



# wwPDB NMR Structure Validation Summary Report ⓘ

Apr 20, 2024 – 01:45 PM EDT

PDB ID : 1GAU  
Title : SOLUTION STRUCTURE OF THE SPECIFIC DNA COMPLEX OF THE ZINC CONTAINING DNA BINDING DOMAIN OF THE ERYTHROID TRANSCRIPTION FACTOR GATA-1 BY MULTIDIMENSIONAL NMR  
Authors : Clore, G.M.; Omichinski, J.G.; Gronenborn, A.M.  
Deposited on : 1993-06-28

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 : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
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) : 2.36.2

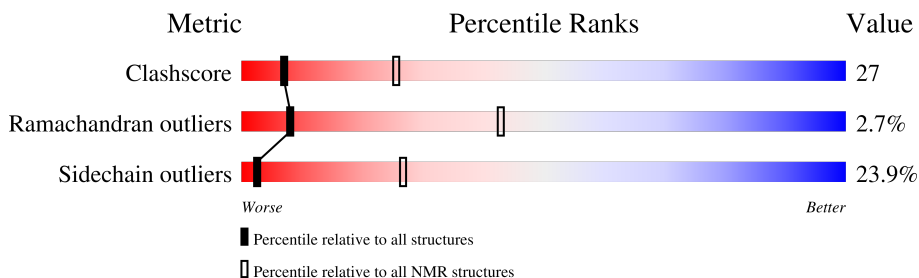
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment was not calculated.

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	B	8	
2	C	8	
3	A	60	

## 2 Ensemble composition and analysis i

This entry contains 30 models. Model 30 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:3-A:57 (55)	0.79	30

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 4 clusters and 2 single-model clusters were found.

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

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

There are 3 unique types of molecules in this entry. The entry contains 1458 atoms, of which 660 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(P\*AP\*GP\*AP\*TP\*AP\*AP\*AP\*C)-3').

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		P
1	B	8	255	79	89	35	44	8	0

- Molecule 2 is a DNA chain called DNA (5'-D(P\*GP\*TP\*TP\*TP\*AP\*TP\*CP\*T)-3').

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		P
2	C	8	255	79	93	23	52	8	0

- Molecule 3 is a protein called ERYTHROID TRANSCRIPTION FACTOR GATA-1.

Mol	Chain	Residues	Atoms					Trace	
			Total	C	H	N	O		S
3	A	60	948	284	478	95	85	6	0


## 4 Residue-property plots

### 4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: DNA (5'-D(P\*AP\*GP\*AP\*TP\*AP\*AP\*AP\*C)-3')

Chain B: 




- Molecule 2: DNA (5'-D(P\*GP\*TP\*TP\*TP\*AP\*TP\*CP\*T)-3')

Chain C: 



- Molecule 3: ERYTHROID TRANSCRIPTION FACTOR GATA-1

Chain A: 

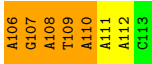


### 4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 30. Colouring as in section 4.1 above.

- Molecule 1: DNA (5'-D(P\*AP\*GP\*AP\*TP\*AP\*AP\*AP\*C)-3')

Chain B: 



- Molecule 2: DNA (5'-D(P\*GP\*TP\*TP\*TP\*AP\*TP\*CP\*T)-3')

Chain C:  25% 75%

G120  
T121  
T122  
T123  
A124  
T125  
C126  
T127

● Molecule 3: ERYTHROID TRANSCRIPTION FACTOR GATA-1

Chain A:  42% 42% 8% 8%

K1  
R2  
V6  
C7  
C10  
Q11  
S13  
T14  
T15  
T16  
L17  
W18  
R19  
R20  
M23  
V27  
C28  
N29  
A30  
G31  
G32  
L33  
Y34  
Y35  
K36  
L37  
H38  
Q39  
V40  
N41  
R42  
P43  
L44  
T45  
M46  
R47  
Q52  
T53  
R54  
N55  
R56  
K57  
V58  
S59  
S60

## 5 Refinement protocol and experimental data overview

The models were refined using the following method: ?.

Of the ? calculated structures, 30 were deposited, based on the following criterion: ?.

The authors did not provide any information on software used for structure solution, optimization or refinement.

No chemical shift data was provided.

## 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	B	1.48±0.04	2±0/187 ( 1.1± 0.1%)	3.06±0.01	20±1/286 ( 6.9± 0.2%)
2	C	1.46±0.02	4±1/179 ( 2.4± 0.4%)	2.56±0.01	16±1/274 ( 5.9± 0.4%)
3	A	1.05±0.01	0±0/439 ( 0.0± 0.0%)	0.91±0.01	0±0/594 ( 0.0± 0.0%)
All	All	1.26	189/24150 ( 0.8%)	2.08	1079/34620 ( 3.1%)

5 of 8 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	B	106	DA	N9-C8	-6.19	1.32	1.37	18	30
2	C	124	DA	N9-C8	-6.04	1.32	1.37	26	30
2	C	122	DT	C5-C7	5.67	1.53	1.50	19	30
1	B	109	DT	C5-C7	5.66	1.53	1.50	25	30
2	C	127	DT	C5-C7	5.46	1.53	1.50	17	30

5 of 41 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
2	C	120	DG	N7-C8-N9	14.53	120.36	113.10	10	30
1	B	107	DG	N7-C8-N9	14.32	120.26	113.10	20	30
1	B	106	DA	N7-C8-N9	13.49	120.55	113.80	12	30
2	C	124	DA	N7-C8-N9	13.45	120.52	113.80	26	30
1	B	111	DA	N7-C8-N9	12.84	120.22	113.80	18	30

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	B	166	89	90	9±3
2	C	162	93	94	10±3
3	A	431	431	435	25±5
All	All	22770	18390	18570	1110

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 27.

5 of 265 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:107:DG:O6	3:A:17:LEU:HD11	0.86	1.69	13	13
2:C:124:DA:OP2	3:A:30:ALA:HB2	0.84	1.73	20	3
3:A:7:CYS:HB2	3:A:12:THR:HG22	0.77	1.55	8	7
2:C:125:DT:H71	3:A:16:THR:HB	0.74	1.57	4	6
3:A:38:HIS:O	3:A:40:VAL:HG22	0.72	1.85	26	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	
3	A	55/60 (92%)	43±2 (78±4%)	11±2 (19±4%)	2±1 (3±2%)	8	43
All	All	1650/1800 (92%)	1287 (78%)	318 (19%)	45 (3%)	8	43

5 of 10 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
3	A	11	GLN	15

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
3	A	5	THR	8
3	A	49	ASP	6
3	A	55	ASN	5
3	A	18	TRP	3

### 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
3	A	49/54 (91%)	37±2 (76±4%)	12±2 (24±4%)	<b>2</b> 27
All	All	1470/1620 (91%)	1118 (76%)	352 (24%)	<b>2</b> 27

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)
3	A	53	THR	30
3	A	44	LEU	24
3	A	47	ARG	23
3	A	54	ARG	21
3	A	19	ARG	20

### 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](#)

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

No chemical shift data were provided