



wwPDB NMR Structure Validation Summary Report ⓘ

Jun 21, 2021 – 12:05 PM BST

PDB ID : 6ZFY
Title : Solution NMR structure of human GATA2 N-terminal zinc finger domain
Authors : Nurmohamed, S.S.; Broadhurst, R.W.; May, G.; Enver, T.
Deposited on : 2020-06-18

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at 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:

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	:	2.20
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.20

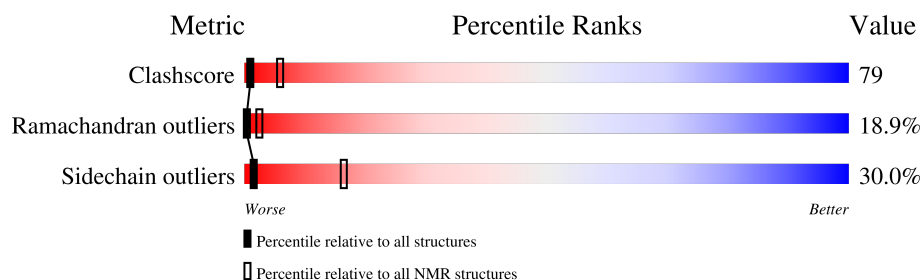
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

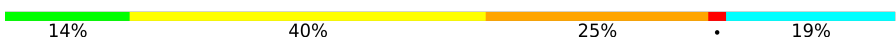
The overall completeness of chemical shifts assignment is 62%.

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 ≥ 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 20 models. Model 2 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:13-A:63 (51)	0.58	2

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 962 atoms, of which 468 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Endothelial transcription factor GATA-2.

Mol	Chain	Residues	Atoms						Trace
1	A	63	Total	C	H	N	O	S	0
			961	302	468	100	85	6	

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P23769
A	2	ALA	-	expression tag	UNP P23769
A	3	HIS	-	expression tag	UNP P23769
A	4	HIS	-	expression tag	UNP P23769
A	5	HIS	-	expression tag	UNP P23769
A	6	HIS	-	expression tag	UNP P23769
A	7	HIS	-	expression tag	UNP P23769
A	8	HIS	-	expression tag	UNP P23769
A	9	SER	-	expression tag	UNP P23769
A	10	SER	-	expression tag	UNP P23769
A	11	GLY	-	expression tag	UNP P23769
A	12	LEU	-	expression tag	UNP P23769
A	13	GLU	-	expression tag	UNP P23769
A	14	VAL	-	expression tag	UNP P23769
A	15	LEU	-	expression tag	UNP P23769
A	16	PHE	-	expression tag	UNP P23769
A	17	GLN	-	expression tag	UNP P23769
A	18	GLY	-	expression tag	UNP P23769
A	19	PRO	-	expression tag	UNP P23769

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

Mol	Chain	Residues	Atoms	
2	A	1	Total	Zn
			1	1

5 Refinement protocol and experimental data overview

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

Of the 50 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	refinement	2.3

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	474
Number of shifts mapped to atoms	474
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	62%

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	0.89±0.12	1±1/405 (0.3± 0.3%)	0.95±0.04	0±0/545 (0.1± 0.1%)
All	All	0.90	26/8100 (0.3%)	0.96	10/10900 (0.1%)

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	0.1±0.4
All	All	0	3

All 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	43	TYR	CE2-CZ	-11.08	1.24	1.38	4	5
1	A	43	TYR	CE1-CZ	10.13	1.51	1.38	4	5
1	A	51	TYR	CE1-CZ	8.51	1.49	1.38	3	8
1	A	51	TYR	CE2-CZ	-8.49	1.27	1.38	3	8

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	35	TRP	N-CA-CB	-6.48	98.93	110.60	1	10

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	26	ASN	Peptide	3

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	396	380	378	61±6
All	All	7940	7600	7560	1223

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 79.

5 of 269 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:24:CYS:SG	1:A:34:LEU:HB3	0.93	2.03	18	3
1:A:33:PRO:HB2	1:A:45:CYS:SG	0.91	2.06	12	3
1:A:34:LEU:HD23	1:A:48:CYS:HB2	0.88	1.43	3	19
1:A:34:LEU:HB2	1:A:45:CYS:SG	0.85	2.10	16	5
1:A:27:CYS:HB3	1:A:48:CYS:SG	0.83	2.14	14	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	50/63 (79%)	25±2 (50±4%)	16±2 (31±4%)	9±1 (19±3%)	0	2
All	All	1000/1260 (79%)	501 (50%)	310 (31%)	189 (19%)	0	2

5 of 24 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	29	ALA	20
1	A	33	PRO	20
1	A	38	ASP	20
1	A	60	PRO	20
1	A	28	GLY	18

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	41/51 (80%)	29±1 (70±3%)	12±1 (30±3%)	1	16
All	All	820/1020 (80%)	574 (70%)	246 (30%)	1	16

5 of 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	34	LEU	20
1	A	35	TRP	20
1	A	43	TYR	20
1	A	50	LEU	20
1	A	27	CYS	19

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 monosaccharides 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 62% for the well-defined parts and 54% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *NF_160209_c.csdep.txt*

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	474
Number of shifts mapped to atoms	474
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
$^{13}\text{C}_\alpha$	46	-0.29 ± 0.20	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	39	-0.02 ± 0.16	None needed (< 0.5 ppm)
$^{13}\text{C}'$	0	—	None (insufficient data)
^{15}N	47	0.27 ± 0.48	None needed (< 0.5 ppm)

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 62%, i.e. 386 atoms were assigned a chemical shift out of a possible 622. 0 out of 7 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	174/249 (70%)	91/99 (92%)	40/102 (39%)	43/48 (90%)
Sidechain	175/320 (55%)	110/190 (58%)	60/110 (55%)	5/20 (25%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	37/53 (70%)	19/27 (70%)	17/21 (81%)	1/5 (20%)
Overall	386/622 (62%)	220/316 (70%)	117/233 (50%)	49/73 (67%)

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:

