



# Full wwPDB NMR Structure Validation Report ⓘ

Apr 26, 2016 – 04:39 PM BST

PDB ID : 1RL1  
Title : Solution structure of human Sgt1 CS domain  
Authors : Lee, Y.-T.; Jacob, J.; Michowski, W.; Nowotny, M.; Kuznicki, J.; Chazin, W.J.  
Deposited on : 2003-11-24

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  
Mogul : unknown  
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-20027457  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027457

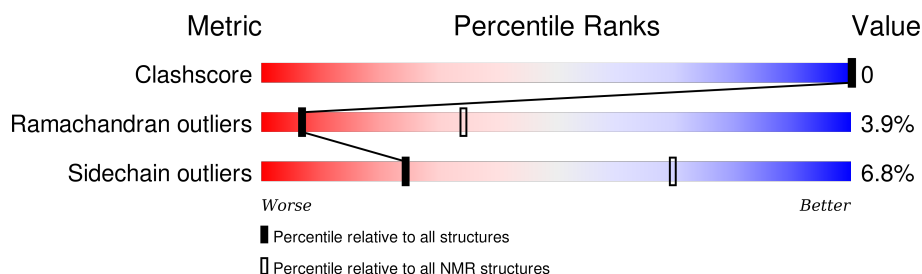
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 88%.

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  $\geq 3$ , 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	114	 71% 6% • 19%

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 18 is the overall representative, medoid model (most similar to other models). The authors have identified model 19 as representative, based on the following criterion: *lowest energy*.

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:140-A:227 (88)	0.43	18

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 6 clusters and 5 single-model clusters were found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1396 atoms, of which 644 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Suppressor of G2 allele of SKP1 homolog.

Mol	Chain	Residues	Atoms						Trace
1	A	92	Total	C	H	N	O	S	0
			1396	487	644	121	143	1	

There are 4 discrepancies between the modelled and reference sequences:

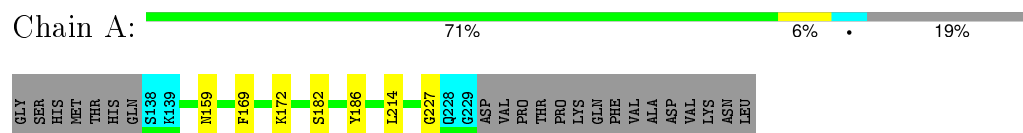
Chain	Residue	Modelled	Actual	Comment	Reference
A	131	GLY	-	CLONING ARTIFACT	UNP Q9Y2Z0
A	132	SER	-	CLONING ARTIFACT	UNP Q9Y2Z0
A	133	HIS	-	CLONING ARTIFACT	UNP Q9Y2Z0
A	134	MET	-	CLONING ARTIFACT	UNP Q9Y2Z0

## 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: Suppressor of G2 allele of SKP1 homolog

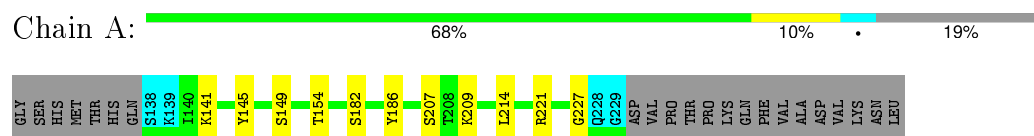


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

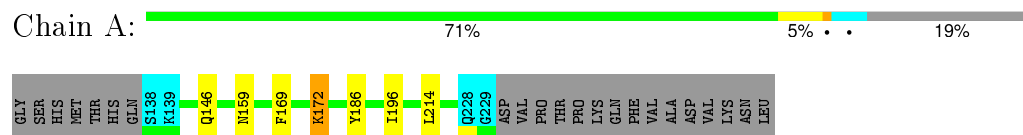
#### 4.2.1 Score per residue for model 1

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



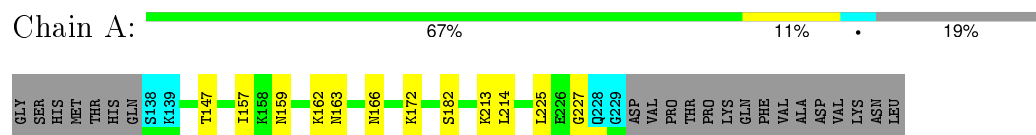
#### 4.2.2 Score per residue for model 2

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



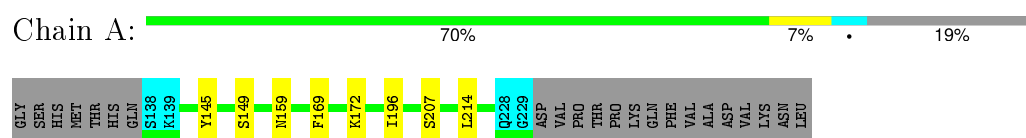
### 4.2.3 Score per residue for model 3

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



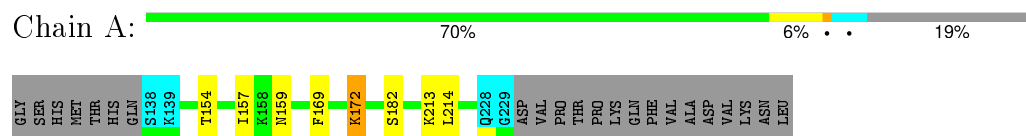
### 4.2.4 Score per residue for model 4

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



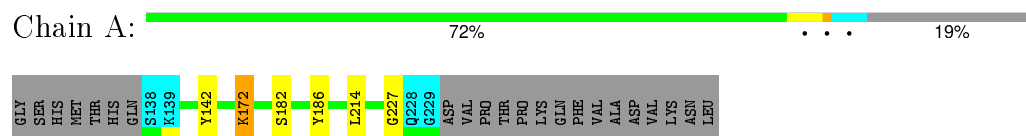
### 4.2.5 Score per residue for model 5

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



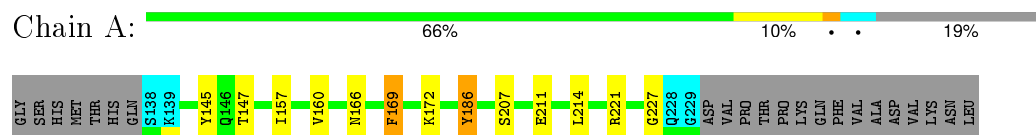
### 4.2.6 Score per residue for model 6

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



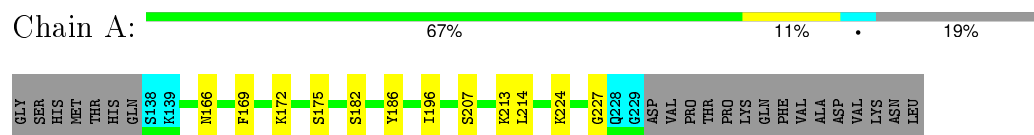
### 4.2.7 Score per residue for model 7

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



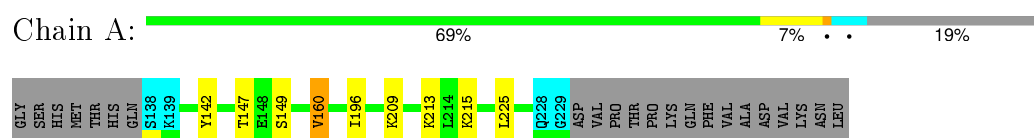
### 4.2.8 Score per residue for model 8

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



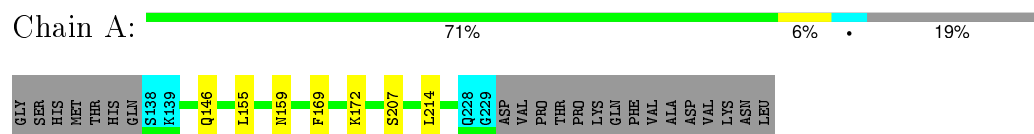
### 4.2.9 Score per residue for model 9

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



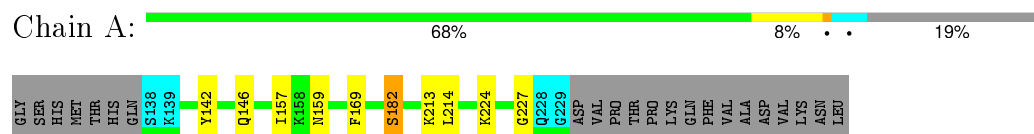
### 4.2.10 Score per residue for model 10

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



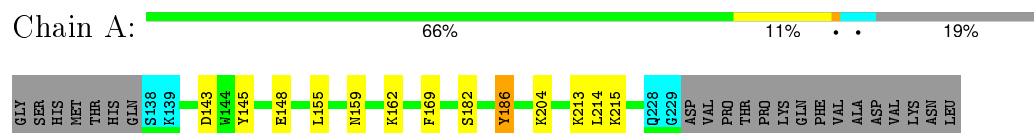
### 4.2.11 Score per residue for model 11

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



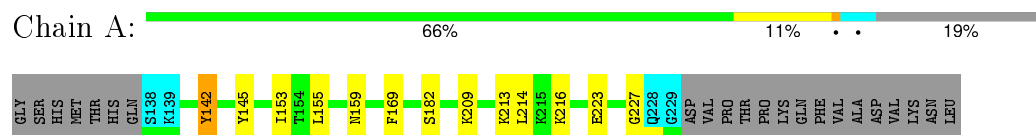
### 4.2.12 Score per residue for model 12

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



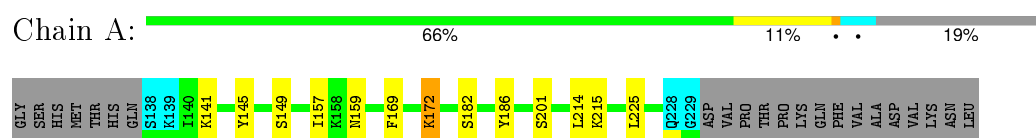
### 4.2.13 Score per residue for model 13

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



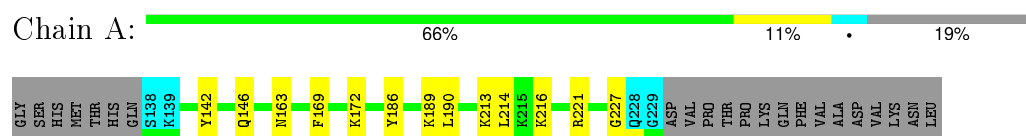
### 4.2.14 Score per residue for model 14

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



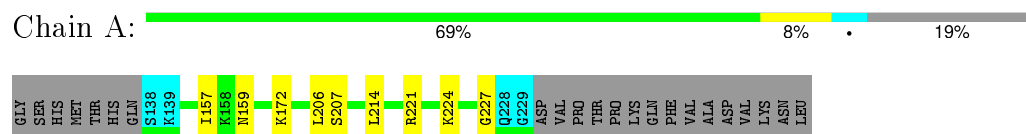
### 4.2.15 Score per residue for model 15

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



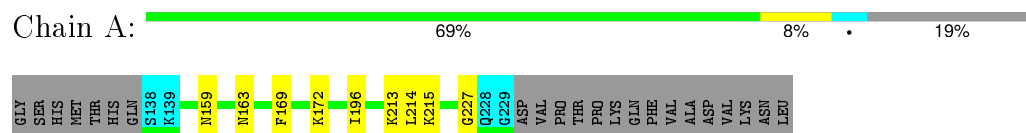
### 4.2.16 Score per residue for model 16

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



### 4.2.17 Score per residue for model 17

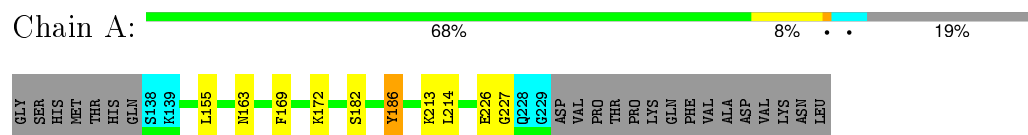
- Molecule 1: Suppressor of G2 allele of SKP1 homolog





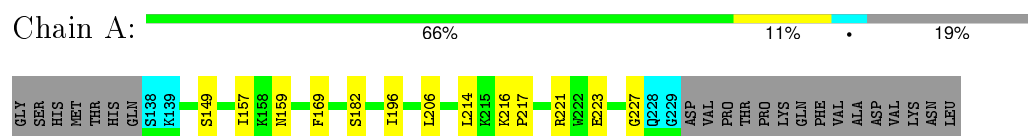
#### 4.2.18 Score per residue for model 18 (medoid)

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



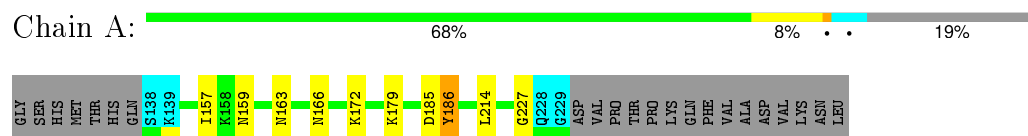
#### 4.2.19 Score per residue for model 19

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



#### 4.2.20 Score per residue for model 20

- Molecule 1: Suppressor of G2 allele of SKP1 homolog



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *Distance Geometry and restrained Molecular Dynamics*.

Of the 50 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
DYANA	structure solution	1.5
AMBER	refinement	7

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)	BMRB entry 6012
Number of chemical shift lists	1
Total number of shifts	1361
Number of shifts mapped to atoms	1361
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	88%

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

## 6 Model quality

### 6.1 Standard geometry

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	0.53±0.00	0±0/738 (0.0±0.0%)	0.90±0.02	0±1/998 (0.0±0.1%)
All	All	0.53	0/14760 (0.0%)	0.90	5/19960 (0.0%)

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	1.4±1.0
All	All	0	28

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	186	TYR	CB-CG-CD2	-7.09	116.74	121.00	18	3
1	A	221	ARG	NE-CZ-NH2	-5.41	117.60	120.30	1	1
1	A	142	TYR	CB-CG-CD2	-5.28	117.83	121.00	6	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	169	PHE	Sidechain	14
1	A	186	TYR	Sidechain	7
1	A	145	TYR	Sidechain	5
1	A	142	TYR	Sidechain	2

## 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	724	622	756	0±0
All	All	14480	12440	15120	1

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:142:TYR:CE1	1:A:153:ILE:CG2	0.40	3.05	13	1

## 6.3 Torsion angles ⓘ

### 6.3.1 Protein backbone ⓘ

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	88/114 (77%)	75±2 (86±2%)	9±2 (10±2%)	3±2 (4±2%)	7	34
All	All	1760/2280 (77%)	1508 (86%)	184 (10%)	68 (4%)	7	34

All 12 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	172	LYS	13
1	A	227	GLY	13
1	A	182	SER	9
1	A	157	ILE	8
1	A	159	ASN	7
1	A	207	SER	6
1	A	149	SER	4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	225	LEU	3
1	A	196	ILE	2
1	A	160	VAL	1
1	A	223	GLU	1
1	A	217	PRO	1

### 6.3.2 Protein sidechains ⓘ

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	84/107 (79%)	78±2 (93±2%)	6±2 (7±2%)	24	70
All	All	1680/2140 (79%)	1565 (93%)	115 (7%)	24	70

All 38 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	214	LEU	19
1	A	213	LYS	10
1	A	159	ASN	6
1	A	172	LYS	5
1	A	163	ASN	5
1	A	221	ARG	4
1	A	146	GLN	4
1	A	155	LEU	4
1	A	215	LYS	4
1	A	186	TYR	4
1	A	196	ILE	4
1	A	166	ASN	4
1	A	224	LYS	3
1	A	147	THR	3
1	A	209	LYS	3
1	A	182	SER	3
1	A	216	LYS	3
1	A	160	VAL	2
1	A	162	LYS	2
1	A	206	LEU	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	154	THR	2
1	A	141	LYS	2
1	A	142	TYR	2
1	A	204	LYS	1
1	A	148	GLU	1
1	A	179	LYS	1
1	A	190	LEU	1
1	A	149	SER	1
1	A	201	SER	1
1	A	226	GLU	1
1	A	145	TYR	1
1	A	175	SER	1
1	A	223	GLU	1
1	A	169	PHE	1
1	A	211	GLU	1
1	A	185	ASP	1
1	A	143	ASP	1
1	A	189	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 carbohydrates 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 ⓘ

There are no chain breaks in this entry.

## 7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 88% for the well-defined parts and 87% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 6012

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	1361
Number of shifts mapped to atoms	1361
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
$^{13}\text{C}_\alpha$	110	$-0.08 \pm 0.10$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	107	$-0.19 \pm 0.12$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	101	$0.54 \pm 0.16$	Should be applied
$^{15}\text{N}$	103	$0.92 \pm 0.39$	Should be applied

#### 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 88%, i.e. 1015 atoms were assigned a chemical shift out of a possible 1152. 19 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	428/432 (99%)	172/172 (100%)	172/176 (98%)	84/84 (100%)
Sidechain	521/647 (81%)	304/378 (80%)	210/246 (85%)	7/23 (30%)

*Continued on next page...*



Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	66/73 (90%)	34/38 (89%)	30/32 (94%)	2/3 (67%)
Overall	1015/1152 (88%)	510/588 (87%)	412/454 (91%)	93/110 (85%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 87%, i.e. 1045 atoms were assigned a chemical shift out of a possible 1198. 19 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	442/452 (98%)	177/180 (98%)	178/184 (97%)	87/88 (99%)
Sidechain	537/673 (80%)	313/394 (79%)	216/254 (85%)	8/25 (32%)
Aromatic	66/73 (90%)	34/38 (89%)	30/32 (94%)	2/3 (67%)
Overall	1045/1198 (87%)	524/612 (86%)	424/470 (90%)	97/116 (84%)

#### 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	225	LEU	HB2	-0.18	3.32 – -0.08	-5.3

#### 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:

