



# wwPDB NMR Structure Validation Summary Report ⓘ

Oct 30, 2024 – 06:58 AM EDT

PDB ID : 2MGS  
BMRB ID : 19601  
Title : Solution structure of CXCL5  
Authors : Sepuru, K.; Rajarathnam, K.  
Deposited on : 2013-11-04

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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
wwPDB-RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
wwPDB-ShiftChecker : v1.2  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.39

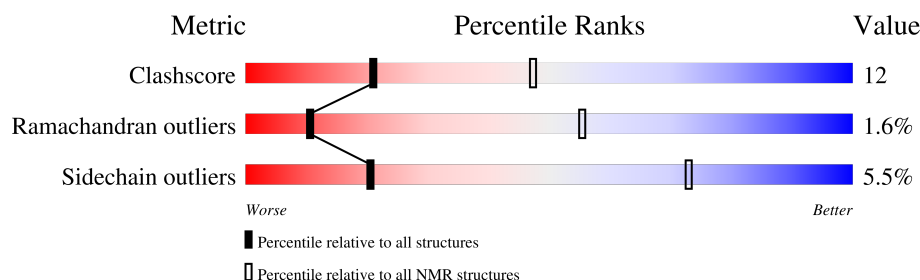
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 20%.

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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	78	
1	B	78	

## 2 Ensemble composition and analysis

This entry contains 20 models. Model 18 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:8-A:70, B:8-B:70 (126)	0.56	18

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

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

There is only 1 type of molecule in this entry. The entry contains 2240 atoms, of which 1150 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called C-X-C motif chemokine 5.

Mol	Chain	Residues	Atoms						Trace
1	A	71	Total	C	H	N	O	S	0
			1120	344	575	96	100	5	
1	B	71	Total	C	H	N	O	S	0
			1120	344	575	96	100	5	



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *DGSA-distance geometry simulated annealing, simulated annealing*.

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
ARIA	structure solution	2.3
CNS	structure solution	1.2
CNS	refinement	1.2

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	438
Number of shifts mapped to atoms	396
Number of unparsed shifts	0
Number of shifts with mapping errors	42
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	20%

## 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	0.46±0.01	0±0/493 ( 0.0± 0.0%)	0.82±0.01	2±0/665 ( 0.3± 0.0%)
1	B	0.46±0.00	0±0/493 ( 0.0± 0.0%)	0.82±0.01	2±0/665 ( 0.3± 0.0%)
All	All	0.46	0/19720 ( 0.0%)	0.82	80/26600 ( 0.3%)

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	1.0±0.0
1	B	0.0±0.0	1.0±0.0
All	All	0	40

There are no bond-length outliers.

All 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	B	45	VAL	CG1-CB-CG2	-7.20	99.38	110.90	5	20
1	A	45	VAL	CG1-CB-CG2	-7.18	99.41	110.90	5	20
1	A	21	GLY	C-N-CA	5.84	136.30	121.70	8	20
1	B	21	GLY	C-N-CA	5.82	136.25	121.70	18	20

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	12	ARG	Sidechain	20
1	B	12	ARG	Sidechain	20

## 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	486	523	522	12±2
1	B	486	523	522	12±1
All	All	19440	20920	20880	470

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 12.

5 of 48 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:56:LEU:HB3	1:A:63:LEU:HD21	0.76	1.56	5	18
1:B:56:LEU:HB3	1:B:63:LEU:HD21	0.73	1.59	3	20
1:A:56:LEU:HB3	1:A:63:LEU:HD11	0.65	1.68	13	18
1:A:44:VAL:HG22	1:A:58:PRO:HG3	0.60	1.74	10	2
1:B:56:LEU:HB3	1:B:63:LEU:HD11	0.59	1.72	5	18

## 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	62/78 (79%)	57±1 (91±1%)	4±1 (7±1%)	1±0 (2±0%)	10	55
1	B	62/78 (79%)	57±1 (91±1%)	4±1 (7±1%)	1±0 (2±0%)	10	55
All	All	2480/3120 (79%)	2266 (91%)	174 (7%)	40 (2%)	10	55

All 2 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	22	VAL	20

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	B	22	VAL	20

### 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	57/65 (88%)	54±1 (94±1%)	3±1 (6±1%)	20	73
1	B	57/65 (88%)	54±0 (94±1%)	3±0 (6±1%)	20	73
All	All	2280/2600 (88%)	2154 (94%)	126 (6%)	20	73

5 of 15 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	25	LYS	20
1	A	63	LEU	20
1	B	25	LYS	20
1	B	63	LEU	20
1	B	45	VAL	17

### 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 oligosaccharides 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 20% for the well-defined parts and 20% for the entire structure.

### 7.1 Chemical shift list 1

File name: working\_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	438
Number of shifts mapped to atoms	396
Number of unparsed shifts	0
Number of shifts with mapping errors	42
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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

- No matching atom found in the structure. First 5 (of 42) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	1	ALA	HA	4.33	0.020	1
1	A	1	ALA	HB1	1.57	0.020	1
1	A	1	ALA	HB2	1.57	0.020	1
1	A	1	ALA	HB3	1.57	0.020	1
1	A	1	ALA	CA	49.2	0.200	1
1	A	1	ALA	CB	16.6	0.200	1
1	A	2	GLY	H	8.486	0.02	1
1	A	2	GLY	CA	43.1	0.200	1
1	A	2	GLY	N	108.774	0.20	1
1	A	3	PRO	HA	4.71	0.020	1
1	A	3	PRO	HB2	2.48	0.020	2
1	A	3	PRO	HB3	2.7	0.020	2
1	A	3	PRO	CA	59.3	0.200	1
1	A	3	PRO	CB	31.8	0.200	1

*Continued on next page...*

*Continued from previous page...*

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	4	ALA	H	8.561	0.02	1
1	A	4	ALA	HA	4.68	0.020	1
1	A	4	ALA	HB1	1.8	0.020	1
1	A	4	ALA	HB2	1.8	0.020	1
1	A	4	ALA	HB3	1.8	0.020	1
1	A	4	ALA	CA	49.6	0.200	1
1	A	4	ALA	CB	16.4	0.200	1
1	A	4	ALA	N	124.234	0.20	1
1	A	5	ALA	H	8.138	0.02	1
1	A	5	ALA	HA	4.5	0.020	1
1	A	5	ALA	HB1	1.83	0.020	1
1	A	5	ALA	HB2	1.83	0.020	1
1	A	5	ALA	HB3	1.83	0.020	1
1	A	5	ALA	CA	51.6	0.200	1
1	A	5	ALA	CB	16.1	0.200	1
1	A	5	ALA	N	123.019	0.20	1
1	A	6	ALA	H	8.551	0.02	1
1	A	6	ALA	HA	4.33	0.020	1
1	A	6	ALA	HB1	1.89	0.020	1
1	A	6	ALA	HB2	1.89	0.020	1
1	A	6	ALA	HB3	1.89	0.020	1
1	A	6	ALA	CA	51.3	0.200	1
1	A	6	ALA	CB	15.9	0.200	1
1	A	6	ALA	N	118.75	0.20	1
1	A	7	VAL	H	7.514	0.02	1
1	A	7	VAL	CA	61.0	0.200	1
1	A	7	VAL	CB	29.3	0.200	1
1	A	7	VAL	N	117.775	0.20	1

### 7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, <i>ppm</i>	Suggested action
$^{13}\text{C}_\alpha$	66	$1.98 \pm 0.29$	Should be checked
$^{13}\text{C}_\beta$	62	$3.00 \pm 0.48$	Should be checked
$^{13}\text{C}'$	0	—	None (insufficient data)
$^{15}\text{N}$	70	$1.60 \pm 0.91$	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 20%, i.e. 349 atoms were assigned a chemical shift out of a possible 1772. 0 out of 28 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	219/620 (35%)	110/250 (44%)	52/252 (21%)	57/118 (48%)
Sidechain	130/1098 (12%)	70/718 (10%)	60/340 (18%)	0/40 (0%)
Aromatic	0/54 (0%)	0/28 (0%)	0/24 (0%)	0/2 (0%)
Overall	349/1772 (20%)	180/996 (18%)	112/616 (18%)	57/160 (36%)

### 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. If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:

