



# Full wwPDB NMR Structure Validation Report ⓘ

Apr 26, 2016 – 08:47 PM BST

PDB ID : 2GO9  
Title : RRM domains 1 and 2 of Prp24 from *S. cerevisiae*  
Authors : Reiter, N.J.; Lee, D.H.; Tonelli, M.; Kwan, S.K.; Brow, D.A.; Butcher, S.E.  
Deposited on : 2006-04-12

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](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

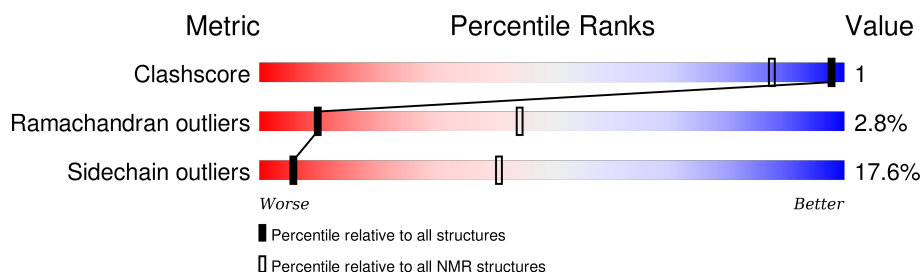
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 84%.

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	161	<div> <div>78%</div> <div>19%</div> <div>.</div> </div>

## 2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 9 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:3-A:158 (156)	1.18	9

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. No single-model clusters were found.

Cluster number	Models
1	2, 7, 9
2	3, 5, 8
3	4, 10
4	1, 6

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2656 atoms, of which 1346 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called U4/U6 snRNA-associated splicing factor PRP24.

Mol	Chain	Residues	Atoms						Trace
1	A	161	Total	C	H	N	O	S	0
			2656	839	1346	231	235	5	

There is a discrepancy between the modelled and reference sequences:

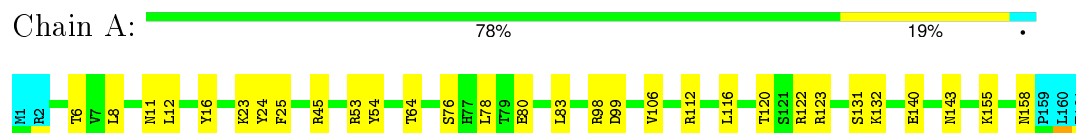
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING METHIONINE	UNP P49960

## 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: U4/U6 snRNA-associated splicing factor PRP24

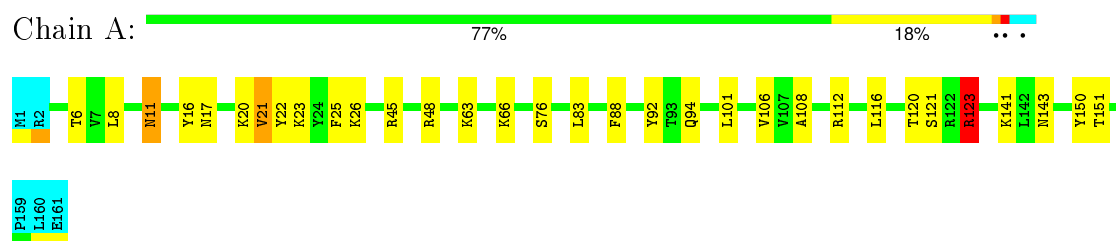


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

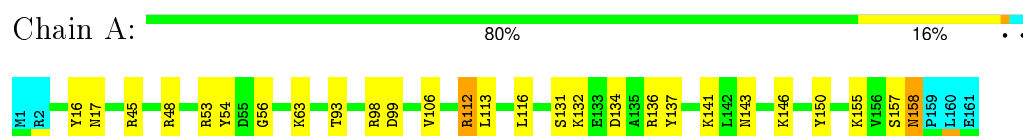
#### 4.2.1 Score per residue for model 1

- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24



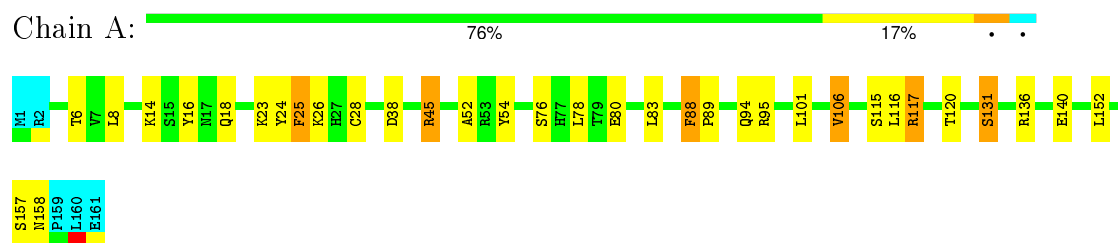
#### 4.2.2 Score per residue for model 2

- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24



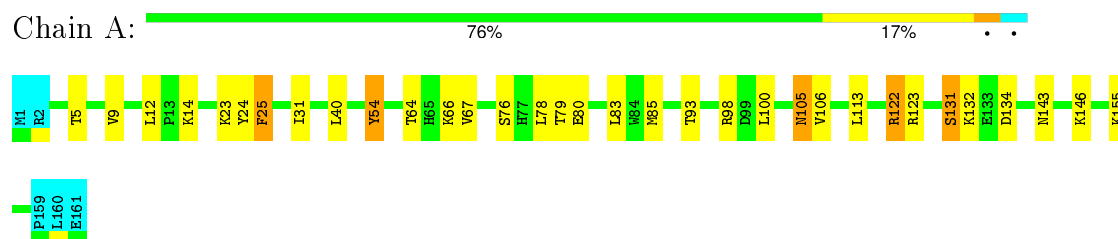
### 4.2.3 Score per residue for model 3

- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24



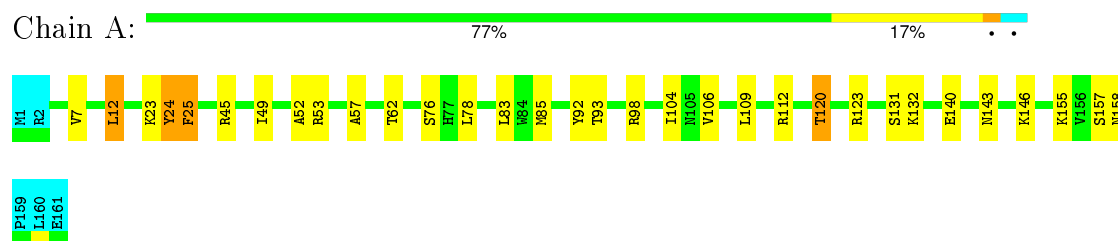
### 4.2.4 Score per residue for model 4

- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24



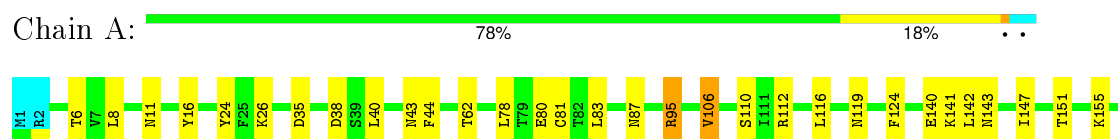
### 4.2.5 Score per residue for model 5

- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24



### 4.2.6 Score per residue for model 6

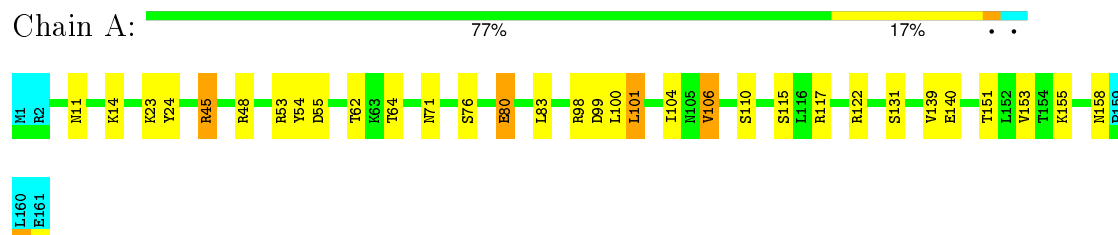
- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24





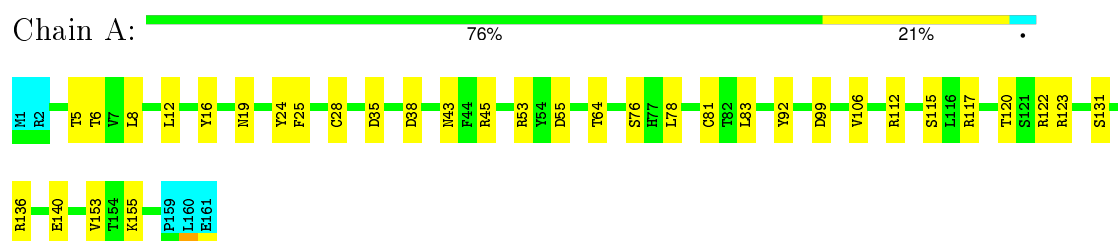
#### 4.2.7 Score per residue for model 7

- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24



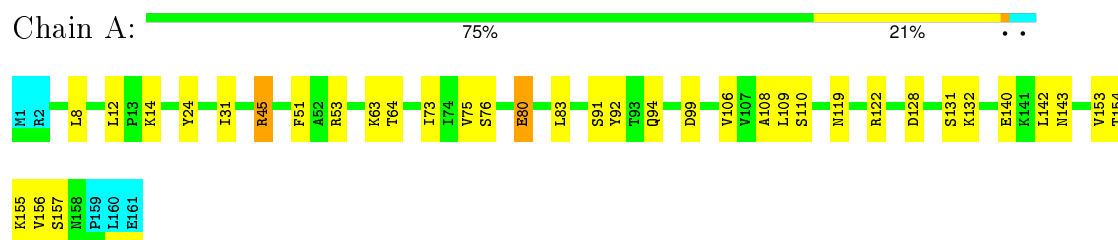
#### 4.2.8 Score per residue for model 8

- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24



#### 4.2.9 Score per residue for model 9 (medoid)

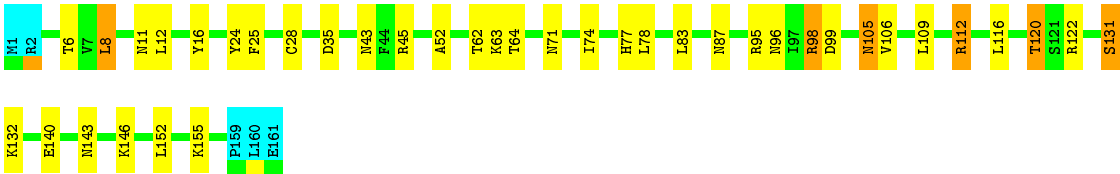
- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24



#### 4.2.10 Score per residue for model 10

- Molecule 1: U4/U6 snRNA-associated splicing factor PRP24







## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *TORSION ANGLE DYNAMICS*.

Of the 200 calculated structures, 10 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	refinement	2.1.4
CYANA	structure solution	2.1.4

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 7070
Number of chemical shift lists	1
Total number of shifts	1958
Number of shifts mapped to atoms	1958
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	84%

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	0.63±0.01	0±0/1293 (0.0±0.0%)	1.04±0.02	1±1/1753 (0.1±0.1%)
All	All	0.63	0/12930 (0.0%)	1.04	9/17530 (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	3.0±1.0
All	All	0	30

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	45	ARG	NE-CZ-NH2	-6.44	117.08	120.30	7	2
1	A	21	VAL	CA-CB-CG2	6.20	120.20	110.90	1	1
1	A	95	ARG	NE-CZ-NH1	5.60	123.10	120.30	10	1
1	A	123	ARG	NE-CZ-NH2	-5.39	117.60	120.30	1	1
1	A	112	ARG	NE-CZ-NH2	-5.33	117.64	120.30	10	1
1	A	64	THR	CA-CB-CG2	5.23	119.72	112.40	4	1
1	A	98	ARG	NE-CZ-NH2	-5.07	117.77	120.30	7	1
1	A	95	ARG	CD-NE-CZ	5.06	130.69	123.60	3	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	45	ARG	Sidechain	4
1	A	54	TYR	Sidechain	4
1	A	136	ARG	Sidechain	3
1	A	112	ARG	Sidechain	2
1	A	98	ARG	Sidechain	2
1	A	16	TYR	Sidechain	2
1	A	24	TYR	Sidechain	2
1	A	53	ARG	Sidechain	2
1	A	154	THR	Peptide	1
1	A	123	ARG	Sidechain	1
1	A	122	ARG	Sidechain	1
1	A	117	ARG	Sidechain	1
1	A	22	TYR	Sidechain	1
1	A	95	ARG	Sidechain	1
1	A	92	TYR	Sidechain	1
1	A	108	ALA	Peptide	1
1	A	137	TYR	Sidechain	1

## 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	1267	1300	1300	2±1
All	All	12670	13000	13000	18

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:88:PHE:CG	1:A:89:PRO:HD2	0.53	2.38	3	1
1:A:16:TYR:CD2	1:A:21:VAL:HG12	0.51	2.40	1	1
1:A:101:LEU:HD13	1:A:108:ALA:HB2	0.50	1.84	1	1
1:A:64:THR:HG21	1:A:74:ILE:HA	0.49	1.83	10	1
1:A:88:PHE:CG	1:A:152:LEU:HD21	0.49	2.43	3	1
1:A:64:THR:HB	1:A:75:VAL:HG22	0.48	1.86	9	1
1:A:31:ILE:HD11	1:A:51:PHE:CE2	0.48	2.44	9	1

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:53:ARG:HG3	1:A:56:GLY:H	0.47	1.69	2	1
1:A:8:LEU:HD21	1:A:109:LEU:HD21	0.44	1.89	10	1
1:A:101:LEU:C	1:A:101:LEU:HD22	0.43	2.34	7	1
1:A:25:PHE:CE2	1:A:49:ILE:HD13	0.42	2.49	5	1
1:A:9:VAL:HG11	1:A:12:LEU:HD21	0.42	1.91	4	1
1:A:147:ILE:HD12	1:A:147:ILE:N	0.42	2.29	6	1
1:A:106:VAL:HG13	1:A:131:SER:HB2	0.42	1.91	10	1
1:A:7:VAL:CG1	1:A:57:ALA:HB1	0.42	2.45	5	1
1:A:25:PHE:CD1	1:A:26:LYS:N	0.41	2.88	3	1
1:A:25:PHE:CD2	1:A:31:ILE:HD13	0.41	2.50	4	1
1:A:101:LEU:HA	1:A:104:ILE:HG22	0.41	1.93	7	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	156/161 (97%)	135±3 (87±2%)	17±4 (11±3%)	4±2 (3±1%)	10	44
All	All	1560/1610 (97%)	1351 (87%)	166 (11%)	43 (3%)	10	44

All 14 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	106	VAL	8
1	A	131	SER	7
1	A	11	ASN	4
1	A	12	LEU	4
1	A	122	ARG	3
1	A	52	ALA	3
1	A	120	THR	3
1	A	80	GLU	3
1	A	81	CYS	2
1	A	105	ASN	2
1	A	63	LYS	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	117	ARG	1
1	A	116	LEU	1
1	A	67	VAL	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	142/147 (97%)	117±2 (82±2%)	25±2 (18±2%)	6	41
All	All	1420/1470 (97%)	1170 (82%)	250 (18%)	6	41

All 85 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	83	LEU	9
1	A	155	LYS	8
1	A	76	SER	7
1	A	24	TYR	7
1	A	140	GLU	7
1	A	143	ASN	7
1	A	78	LEU	6
1	A	25	PHE	6
1	A	8	LEU	6
1	A	132	LYS	5
1	A	112	ARG	5
1	A	6	THR	5
1	A	45	ARG	5
1	A	99	ASP	5
1	A	23	LYS	5
1	A	14	LYS	4
1	A	120	THR	4
1	A	80	GLU	4
1	A	62	THR	4
1	A	116	LEU	4
1	A	106	VAL	4
1	A	123	ARG	4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	158	ASN	4
1	A	157	SER	4
1	A	146	LYS	4
1	A	48	ARG	3
1	A	151	THR	3
1	A	153	VAL	3
1	A	93	THR	3
1	A	43	ASN	3
1	A	141	LYS	3
1	A	63	LYS	3
1	A	28	CYS	3
1	A	35	ASP	3
1	A	115	SER	3
1	A	131	SER	3
1	A	110	SER	3
1	A	92	TYR	3
1	A	16	TYR	3
1	A	98	ARG	3
1	A	38	ASP	3
1	A	94	GLN	3
1	A	105	ASN	2
1	A	17	ASN	2
1	A	5	THR	2
1	A	100	LEU	2
1	A	109	LEU	2
1	A	71	ASN	2
1	A	55	ASP	2
1	A	119	ASN	2
1	A	40	LEU	2
1	A	85	MET	2
1	A	150	TYR	2
1	A	26	LYS	2
1	A	87	ASN	2
1	A	66	LYS	2
1	A	134	ASP	2
1	A	64	THR	2
1	A	122	ARG	2
1	A	113	LEU	2
1	A	101	LEU	2
1	A	117	ARG	2
1	A	88	PHE	2
1	A	142	LEU	2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	53	ARG	2
1	A	96	ASN	1
1	A	139	VAL	1
1	A	77	HIS	1
1	A	121	SER	1
1	A	12	LEU	1
1	A	128	ASP	1
1	A	79	THR	1
1	A	95	ARG	1
1	A	104	ILE	1
1	A	152	LEU	1
1	A	44	PHE	1
1	A	91	SER	1
1	A	54	TYR	1
1	A	18	GLN	1
1	A	156	VAL	1
1	A	124	PHE	1
1	A	19	ASN	1
1	A	11	ASN	1
1	A	73	ILE	1
1	A	20	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 [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 84% for the well-defined parts and 83% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 7070

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	1958
Number of shifts mapped to atoms	1958
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

#### 7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	158	$-0.70 \pm 0.13$	Should be applied
$^{13}\text{C}_\beta$	153	$0.11 \pm 0.16$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	147	$-0.15 \pm 0.26$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	151	$0.84 \pm 0.27$	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 84%, i.e. 1692 atoms were assigned a chemical shift out of a possible 2013. 29 out of 29 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	752/770 (98%)	303/307 (99%)	300/312 (96%)	149/151 (99%)
Sidechain	839/1076 (78%)	501/628 (80%)	327/390 (84%)	11/58 (19%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	101/167 (60%)	66/89 (74%)	34/73 (47%)	1/5 (20%)
Overall	1692/2013 (84%)	870/1024 (85%)	661/775 (85%)	161/214 (75%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 83%, i.e. 1728 atoms were assigned a chemical shift out of a possible 2087. 30 out of 30 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	764/793 (96%)	308/316 (97%)	305/322 (95%)	151/155 (97%)
Sidechain	863/1127 (77%)	516/659 (78%)	336/407 (83%)	11/61 (18%)
Aromatic	101/167 (60%)	66/89 (74%)	34/73 (47%)	1/5 (20%)
Overall	1728/2087 (83%)	890/1064 (84%)	675/802 (84%)	163/221 (74%)

#### 7.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

#### 7.1.5 Random Coil Index (RCI) plots ⓘ

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:

