



# wwPDB NMR Structure Validation Summary Report ⓘ

May 28, 2020 – 11:14 pm BST

PDB ID : 2LP1  
Title : The solution NMR structure of the transmembrane C-terminal domain of the amyloid precursor protein (C99)  
Authors : Barrett, P.J.; Song, Y.; Van Horn, W.D.; Hustedt, E.J.; Schafer, J.M.; Hadzise-  
limovic, A.; Beel, A.J.; Sanders, C.R.  
Deposited on : 2012-01-30

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

<https://www.wwpdb.org/validation/2017/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 : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : 2.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

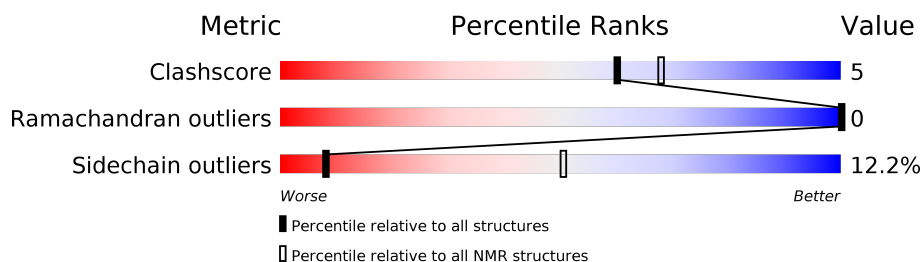
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 43%.

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	122	

## 2 Ensemble composition and analysis ⓘ

This entry contains 30 models. Model 24 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:701-A:723 (23)	0.32	24

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 3 clusters and 1 single-model cluster was found.

Cluster number	Models
1	2, 3, 4, 5, 9, 11, 12, 14, 19, 22, 24, 25, 26, 28, 29, 30
2	6, 8, 10, 13, 16, 20, 23
3	1, 7, 15, 17, 21, 27
Single-model clusters	18

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 732 atoms, of which 384 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called C99.

Mol	Chain	Residues	Atoms						Trace
1	A	46	Total	C	H	N	O	S	0
			732	231	384	58	57	2	

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	771	GLN	-	EXPRESSION TAG	UNP P05067
A	772	GLY	-	EXPRESSION TAG	UNP P05067
A	773	ARG	-	EXPRESSION TAG	UNP P05067
A	774	ILE	-	EXPRESSION TAG	UNP P05067
A	775	LEU	-	EXPRESSION TAG	UNP P05067
A	776	GLN	-	EXPRESSION TAG	UNP P05067
A	777	ILE	-	EXPRESSION TAG	UNP P05067
A	778	SER	-	EXPRESSION TAG	UNP P05067
A	779	ILE	-	EXPRESSION TAG	UNP P05067
A	780	THR	-	EXPRESSION TAG	UNP P05067
A	781	LEU	-	EXPRESSION TAG	UNP P05067
A	782	ALA	-	EXPRESSION TAG	UNP P05067
A	783	ALA	-	EXPRESSION TAG	UNP P05067
A	784	ALA	-	EXPRESSION TAG	UNP P05067
A	785	LEU	-	EXPRESSION TAG	UNP P05067
A	786	GLU	-	EXPRESSION TAG	UNP P05067
A	787	HIS	-	EXPRESSION TAG	UNP P05067
A	788	HIS	-	EXPRESSION TAG	UNP P05067
A	789	HIS	-	EXPRESSION TAG	UNP P05067
A	790	HIS	-	EXPRESSION TAG	UNP P05067
A	791	HIS	-	EXPRESSION TAG	UNP P05067
A	792	HIS	-	EXPRESSION TAG	UNP P05067



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *molecular dynamics, simulated annealing*.

Of the 100 calculated structures, 30 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
X-PLOR NIH	structure solution	2.24
X-PLOR NIH	refinement	2.24
TALOS	geometry optimization	Linux9
TALOS	structure solution	Linux9
ProcheckNMR	structure solution	3.5.4

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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	179
Number of shifts mapped to atoms	179
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	43%

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	158	193	193	2±1
All	All	4740	5790	5790	49

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 5.

5 of 15 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:720:LEU:HD13	1:A:723:LEU:HD12	0.79	1.52	12	1
1:A:707:VAL:O	1:A:710:VAL:HG23	0.61	1.96	21	2
1:A:711:VAL:O	1:A:715:VAL:HG23	0.55	2.01	2	12
1:A:717:VAL:O	1:A:721:VAL:HG23	0.55	2.01	22	2
1:A:720:LEU:CD1	1:A:723:LEU:HD12	0.54	2.29	12	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	23/122 (19%)	23±0 (99±2%)	0±0 (1±2%)	0±0 (0±0%)	100	100
All	All	690/3660 (19%)	685 (99%)	5 (1%)	0 (0%)	100	100

There are no Ramachandran outliers.

### 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	18/104 (17%)	16±1 (88±6%)	2±1 (12±6%)	8	50
All	All	540/3120 (17%)	474 (88%)	66 (12%)	8	50

5 of 10 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	720	LEU	14
1	A	714	THR	12
1	A	719	THR	9
1	A	711	VAL	8
1	A	706	MET	7

### 6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.



## 6.6 Ligand geometry

There are no ligands in this entry.

## 6.7 Other polymers

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 43% for the well-defined parts and 32% for the entire structure.

### 7.1 Chemical shift list 1

File name: input\_cs.cif

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	179
Number of shifts mapped to atoms	179
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
$^{13}\text{C}_\alpha$	37	$-0.71 \pm 0.18$	Should be applied
$^{13}\text{C}_\beta$	30	$1.27 \pm 0.07$	Should be applied
$^{13}\text{C}'$	38	$-0.36 \pm 0.13$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	37	$0.57 \pm 0.28$	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 43%, i.e. 109 atoms were assigned a chemical shift out of a possible 251. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	91/115 (79%)	23/46 (50%)	45/46 (98%)	23/23 (100%)
Sidechain	18/136 (13%)	0/74 (0%)	18/62 (29%)	0/0 (—%)

*Continued on next page...*

*Continued from previous page...*

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	0/0 (—%)	0/0 (—%)	0/0 (—%)	0/0 (—%)
Overall	109/251 (43%)	23/120 (19%)	63/108 (58%)	23/23 (100%)

#### 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:

