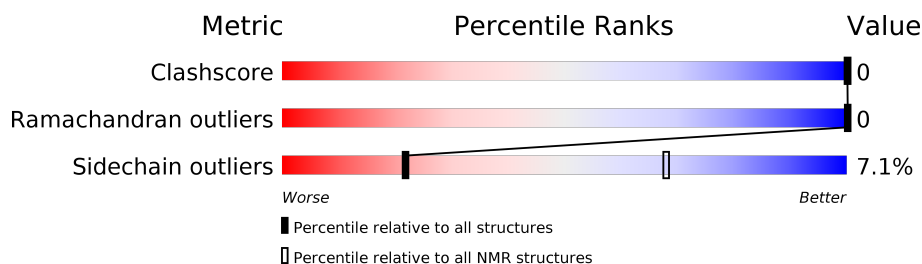
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 86%.

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	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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	21	<div style="background-color: cyan; width: 100%; text-align: center;">100%</div>
2	B	88	<div style="display: flex; align-items: center;"> <div style="background-color: green; width: 85%;"></div> <div style="background-color: yellow; width: 13%;"></div> <div style="background-color: red; width: 2%;"></div> <div style="margin-left: 10px;">•</div> </div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>85%</span> <span>13%</span> </div>

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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	B:348-B:424 (77)	0.31	2

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 3 clusters and 8 single-model clusters were found.

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

### 3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1750 atoms, of which 880 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Histone-lysine N-methyltransferase 2A.

Mol	Chain	Residues	Atoms					Trace
1	A	21	Total	C	H	N	O	0
			296	95	137	28	36	

- Molecule 2 is a protein called PC4 and SFRS1-interacting protein.

Mol	Chain	Residues	Atoms						Trace
2	B	88	Total	C	H	N	O	S	0
			1454	438	743	131	135	7	

There are 5 discrepancies between the modelled and reference sequences:

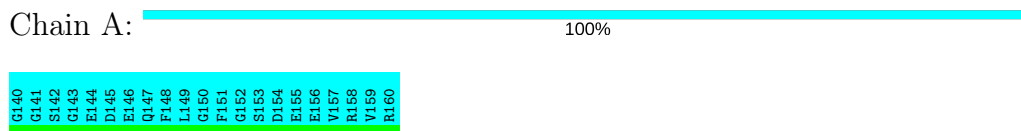
Chain	Residue	Modelled	Actual	Comment	Reference
B	339	SER	-	EXPRESSION TAG	UNP O75475
B	340	ASN	-	EXPRESSION TAG	UNP O75475
B	341	ALA	-	EXPRESSION TAG	UNP O75475
B	342	ALA	-	EXPRESSION TAG	UNP O75475
B	343	SER	-	EXPRESSION TAG	UNP O75475

## 4 Residue-property plots

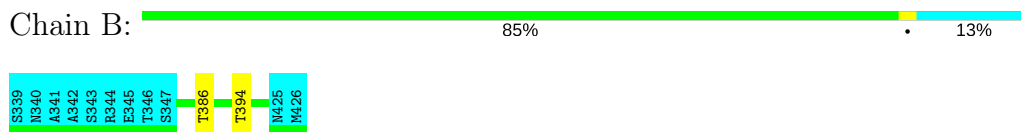
### 4.1 Average score per residue in the NMR ensemble

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: Histone-lysine N-methyltransferase 2A



- Molecule 2: PC4 and SFRS1-interacting protein

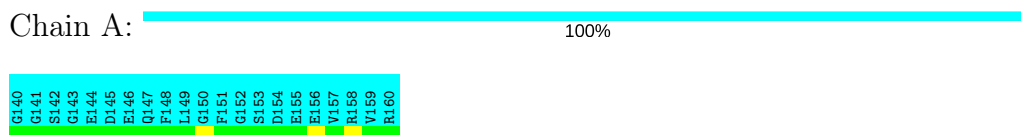


### 4.2 Scores per residue for each member of the ensemble

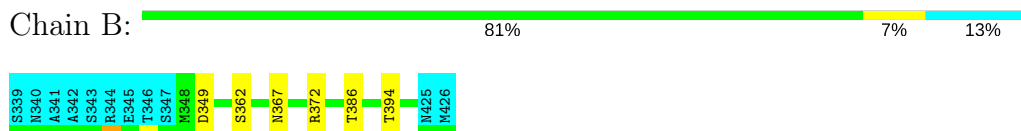
Colouring as in section 4.1 above.

#### 4.2.1 Score per residue for model 1

- Molecule 1: Histone-lysine N-methyltransferase 2A



- Molecule 2: PC4 and SFRS1-interacting protein




### 4.2.2 Score per residue for model 2 (medoid)

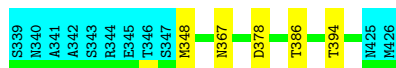
- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

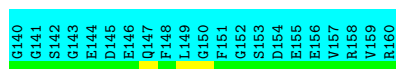
Chain B:  82% 6% 13%




### 4.2.3 Score per residue for model 3

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

Chain B:  81% 7% 13%



### 4.2.4 Score per residue for model 4

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



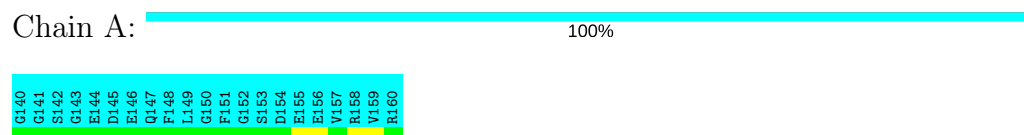
- Molecule 2: PC4 and SFRS1-interacting protein

Chain B:  80% 8% 13%

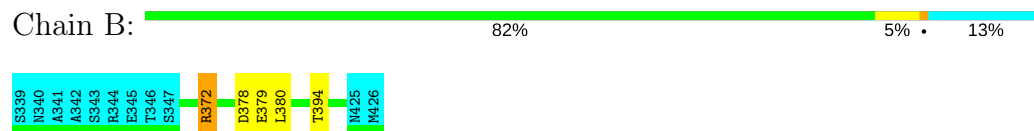


### 4.2.5 Score per residue for model 5

- Molecule 1: Histone-lysine N-methyltransferase 2A

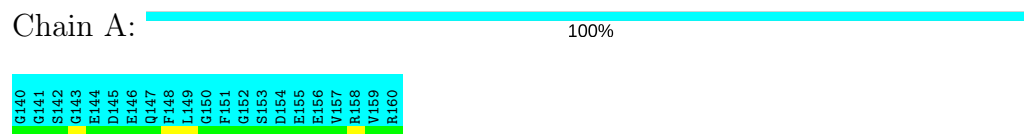


- Molecule 2: PC4 and SFRS1-interacting protein

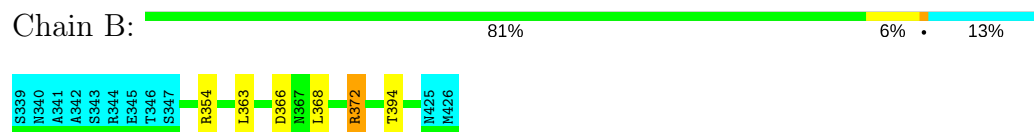


### 4.2.6 Score per residue for model 6

- Molecule 1: Histone-lysine N-methyltransferase 2A

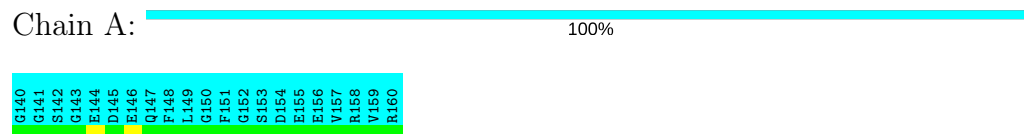


- Molecule 2: PC4 and SFRS1-interacting protein

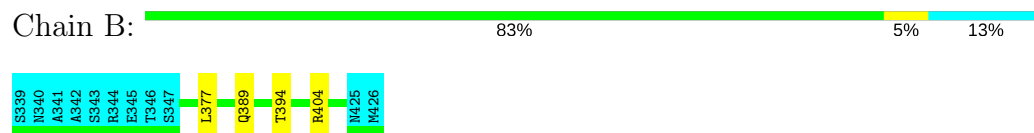


### 4.2.7 Score per residue for model 7

- Molecule 1: Histone-lysine N-methyltransferase 2A



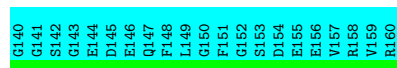
- Molecule 2: PC4 and SFRS1-interacting protein




### 4.2.8 Score per residue for model 8

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

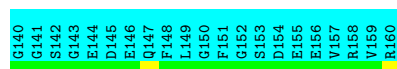
Chain B:  81% 6% 13%




### 4.2.9 Score per residue for model 9

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

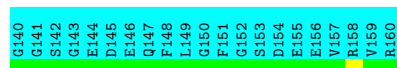
Chain B:  81% 7% 13%




### 4.2.10 Score per residue for model 10

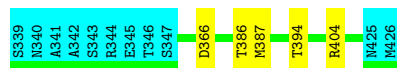
- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

Chain B:  82% 6% 13%

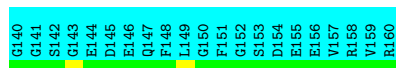




### 4.2.11 Score per residue for model 11

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

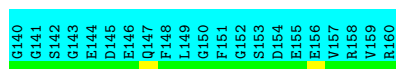
Chain B:  78% 9% 13%




### 4.2.12 Score per residue for model 12

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

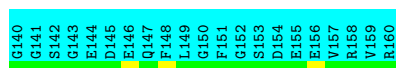
Chain B:  82% 6% 13%



### 4.2.13 Score per residue for model 13

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

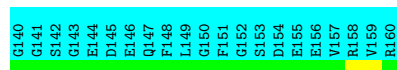
Chain B:  82% 6% 13%




#### 4.2.14 Score per residue for model 14

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

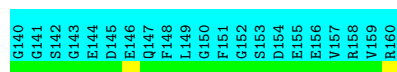
Chain B:  80% 8% 13%




#### 4.2.15 Score per residue for model 15

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

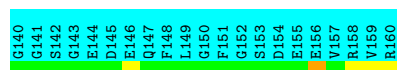
Chain B:  82% 6% 13%




#### 4.2.16 Score per residue for model 16

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

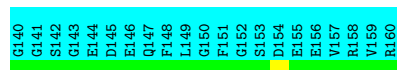
Chain B:  78% 9% 13%




#### 4.2.17 Score per residue for model 17

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

Chain B:  76% 11% 13%




#### 4.2.18 Score per residue for model 18

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



- Molecule 2: PC4 and SFRS1-interacting protein

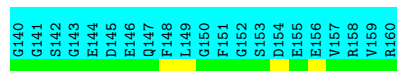
Chain B:  82% 6% 13%




#### 4.2.19 Score per residue for model 19

- Molecule 1: Histone-lysine N-methyltransferase 2A

Chain A:  100%



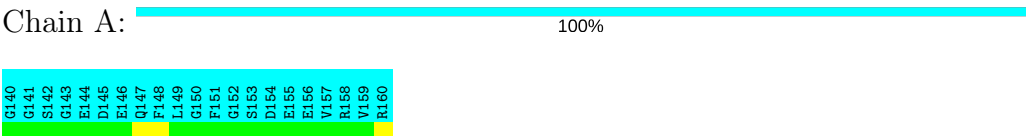
- Molecule 2: PC4 and SFRS1-interacting protein

Chain B:  83% 5% 13%

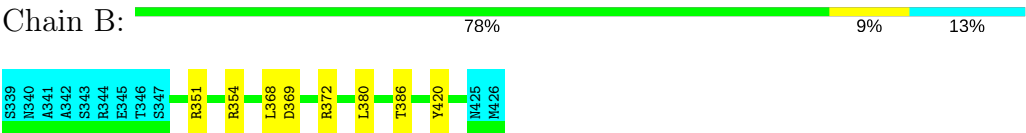


4.2.20 Score per residue for model 20

- Molecule 1: Histone-lysine N-methyltransferase 2A



- Molecule 2: PC4 and SFRS1-interacting protein



## 5 Refinement protocol and experimental data overview

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

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

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

Software name	Classification	Version
CYANA	structure solution	
YASARA	refinement	
CYANA	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)	2msr_cs.str
Number of chemical shift lists	1
Total number of shifts	1195
Number of shifts mapped to atoms	1195
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	86%

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

## 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
2	B	0.66±0.16	0±0/637 (0.0±0.1%)	0.69±0.07	1±1/850 (0.1±0.1%)
All	All	0.67	3/12740 (0.0%)	0.70	17/17000 (0.1%)

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
2	B	375	GLU	CD-OE1	-5.20	1.20	1.25	16	1
2	B	420	TYR	CG-CD2	5.16	1.45	1.39	20	1
2	B	420	TYR	CD2-CE2	5.06	1.47	1.39	16	1

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	B	404	ARG	NE-CZ-NH1	6.13	123.37	120.30	13	5
2	B	372	ARG	NE-CZ-NH1	6.08	123.34	120.30	20	5
2	B	354	ARG	NE-CZ-NH1	5.88	123.24	120.30	14	5
2	B	351	ARG	NE-CZ-NH1	5.66	123.13	120.30	20	2

There are no chirality outliers.

There are no planarity outliers.

### 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
All	All	12640	13440	13440	-

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 -.

There are no clashes.

## 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	0	-	-	-	-	
2	B	77/88 (88%)	74±1 (96±1%)	3±1 (4±1%)	0±0 (0±0%)	100	100
All	All	1540/2180 (71%)	1485 (96%)	55 (4%)	0 (0%)	100	100

There are no Ramachandran outliers.

### 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	0	-	-	-	
2	B	73/82 (89%)	68±1 (93±2%)	5±1 (7±2%)	22	68
All	All	1460/1960 (74%)	1356 (93%)	104 (7%)	22	68

All 28 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)
2	B	394	THR	19
2	B	386	THR	14

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
2	B	367	ASN	8
2	B	380	LEU	7
2	B	378	ASP	5
2	B	372	ARG	5
2	B	366	ASP	4
2	B	368	LEU	4
2	B	354	ARG	3
2	B	418	MET	3
2	B	424	LYS	3
2	B	406	PHE	3
2	B	349	ASP	3
2	B	405	ARG	3
2	B	387	MET	2
2	B	389	GLN	2
2	B	363	LEU	2
2	B	379	GLU	2
2	B	351	ARG	2
2	B	369	ASP	2
2	B	362	SER	1
2	B	404	ARG	1
2	B	410	GLN	1
2	B	388	GLN	1
2	B	353	GLN	1
2	B	348	MET	1
2	B	408	VAL	1
2	B	377	LEU	1

### 6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

### 6.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.



## 6.6 Ligand geometry

There are no ligands in this entry.

## 6.7 Other polymers

There are no such molecules in this entry.

## 6.8 Polymer linkage issues

There are no chain breaks in this entry.

## 7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 86% for the well-defined parts and 76% for the entire structure.

### 7.1 Chemical shift list 1

File name: 2msr\_cs.str

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

#### 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	1195
Number of shifts mapped to atoms	1195
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

#### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	87	$-0.52 \pm 0.08$	Should be applied
$^{13}\text{C}_\beta$	87	$0.37 \pm 0.12$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	82	$-0.12 \pm 0.14$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	83	$-0.19 \pm 0.39$	None needed ( $< 0.5$ ppm)

#### 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 86%, i.e. 886 atoms were assigned a chemical shift out of a possible 1031. 12 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	380/385 (99%)	152/154 (99%)	152/154 (99%)	76/77 (99%)
Sidechain	491/606 (81%)	304/355 (86%)	182/217 (84%)	5/34 (15%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	15/40 (38%)	14/22 (64%)	1/16 (6%)	0/2 (0%)
Overall	886/1031 (86%)	470/531 (89%)	335/387 (87%)	81/113 (72%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 76%, i.e. 1057 atoms were assigned a chemical shift out of a possible 1388. 12 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	454/545 (83%)	202/218 (93%)	169/218 (78%)	83/109 (76%)
Sidechain	578/785 (74%)	376/460 (82%)	196/279 (70%)	6/46 (13%)
Aromatic	25/58 (43%)	24/32 (75%)	1/24 (4%)	0/2 (0%)
Overall	1057/1388 (76%)	602/710 (85%)	366/521 (70%)	89/157 (57%)

#### 7.1.4 Statistically unusual chemical shifts ⓘ

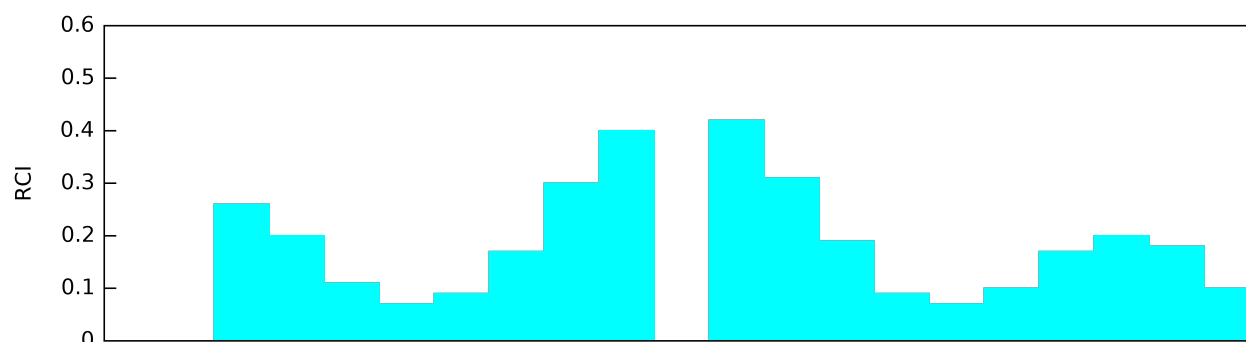
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
2	B	360	LYS	HE3	1.93	3.86 – 1.96	-5.2

#### 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 B:

