



# Full wwPDB NMR Structure Validation Report ⓘ

Feb 12, 2017 – 10:37 pm GMT

PDB ID : 2K27  
Title : Solution structure of Human Pax8 Paired Box Domain  
Authors : Codutti, L.; Esposito, G.; Corazza, A.; Fogolari, F.; Tell, G.; Vascotto, C.; van Ingen, H.; Boelens, R.; Viglino, P.; Quadrifoglio, F.  
Deposited on : 2008-03-26

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

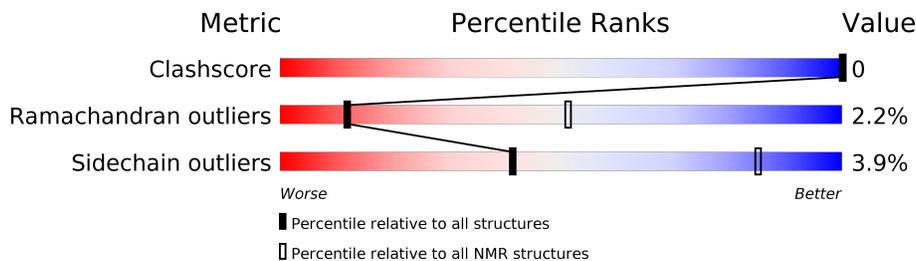
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 76%.

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	159	 62% 38%

## 2 Ensemble composition and analysis i

This entry contains 20 models. Model 15 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:3-A:16 (14)	1.10	19
2	A:27-A:63 (37)	0.78	6
3	A:87-A:134 (48)	0.75	15

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 5 single-model clusters were found.

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

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2453 atoms, of which 1228 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Paired box protein Pax-8.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	159	2453	758	1228	243	218	6	0

There are 16 discrepancies between the modelled and reference sequences:

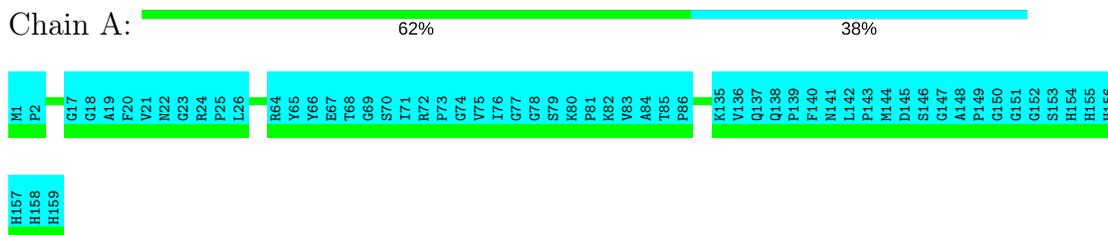
Chain	Residue	Modelled	Actual	Comment	Reference
A	144	MET	-	EXPRESSION TAG	UNP Q06710
A	145	ASP	-	EXPRESSION TAG	UNP Q06710
A	146	SER	-	EXPRESSION TAG	UNP Q06710
A	147	GLY	-	EXPRESSION TAG	UNP Q06710
A	148	ALA	-	EXPRESSION TAG	UNP Q06710
A	149	PRO	-	EXPRESSION TAG	UNP Q06710
A	150	GLY	-	EXPRESSION TAG	UNP Q06710
A	151	GLY	-	EXPRESSION TAG	UNP Q06710
A	152	GLY	-	EXPRESSION TAG	UNP Q06710
A	153	SER	-	EXPRESSION TAG	UNP Q06710
A	154	HIS	-	EXPRESSION TAG	UNP Q06710
A	155	HIS	-	EXPRESSION TAG	UNP Q06710
A	156	HIS	-	EXPRESSION TAG	UNP Q06710
A	157	HIS	-	EXPRESSION TAG	UNP Q06710
A	158	HIS	-	EXPRESSION TAG	UNP Q06710
A	159	HIS	-	EXPRESSION TAG	UNP Q06710

## 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: Paired box protein Pax-8

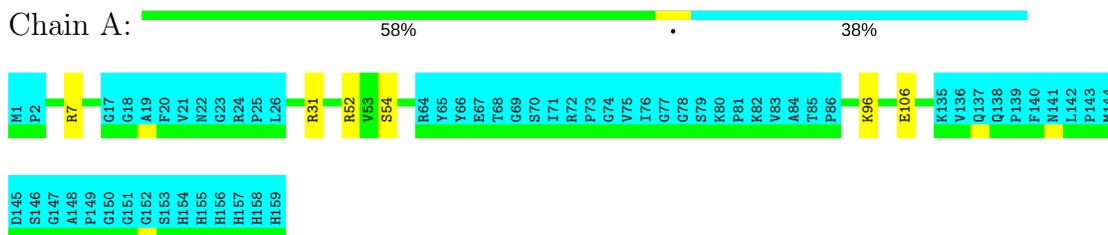


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

#### 4.2.1 Score per residue for model 1

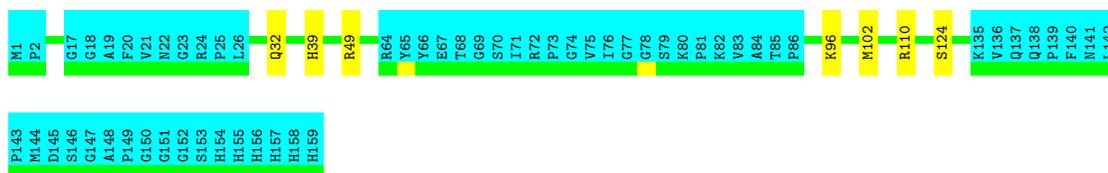
- Molecule 1: Paired box protein Pax-8



#### 4.2.2 Score per residue for model 2

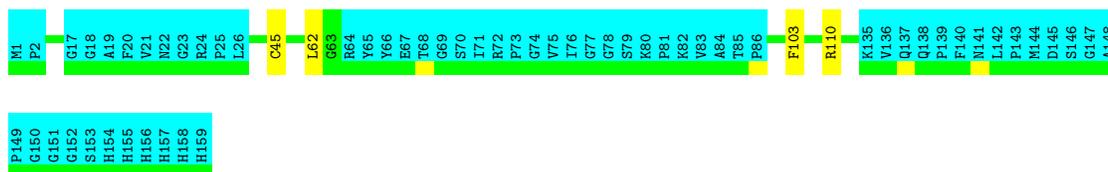
- Molecule 1: Paired box protein Pax-8





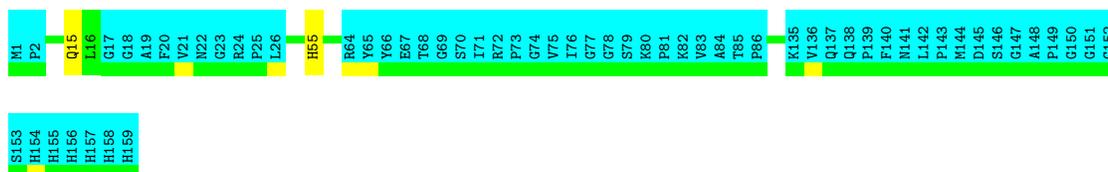
### 4.2.3 Score per residue for model 3

- Molecule 1: Paired box protein Pax-8



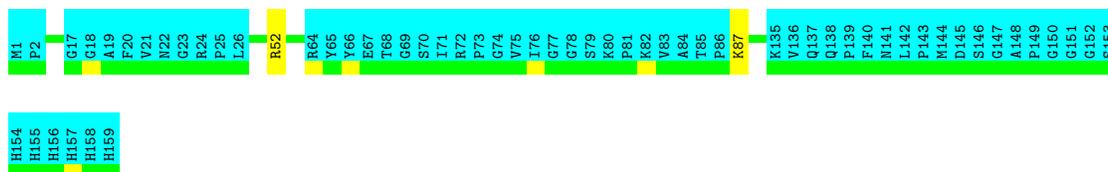
### 4.2.4 Score per residue for model 4

- Molecule 1: Paired box protein Pax-8



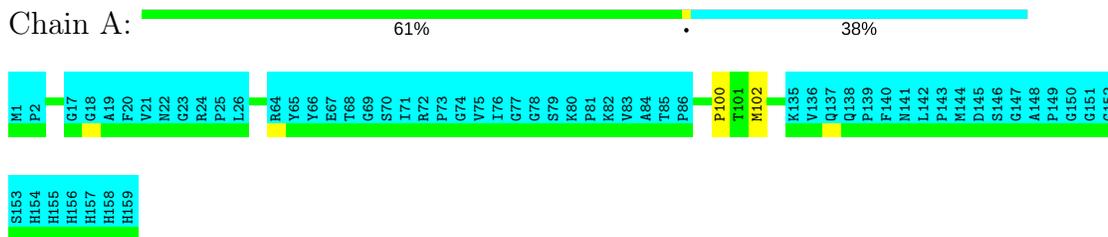
### 4.2.5 Score per residue for model 5

- Molecule 1: Paired box protein Pax-8



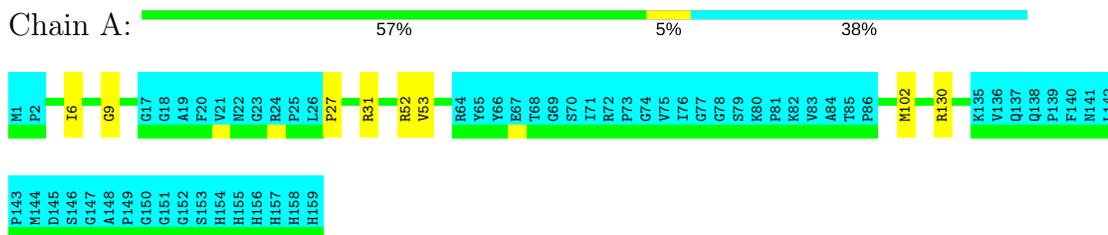
### 4.2.6 Score per residue for model 6

- Molecule 1: Paired box protein Pax-8



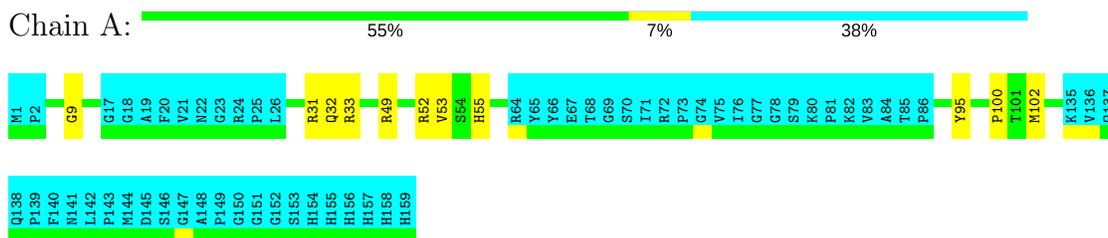
#### 4.2.7 Score per residue for model 7

- Molecule 1: Paired box protein Pax-8



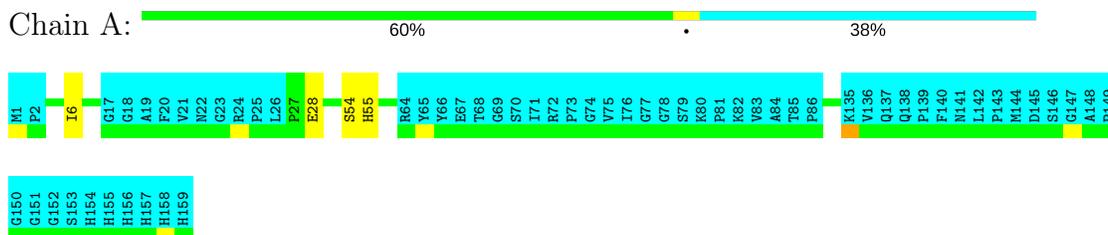
#### 4.2.8 Score per residue for model 8

- Molecule 1: Paired box protein Pax-8



#### 4.2.9 Score per residue for model 9

- Molecule 1: Paired box protein Pax-8





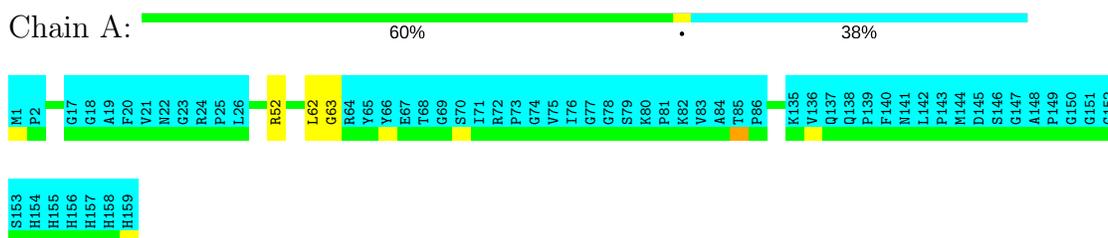
#### 4.2.14 Score per residue for model 14

- Molecule 1: Paired box protein Pax-8



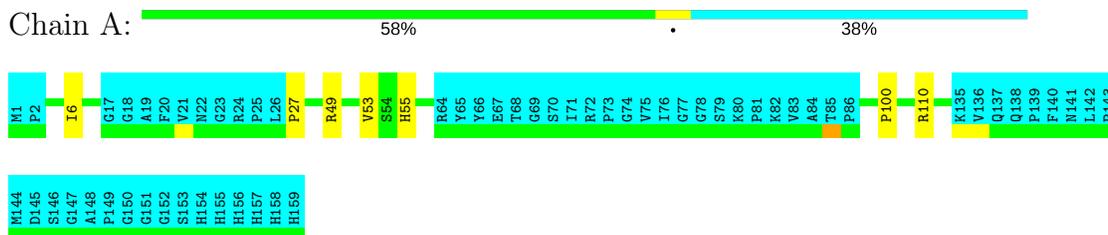
#### 4.2.15 Score per residue for model 15 (medoid)

- Molecule 1: Paired box protein Pax-8



#### 4.2.16 Score per residue for model 16

- Molecule 1: Paired box protein Pax-8



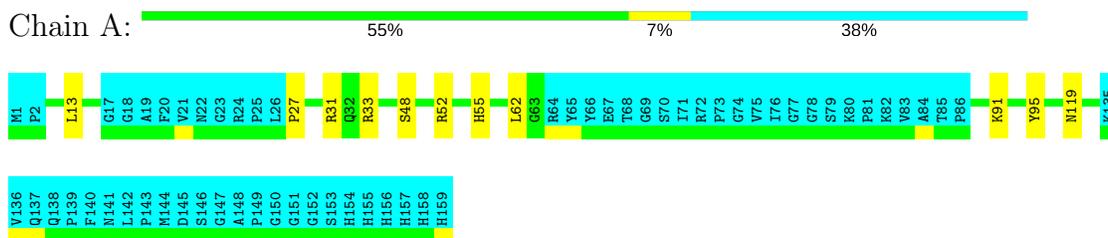
#### 4.2.17 Score per residue for model 17

- Molecule 1: Paired box protein Pax-8



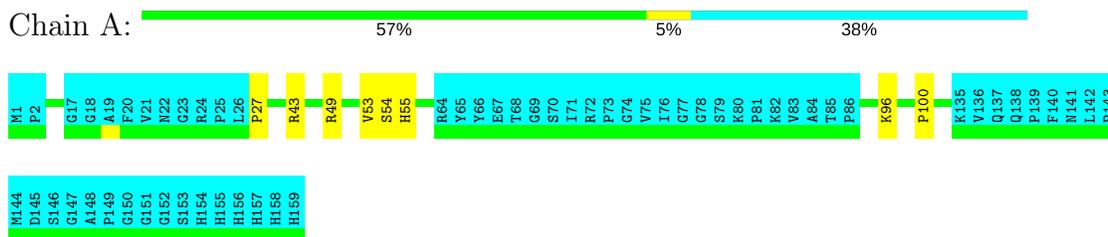
### 4.2.18 Score per residue for model 18

- Molecule 1: Paired box protein Pax-8



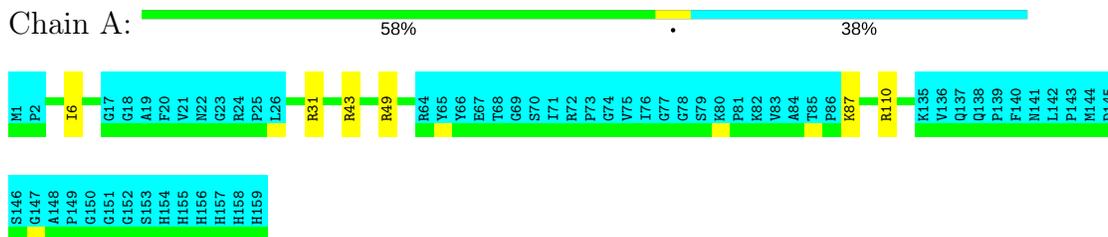
### 4.2.19 Score per residue for model 19

- Molecule 1: Paired box protein Pax-8



### 4.2.20 Score per residue for model 20

- Molecule 1: Paired box protein Pax-8



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *molecular mechanics*.

Of the 300 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
CYANA	structure solution	2.1
DISCOVER	refinement	2.98

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 15693
Number of chemical shift lists	1
Total number of shifts	1612
Number of shifts mapped to atoms	1612
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	76%

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
1	A	0.51±0.00	0±0/790 (0.0±0.0%)	0.79±0.02	0±0/1066 (0.0±0.0%)
All	All	0.51	0/15800 (0.0%)	0.79	2/21320 (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	0.3±0.7
All	All	0	6

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	95	TYR	CB-CG-CD2	-5.29	117.83	121.00	12	2

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	95	TYR	Sidechain	3
1	A	43	ARG	Sidechain	1
1	A	33	ARG	Sidechain	1
1	A	119	ASN	Peptide	1

## 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	15560	15920	15920	-

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	99/159 (62%)	86±4 (87±4%)	11±4 (11±4%)	2±2 (2±2%)	12	51
All	All	1980/3180 (62%)	1716 (87%)	221 (11%)	43 (2%)	12	51

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	55	HIS	9
1	A	27	PRO	6
1	A	53	VAL	5
1	A	6	ILE	4
1	A	100	PRO	4
1	A	54	SER	4
1	A	102	MET	2
1	A	9	GLY	2
1	A	63	GLY	2
1	A	12	GLY	1
1	A	124	SER	1
1	A	52	ARG	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	45	CYS	1
1	A	28	GLU	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	88/134 (66%)	85±2 (96±2%)	3±2 (4±2%)	41	85
All	All	1760/2680 (66%)	1691 (96%)	69 (4%)	41	85

All 26 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	31	ARG	7
1	A	52	ARG	6
1	A	49	ARG	6
1	A	110	ARG	6
1	A	102	MET	5
1	A	96	LYS	4
1	A	7	ARG	3
1	A	33	ARG	3
1	A	108	ARG	3
1	A	62	LEU	3
1	A	130	ARG	3
1	A	32	GLN	3
1	A	119	ASN	2
1	A	6	ILE	2
1	A	87	LYS	2
1	A	106	GLU	1
1	A	43	ARG	1
1	A	91	LYS	1
1	A	39	HIS	1
1	A	13	LEU	1
1	A	48	SER	1
1	A	15	GLN	1
1	A	105	TRP	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	114	GLU	1
1	A	109	ASP	1
1	A	103	PHE	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 [i](#)

The completeness of assignment taking into account all chemical shift lists is 76% for the well-defined parts and 73% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 15693

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

#### 7.1.1 Bookkeeping [i](#)

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	1612
Number of shifts mapped to atoms	1612
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 [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	149	0.02 $\pm$ 0.07	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	129	0.50 $\pm$ 0.09	Should be applied
$^{13}\text{C}'$	92	0.02 $\pm$ 0.10	None needed (< 0.5 ppm)
$^{15}\text{N}$	139	-0.08 $\pm$ 0.19	None needed (< 0.5 ppm)

#### 7.1.3 Completeness of resonance assignments [i](#)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 76%, i.e. 960 atoms were assigned a chemical shift out of a possible 1257. 16 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	434/487 (89%)	187/194 (96%)	152/198 (77%)	95/95 (100%)
Sidechain	504/713 (71%)	294/418 (70%)	202/248 (81%)	8/47 (17%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	22/57 (39%)	12/31 (39%)	9/21 (43%)	1/5 (20%)
Overall	960/1257 (76%)	493/643 (77%)	363/467 (78%)	104/147 (71%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 73%, i.e. 1429 atoms were assigned a chemical shift out of a possible 1959. 18 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	655/771 (85%)	275/306 (90%)	241/318 (76%)	139/147 (95%)
Sidechain	730/1055 (69%)	430/626 (69%)	289/366 (79%)	11/63 (17%)
Aromatic	44/133 (33%)	24/73 (33%)	19/49 (39%)	1/11 (9%)
Overall	1429/1959 (73%)	729/1005 (73%)	549/733 (75%)	151/221 (68%)

#### 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:

