



# Full wwPDB NMR Structure Validation Report ⓘ

Apr 26, 2016 – 07:35 PM BST

PDB ID : 2DJA  
Title : Solution structure of the B-box domain of the human Midline-2 protein  
Authors : Miyamoto, K.; Kigawa, T.; Tomizawa, T.; Koshiba, S.; Inoue, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)  
Deposited on : 2006-03-31

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

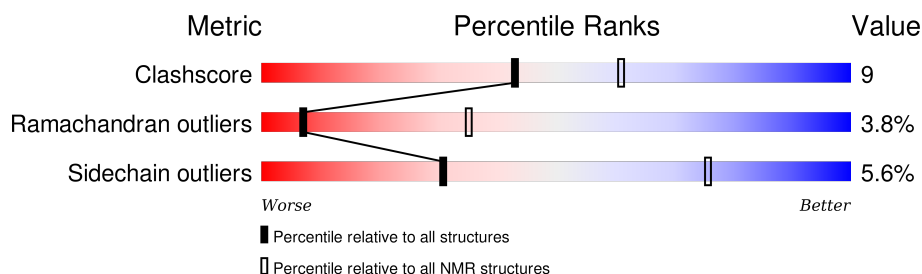
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 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  $\geq 3$ , 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	84	

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 11 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 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:20-A:57 (38)	0.28	11

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 1 single-model cluster was found.

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

### 3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1265 atoms, of which 625 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Midline-2.

Mol	Chain	Residues	Atoms						Trace
1	A	84	Total	C	H	N	O	S	0
			1263	387	625	117	128	6	

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP Q9UJV3
A	2	SER	-	CLONING ARTIFACT	UNP Q9UJV3
A	3	SER	-	CLONING ARTIFACT	UNP Q9UJV3
A	4	GLY	-	CLONING ARTIFACT	UNP Q9UJV3
A	5	SER	-	CLONING ARTIFACT	UNP Q9UJV3
A	6	SER	-	CLONING ARTIFACT	UNP Q9UJV3
A	7	GLY	-	CLONING ARTIFACT	UNP Q9UJV3
A	79	SER	-	CLONING ARTIFACT	UNP Q9UJV3
A	80	GLY	-	CLONING ARTIFACT	UNP Q9UJV3
A	81	PRO	-	CLONING ARTIFACT	UNP Q9UJV3
A	82	SER	-	CLONING ARTIFACT	UNP Q9UJV3
A	83	SER	-	CLONING ARTIFACT	UNP Q9UJV3
A	84	GLY	-	CLONING ARTIFACT	UNP Q9UJV3

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

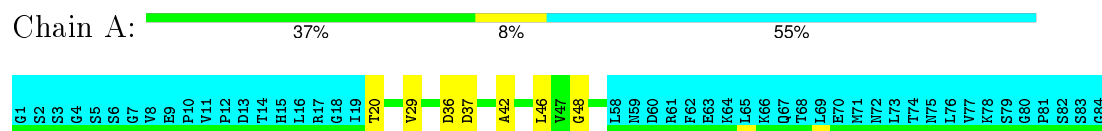
Mol	Chain	Residues	Atoms	
2	A	2	Total	Zn
			2	2

## 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: Midline-2

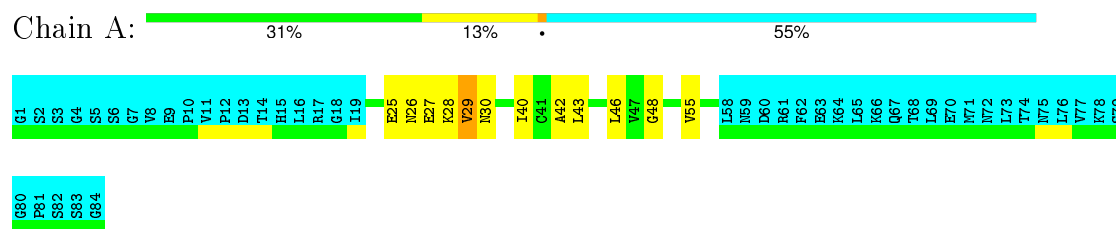


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

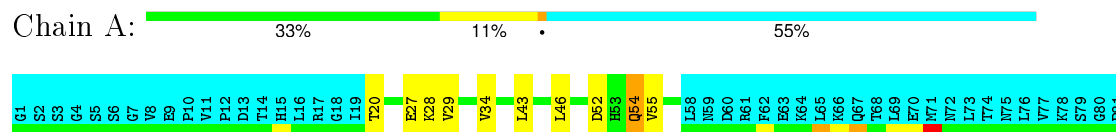
#### 4.2.1 Score per residue for model 1

- Molecule 1: Midline-2



#### 4.2.2 Score per residue for model 2

- Molecule 1: Midline-2

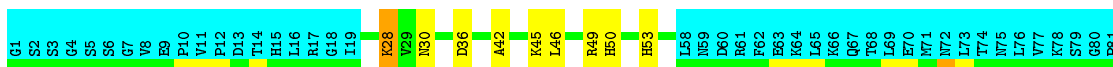


S82  
S83  
G84

### 4.2.3 Score per residue for model 3

- Molecule 1: Midline-2

Chain A: 35% 10% 55%



S82  
S83  
G84

### 4.2.4 Score per residue for model 4

- Molecule 1: Midline-2

Chain A: 27% 17% 55%

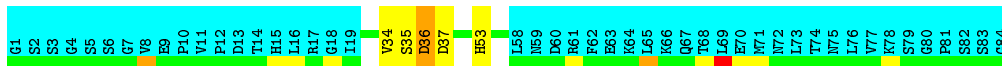


T74  
M75  
L76  
V77  
K78  
S79  
G80  
P81  
S82  
S83  
G84

### 4.2.5 Score per residue for model 5

- Molecule 1: Midline-2

Chain A: 39% 5% 55%



### 4.2.6 Score per residue for model 6

- Molecule 1: Midline-2

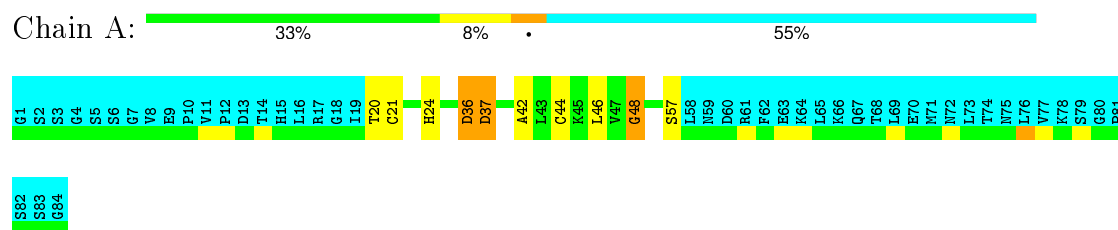
Chain A: 35% 11% 55%



G84

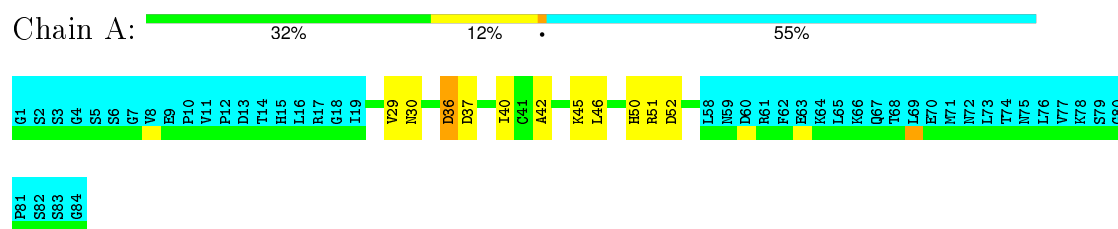
### 4.2.7 Score per residue for model 7

- Molecule 1: Midline-2



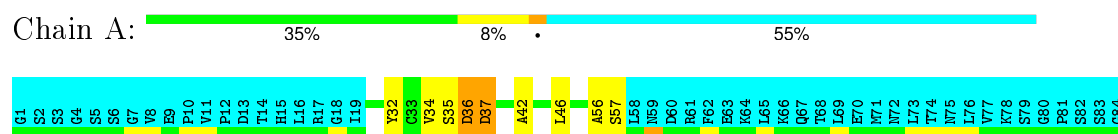
### 4.2.8 Score per residue for model 8

- Molecule 1: Midline-2



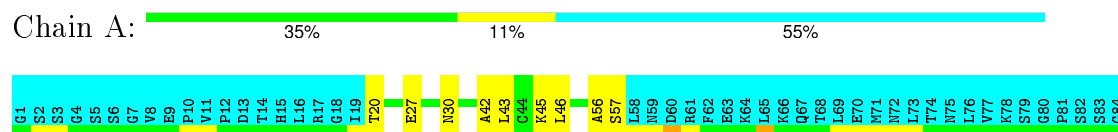
### 4.2.9 Score per residue for model 9

- Molecule 1: Midline-2




### 4.2.10 Score per residue for model 10

- Molecule 1: Midline-2



### 4.2.11 Score per residue for model 11 (medoid)

- Molecule 1: Midline-2

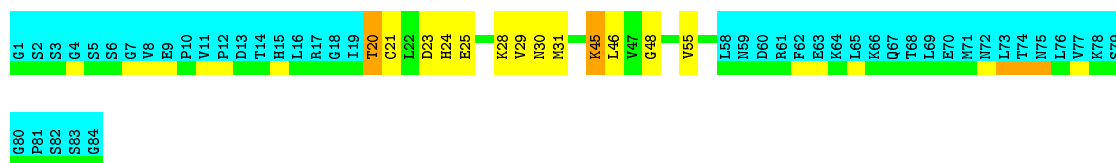
Chain A:  37% 6% • 55%



#### 4.2.12 Score per residue for model 12

- Molecule 1: Midline-2

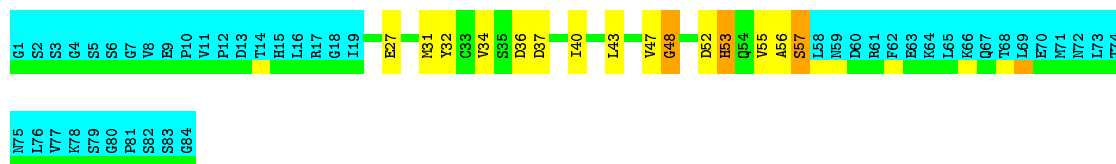
Chain A:  30% 13% • 55%



#### 4.2.13 Score per residue for model 13


- Molecule 1: Midline-2

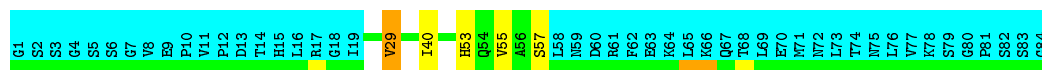
Chain A:  27% 14% • 55%



#### 4.2.14 Score per residue for model 14


- Molecule 1: Midline-2

Chain A:  39% 5% • 55%



#### 4.2.15 Score per residue for model 15

- Molecule 1: Midline-2

Chain A:  32% 12% • 55%





S79  
G80  
P81  
S82  
S83  
G84

#### 4.2.16 Score per residue for model 16

- Molecule 1: Midline-2

Chain A: 31% 13% 55%

G1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40 S41 S42 S43 S44 S45 S46 S47 S48 S49 S50 S51 S52 S53 S54 S55 S56 S57 S58 S59 S60 S61 S62 S63 S64 S65 S66 S67 S68 S69 S70 S71 S72 S73 S74 S75 S76 S77 S78 S79

G80  
P81  
S82  
S83  
G84

#### 4.2.17 Score per residue for model 17

- Molecule 1: Midline-2

Chain A: 32% 12% 55%

G1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40 S41 S42 S43 S44 S45 S46 S47 S48 S49 S50 S51 S52 S53 S54 S55 S56 S57 S58 S59 S60 S61 S62 S63 S64 S65 S66 S67 S68 S69 S70 S71 S72 S73 S74 S75 S76 S77 S78 S79

S79  
G80  
P81  
S82  
S83  
G84

#### 4.2.18 Score per residue for model 18

- Molecule 1: Midline-2

Chain A: 30% 14% 55%

G1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40 S41 S42 S43 S44 S45 S46 S47 S48 S49 S50 S51 S52 S53 S54 S55 S56 S57 S58 S59 S60 S61 S62 S63 S64 S65 S66 S67 S68 S69 S70 S71 S72 S73 S74

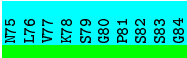
S75  
L76  
L77  
K78  
S79  
G80  
P81  
S82  
S83  
G84

#### 4.2.19 Score per residue for model 19

- Molecule 1: Midline-2

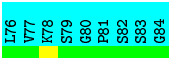
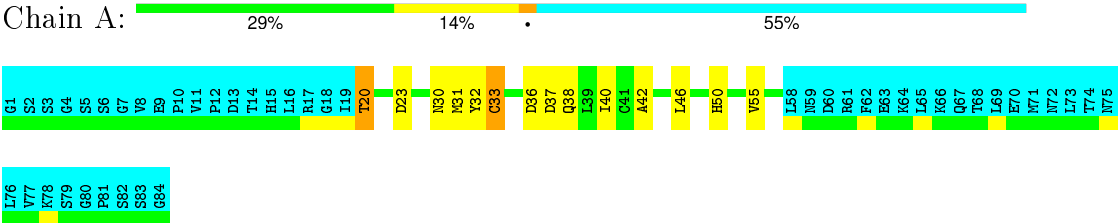
Chain A: 29% 14% 55%

G1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28 S29 S30 S31 S32 S33 S34 S35 S36 S37 S38 S39 S40 S41 S42 S43 S44 S45 S46 S47 S48 S49 S50 S51 S52 S53 S54 S55 S56 S57 S58 S59 S60 S61 S62 S63 S64 S65 S66 S67 S68 S69 S70 S71 S72 S73 S74



4.2.20 Score per residue for model 20

- Molecule 1: Midline-2



## 5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function, 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	2.0.17
CYANA	refinement	2.0.17

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 11331
Number of chemical shift lists	1
Total number of shifts	942
Number of shifts mapped to atoms	942
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

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

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	299	284	284	5±2
All	All	6020	5680	5680	101

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 9.

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:ALA:O	1:A:46:LEU:HD13	0.82	1.74	15	11
1:A:40:ILE:HD11	1:A:55:VAL:HG11	0.79	1.54	18	7
1:A:27:GLU:HG2	1:A:43:LEU:HD12	0.76	1.57	17	1
1:A:34:VAL:HG21	1:A:54:GLN:OE1	0.74	1.82	2	2
1:A:27:GLU:CG	1:A:43:LEU:HD12	0.66	2.19	17	1
1:A:42:ALA:C	1:A:46:LEU:HD23	0.59	2.18	7	3
1:A:45:LYS:HG3	1:A:55:VAL:HG21	0.58	1.74	12	1
1:A:49:ARG:CG	1:A:50:HIS:CD2	0.58	2.86	3	1
1:A:27:GLU:HG3	1:A:43:LEU:HD12	0.58	1.75	2	6
1:A:31:MET:SD	1:A:55:VAL:HG21	0.58	2.39	13	2
1:A:42:ALA:O	1:A:46:LEU:HD23	0.56	2.00	7	1
1:A:49:ARG:HG3	1:A:50:HIS:CD2	0.53	2.39	3	1

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:31:MET:HE1	1:A:42:ALA:HA	0.50	1.82	20	1
1:A:30:ASN:ND2	1:A:31:MET:SD	0.50	2.85	12	3
1:A:27:GLU:N	1:A:27:GLU:OE1	0.50	2.45	17	1
1:A:29:VAL:HG23	1:A:29:VAL:O	0.49	2.07	6	2
1:A:45:LYS:O	1:A:46:LEU:C	0.49	2.51	16	1
1:A:40:ILE:HG21	1:A:50:HIS:CE1	0.49	2.43	8	2
1:A:27:GLU:CD	1:A:27:GLU:N	0.49	2.66	17	1
1:A:34:VAL:HG23	1:A:54:GLN:O	0.48	2.08	19	2
1:A:36:ASP:OD1	1:A:53:HIS:CE1	0.48	2.65	3	1
1:A:20:THR:HG21	1:A:25:GLU:O	0.48	2.08	17	1
1:A:33:CYS:N	1:A:38:GLN:O	0.47	2.47	20	1
1:A:29:VAL:O	1:A:29:VAL:HG23	0.47	2.09	19	1
1:A:29:VAL:HG11	1:A:39:LEU:HB3	0.46	1.87	16	1
1:A:44:CYS:O	1:A:48:GLY:CA	0.46	2.64	7	2
1:A:32:TYR:CE1	1:A:37:ASP:HA	0.46	2.46	20	2
1:A:52:ASP:O	1:A:53:HIS:C	0.46	2.54	13	1
1:A:40:ILE:HG21	1:A:50:HIS:CD2	0.46	2.46	20	1
1:A:38:GLN:NE2	1:A:39:LEU:O	0.45	2.49	11	1
1:A:36:ASP:O	1:A:37:ASP:C	0.45	2.55	5	7
1:A:28:LYS:O	1:A:29:VAL:C	0.45	2.55	11	6
1:A:53:HIS:CD2	1:A:53:HIS:H	0.44	2.30	19	1
1:A:43:LEU:O	1:A:44:CYS:C	0.44	2.55	16	1
1:A:32:TYR:CE2	1:A:34:VAL:HG22	0.44	2.48	15	2
1:A:40:ILE:CD1	1:A:55:VAL:HG11	0.43	2.42	19	1
1:A:34:VAL:O	1:A:35:SER:C	0.43	2.55	9	2
1:A:56:ALA:O	1:A:57:SER:C	0.43	2.55	9	2
1:A:36:ASP:C	1:A:37:ASP:OD1	0.43	2.57	15	2
1:A:52:ASP:O	1:A:53:HIS:CD2	0.42	2.73	15	1
1:A:30:ASN:OD1	1:A:30:ASN:N	0.42	2.52	8	1
1:A:51:ARG:HB3	1:A:51:ARG:CZ	0.41	2.45	19	1
1:A:40:ILE:HD11	1:A:55:VAL:CG1	0.41	2.44	14	1
1:A:51:ARG:O	1:A:52:ASP:CG	0.41	2.59	8	1
1:A:47:VAL:O	1:A:48:GLY:C	0.41	2.58	13	1
1:A:21:CYS:O	1:A:24:HIS:O	0.41	2.39	18	3
1:A:52:ASP:CG	1:A:52:ASP:O	0.41	2.58	2	1
1:A:49:ARG:HA	1:A:49:ARG:CZ	0.41	2.46	19	1
1:A:54:GLN:HE21	1:A:55:VAL:N	0.41	2.14	2	1
1:A:28:LYS:N	1:A:28:LYS:HD2	0.40	2.31	3	1
1:A:42:ALA:C	1:A:46:LEU:HD13	0.40	2.37	17	1
1:A:21:CYS:O	1:A:24:HIS:C	0.40	2.60	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	38/84 (45%)	28±2 (74±6%)	9±2 (23±6%)	1±1 (4±2%)	7	35
All	All	760/1680 (45%)	559 (74%)	172 (23%)	29 (4%)	7	35

All 6 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	48	GLY	10
1	A	29	VAL	6
1	A	20	THR	5
1	A	53	HIS	4
1	A	57	SER	3
1	A	33	CYS	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	35/75 (47%)	33±1 (94±3%)	2±1 (6±3%)	31	76
All	All	700/1500 (47%)	661 (94%)	39 (6%)	31	76

All 16 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	36	ASP	6
1	A	45	LYS	5
1	A	23	ASP	4
1	A	25	GLU	3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	37	ASP	3
1	A	30	ASN	3
1	A	20	THR	2
1	A	29	VAL	2
1	A	46	LEU	2
1	A	28	LYS	2
1	A	49	ARG	2
1	A	27	GLU	1
1	A	54	GLN	1
1	A	26	ASN	1
1	A	38	GLN	1
1	A	52	ASP	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](#)

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 85% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 11331

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	942
Number of shifts mapped to atoms	942
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
$^{13}\text{C}_\alpha$	75	$0.20 \pm 0.14$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	71	$-0.13 \pm 0.09$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}'$	73	$0.19 \pm 0.12$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	71	$-0.08 \pm 0.66$	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 91%, i.e. 417 atoms were assigned a chemical shift out of a possible 460. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	188/190 (99%)	76/76 (100%)	74/76 (97%)	38/38 (100%)
Sidechain	209/241 (87%)	129/140 (92%)	76/89 (85%)	4/12 (33%)

*Continued on next page...*



Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	20/29 (69%)	10/16 (62%)	10/10 (100%)	0/3 (0%)
Overall	417/460 (91%)	215/232 (93%)	160/175 (91%)	42/53 (79%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 85%, i.e. 837 atoms were assigned a chemical shift out of a possible 990. 1 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	365/414 (88%)	146/165 (88%)	148/168 (88%)	71/81 (88%)
Sidechain	439/531 (83%)	271/311 (87%)	160/195 (82%)	8/25 (32%)
Aromatic	33/45 (73%)	17/25 (68%)	16/16 (100%)	0/4 (0%)
Overall	837/990 (85%)	434/501 (87%)	324/379 (85%)	79/110 (72%)

#### 7.1.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

#### 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:

