



# wwPDB NMR Structure Validation Summary Report ⓘ

Feb 12, 2017 – 07:28 pm GMT

PDB ID : 1PXE  
Title : Solution Structure of a CCHHC Domain of Neural Zinc Finger Factor-1  
Authors : Berkovits-Cymet, H.J.; Amann, B.T.; Berg, J.M.  
Deposited on : 2003-07-03

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

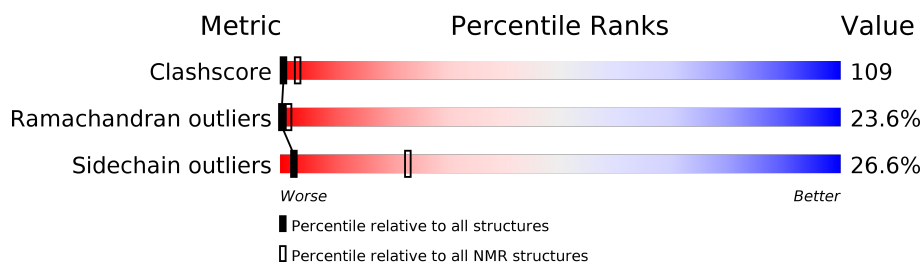
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 39%.

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	63	

## 2 Ensemble composition and analysis

This entry contains 21 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:19-A:55 (37)	0.41	1

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 5 clusters and 3 single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 10, 13, 16, 18, 20
2	7, 9, 14
3	8, 11
4	12, 15
5	5, 6
Single-model clusters	17; 19; 21

### 3 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 723 atoms, of which 355 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called neural zinc finger transcription factor 1.

Mol	Chain	Residues	Atoms						Trace
1	A	48	Total	C	H	N	O	S	0
			722	225	355	71	67	4	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING MET	UNP P70475
A	32	THR	ASN	SEE REMARK 999	UNP P70475

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

Mol	Chain	Residues	Atoms	
2	A	1	Total	Zn
			1	1

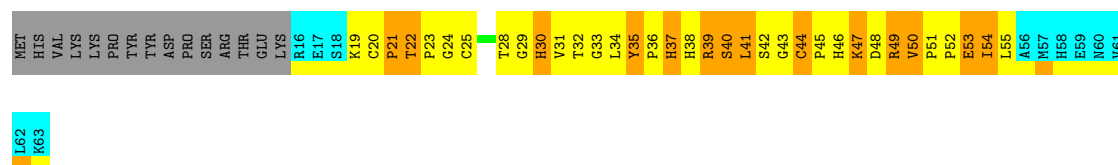
## 4 Residue-property plots [i](#)

### 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: neural zinc finger transcription factor 1

Chain A: 

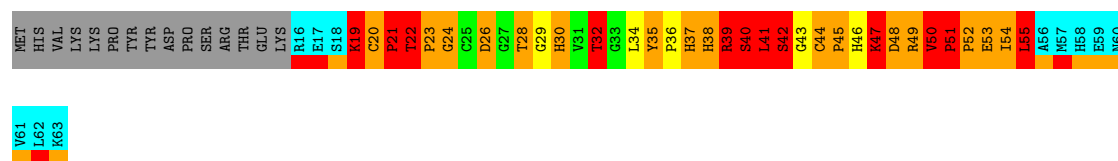


### 4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: neural zinc finger transcription factor 1

Chain A: 



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *Torsion angle dynamics Simulated annealing*.

Of the 30 calculated structures, 21 were deposited, based on the following criterion: *The submitted conformer models are those with the lowest energies..*

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

Software name	Classification	Version
CNS	structure solution	1.1
CNS	refinement	1.1

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 5901
Number of chemical shift lists	1
Total number of shifts	256
Number of shifts mapped to atoms	256
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	39%

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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	1.10±3.33	5±22/287 (1.7±7.5%)	0.93±1.90	3±15/393 (0.9±3.8%)
All	All	3.51	101/6027 (1.7%)	2.12	71/8253 (0.9%)

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	0.1±0.4	0.0±0.0
All	All	2	0

5 of 101 unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	53	GLU	CD-OE1	-83.51	0.33	1.25	1	1
1	A	49	ARG	CZ-NH2	-66.12	0.47	1.33	1	1
1	A	40	SER	CB-OG	-62.75	0.60	1.42	1	1
1	A	53	GLU	CD-OE2	-59.09	0.60	1.25	1	1
1	A	49	ARG	CZ-NH1	-58.39	0.57	1.33	1	1

5 of 71 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	38	HIS	O-C-N	-71.11	8.93	122.70	1	1
1	A	26	ASP	CB-CG-OD1	48.55	162.00	118.30	1	1
1	A	26	ASP	CB-CG-OD2	47.27	160.85	118.30	1	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	48	ASP	CB-CG-OD1	47.07	160.66	118.30	1	1
1	A	26	ASP	OD1-CG-OD2	-45.34	37.16	123.30	1	1

All unique chiral outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
1	A	19	LYS	CA	1
1	A	22	THR	CB	1

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	276	265	265	59±23
All	All	5817	5565	5554	1244

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 109.

5 of 418 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:55:LEU:CB	1:A:55:LEU:CD2	1.56	1.77	1	1
1:A:47:LYS:CE	1:A:47:LYS:CG	1.55	1.85	1	1
1:A:47:LYS:CD	1:A:47:LYS:CB	1.46	1.90	1	1
1:A:39:ARG:CG	1:A:39:ARG:CA	1.44	1.95	1	1
1:A:47:LYS:CA	1:A:47:LYS:CG	1.44	1.95	1	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	37/63 (59%)	15±2 (42±7%)	13±3 (35±7%)	9±2 (24±4%)	0	1
All	All	777/1323 (59%)	324 (42%)	270 (35%)	183 (24%)	0	1

5 of 25 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	44	CYS	21
1	A	30	HIS	21
1	A	37	HIS	20
1	A	21	PRO	16
1	A	40	SER	13

### 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	32/57 (56%)	23±1 (73±4%)	9±1 (27±4%)	2	23
All	All	672/1197 (56%)	493 (73%)	179 (27%)	2	23

5 of 21 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	54	ILE	21
1	A	35	TYR	21
1	A	28	THR	20
1	A	41	LEU	17
1	A	22	THR	16

### 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 1 ligands modelled in this entry, 1 is 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 39% for the well-defined parts and 34% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 5901

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	256
Number of shifts mapped to atoms	256
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
$^{13}\text{C}_\alpha$	0	—	None (insufficient data)
$^{13}\text{C}_\beta$	0	—	None (insufficient data)
$^{13}\text{C}'$	0	—	None (insufficient data)
$^{15}\text{N}$	40	$2.20 \pm 1.13$	None needed (imprecise)

#### 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 39%, i.e. 168 atoms were assigned a chemical shift out of a possible 430. 0 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	85/173 (49%)	57/68 (84%)	0/74 (0%)	28/31 (90%)
Sidechain	70/221 (32%)	70/134 (52%)	0/79 (0%)	0/8 (0%)

*Continued on next page...*

*Continued from previous page...*

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	13/36 (36%)	13/20 (65%)	0/12 (0%)	0/4 (0%)
Overall	168/430 (39%)	140/222 (63%)	0/165 (0%)	28/43 (65%)

#### 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:

