



# wwPDB NMR Structure Validation Summary Report ⓘ

Feb 12, 2017 – 06:12 pm GMT

PDB ID : 1H5P  
Title : SOLUTION STRUCTURE OF THE HUMAN SP100B SAND DOMAIN BY HETERONUCLEAR NMR.  
Authors : Bottomley, M.J.; Liu, Z.; Collard, M.W.; Huggenvik, J.I.; Gibson, T.J.; Sattler, M.  
Deposited on : 2001-05-24

This is a wwPDB NMR Structure Validation Summary 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

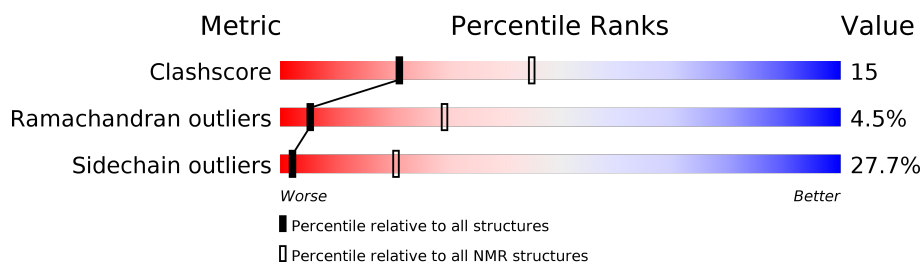
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 68%.

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	95	

## 2 Ensemble composition and analysis

This entry contains 10 models. Model 10 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:598-A:632, A:636-A:673 (73)	0.29	10

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 2 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 4, 5, 6, 8, 9, 10
2	3, 7
Single-model clusters	2

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1569 atoms, of which 795 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called NUCLEAR AUTOANTIGEN SP100-B.

Mol	Chain	Residues	Atoms						Trace
1	A	95	Total	C	H	N	O	S	0
			1569	491	795	134	144	5	

There are 4 discrepancies between the modelled and reference sequences:

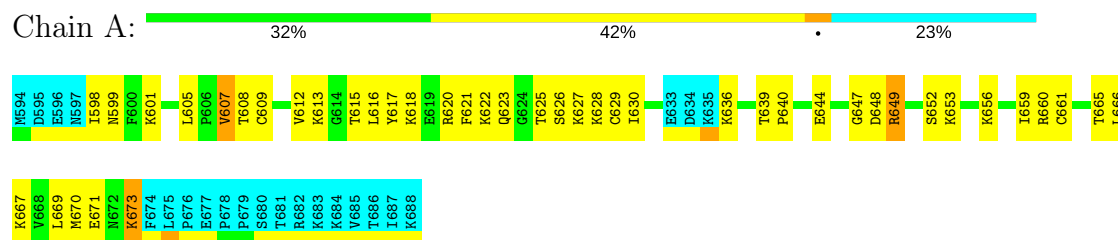
Chain	Residue	Modelled	Actual	Comment	Reference
A	685	VAL	ARG	VARIANT	UNP P23497
A	686	THR	ILE	VARIANT	UNP P23497
A	687	ILE	LEU	VARIANT	UNP P23497
A	688	LYS	GLU	VARIANT	UNP P23497

## 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.

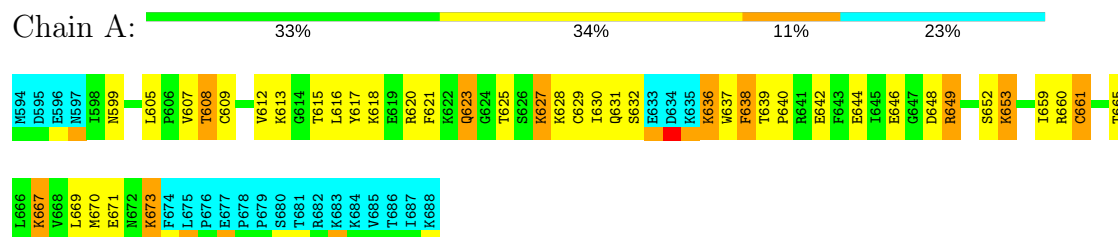
- Molecule 1: NUCLEAR AUTOANTIGEN SP100-B



### 4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 10. Colouring as in section 4.1 above.

- Molecule 1: NUCLEAR AUTOANTIGEN SP100-B



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *ARIA AMBIGUOUS DISTANCE RESTRAINTS SIMULATED ANNEALING WITH TORSION ANGLE DYNAMICS*.

Of the 100 calculated structures, 10 were deposited, based on the following criterion: *ACCEPTABLE COVALENT GEOMETRY, FAVOURABLE NON-BOND ENERGY, FEWEST RESTRAINT VIOLATIONS, LOWEST OVERALL ENERGIES*.

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

Software name	Classification	Version
CNS 0.3	refinement	
NMRPIPE	structure solution	
XEASY	structure solution	
ARIA 1.0	structure solution	
CNS 0.9	structure solution	

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 5558
Number of chemical shift lists	1
Total number of shifts	878
Number of shifts mapped to atoms	870
Number of unparsed shifts	0
Number of shifts with mapping errors	8
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	68%

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](#)

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 [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
1	A	592	607	607	18±2
All	All	5920	6070	6070	179

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 15.

5 of 80 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:620:ARG:HB2	1:A:629:CYS:HB3	0.73	1.61	7	5
1:A:620:ARG:HB3	1:A:627:LYS:HB3	0.66	1.66	7	2
1:A:623:GLN:HB2	1:A:626:SER:HB2	0.61	1.71	3	1
1:A:623:GLN:HG3	1:A:627:LYS:HG3	0.61	1.71	10	2
1:A:600:PHE:HA	1:A:605:LEU:HD13	0.61	1.72	2	3

### 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	73/95 (77%)	60±2 (82±2%)	10±1 (13±2%)	3±1 (5±2%)	5	29
All	All	730/950 (77%)	600 (82%)	97 (13%)	33 (5%)	5	29

5 of 8 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	673	LYS	10
1	A	649	ARG	9
1	A	653	LYS	7
1	A	652	SER	2
1	A	663	GLY	2

### 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	65/87 (75%)	47±4 (72±6%)	18±4 (28±6%)	2	21
All	All	650/870 (75%)	470 (72%)	180 (28%)	2	21

5 of 42 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	667	LYS	10
1	A	607	VAL	10
1	A	615	THR	10
1	A	636	LYS	9
1	A	652	SER	9

### 6.3.3 RNA ⓘ

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 68% for the well-defined parts and 64% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 5558

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	878
Number of shifts mapped to atoms	870
Number of unparsed shifts	0
Number of shifts with mapping errors	8
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Residue not found in structure. First 5 (of 8) occurrences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	9	PRO	CD	48.284	0.02	1
A	9	PRO	HA	4.33	0.02	1
A	9	PRO	CA	60.709	0.02	1
A	9	PRO	CG	24.986	0.02	1
A	9	PRO	HD3	3.501	0.02	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
$^{13}\text{C}_\alpha$	91	$2.31 \pm 0.13$	Should be applied
$^{13}\text{C}_\beta$	81	$2.57 \pm 0.19$	Should be applied
$^{13}\text{C}'$	0	—	None (insufficient data)

*Continued on next page...*

Continued from previous page...

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{15}\text{N}$	88	$0.73 \pm 0.43$	None needed (imprecise)

### 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 68%, i.e. 644 atoms were assigned a chemical shift out of a possible 951. 6 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	288/361 (80%)	144/144 (100%)	73/146 (50%)	71/71 (100%)
Sidechain	309/514 (60%)	157/306 (51%)	146/179 (82%)	6/29 (21%)
Aromatic	47/76 (62%)	27/40 (68%)	18/34 (53%)	2/2 (100%)
Overall	644/951 (68%)	328/490 (67%)	237/359 (66%)	79/102 (77%)

### 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	631	GLN	HB3	-0.52	3.37 – 0.67	-9.4
1	A	631	GLN	HG3	0.47	3.75 – 0.85	-6.3
1	A	656	LYS	HG3	-0.23	2.76 – -0.04	-5.7
1	A	641	ARG	HD3	1.83	4.36 – 1.86	-5.1
1	A	637	TRP	CD1	116.64	136.18 – 116.78	-5.1

### 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:

