Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 11:33 AM EDT

PDB ID	:	1GAT
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 Full wwPDB NMR Structure Validation 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 (i) 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 (i)) were used in the production of this report:

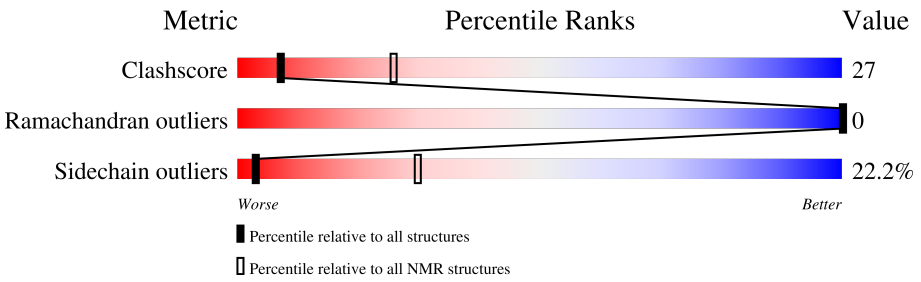
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

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 >=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 <=5%

Mol	Chain	Length	Quality of chain
1	B	8	12% 38% 50%
2	C	8	50% 50%
3	A	60	57% 38% 5%

2 Ensemble composition and analysis (i)

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1459 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

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
4	A	1	1	1

4 Residue-property plots (i)

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')

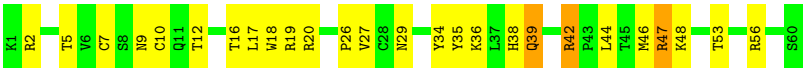
Chain B:

• 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:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *?*.

Of the ? calculated structures, 1 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)

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	B	1.44	2/187 (1.1%)	3.06	20/286 (7.0%)
2	C	1.46	3/179 (1.7%)	2.56	14/274 (5.1%)
3	A	1.05	0/477 (0.0%)	0.95	0/642 (0.0%)
All	All	1.24	5/843 (0.6%)	2.05	34/1202 (2.8%)

All bond outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	124	DA	N9-C8	-5.42	1.33	1.37
1	B	109	DT	C5-C7	5.35	1.53	1.50
2	C	122	DT	C5-C7	5.33	1.53	1.50
2	C	127	DT	C5-C7	5.14	1.53	1.50
1	B	106	DA	N9-C8	-5.12	1.33	1.37

All angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	120	DG	N7-C8-N9	14.21	120.20	113.10
1	B	107	DG	N7-C8-N9	13.67	119.93	113.10
2	C	124	DA	N7-C8-N9	13.22	120.41	113.80
1	B	106	DA	N7-C8-N9	13.08	120.34	113.80
1	B	111	DA	N7-C8-N9	12.54	120.07	113.80
1	B	108	DA	N7-C8-N9	12.38	119.99	113.80
1	B	110	DA	N7-C8-N9	12.24	119.92	113.80
1	B	112	DA	N7-C8-N9	12.12	119.86	113.80
1	B	107	DG	C8-N9-C4	-9.76	102.49	106.40
2	C	120	DG	C8-N9-C4	-9.51	102.59	106.40
1	B	110	DA	C8-N9-C4	-8.38	102.45	105.80
1	B	112	DA	C8-N9-C4	-8.20	102.52	105.80
1	B	111	DA	C8-N9-C4	-8.20	102.52	105.80

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	124	DA	C5-N7-C8	-7.99	99.91	103.90
2	C	120	DG	C5-N7-C8	-7.93	100.34	104.30
1	B	108	DA	C8-N9-C4	-7.89	102.65	105.80
2	C	124	DA	C8-N9-C4	-7.83	102.67	105.80
1	B	108	DA	C5-N7-C8	-7.76	100.02	103.90
1	B	107	DG	C5-N7-C8	-7.75	100.42	104.30
1	B	106	DA	C5-N7-C8	-7.75	100.03	103.90
1	B	106	DA	C8-N9-C4	-7.68	102.73	105.80
1	B	111	DA	C5-N7-C8	-7.55	100.12	103.90
1	B	112	DA	C5-N7-C8	-7.53	100.14	103.90
1	B	110	DA	C5-N7-C8	-7.33	100.23	103.90
2	C	125	DT	O4'-C1'-N1	6.62	112.64	108.00
2	C	121	DT	C6-C5-C7	-6.46	119.02	122.90
2	C	123	DT	C6-C5-C7	-5.69	119.49	122.90
2	C	122	DT	C6-C5-C7	-5.61	119.54	122.90
2	C	127	DT	C6-C5-C7	-5.57	119.56	122.90
1	B	107	DG	O4'-C1'-N9	5.54	111.88	108.00
1	B	109	DT	C6-C5-C7	-5.52	119.59	122.90
2	C	121	DT	O4'-C1'-N1	5.40	111.78	108.00
2	C	125	DT	C4-C5-C6	5.14	121.09	118.00
2	C	121	DT	C4-C5-C6	5.11	121.07	118.00

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	6
2	C	162	93	94	5
3	A	470	478	478	34
All	All	799	660	662	39

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.

All clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
3:A:34:TYR:CD1	3:A:42:ARG:HD3	0.86	2.05
1:B:107:DG:O6	3:A:17:LEU:HD11	0.84	1.73
3:A:10:CYS:SG	3:A:47:ARG:HD3	0.76	2.21
3:A:34:TYR:CE1	3:A:42:ARG:NH1	0.69	2.59
3:A:34:TYR:CE1	3:A:42:ARG:CZ	0.64	2.81
3:A:42:ARG:CZ	3:A:42:ARG:HA	0.62	2.24
3:A:39:GLN:CG	3:A:39:GLN:O	0.61	2.47
2:C:125:DT:C7	3:A:16:THR:HB	0.60	2.27
3:A:34:TYR:CD1	3:A:34:TYR:C	0.58	2.77
3:A:34:TYR:CZ	3:A:42:ARG:NH1	0.56	2.73
2:C:120:DG:H2"	2:C:121:DT:O5'	0.54	2.00
3:A:42:ARG:NH1	3:A:42:ARG:HG3	0.54	2.18
3:A:10:CYS:SG	3:A:12:THR:HG22	0.54	2.43
3:A:34:TYR:HD1	3:A:42:ARG:HD3	0.53	1.59
1:B:106:DA:H2"	1:B:107:DG:O5'	0.52	2.05
3:A:42:ARG:HG3	3:A:42:ARG:HH11	0.51	1.64
3:A:34:TYR:CD1	3:A:42:ARG:CD	0.50	2.88
3:A:17:LEU:HB2	3:A:29:ASN:HB2	0.50	1.82
3:A:35:TYR:O	3:A:39:GLN:N	0.48	2.45
2:C:121:DT:H2'	2:C:122:DT:H71	0.47	1.86
3:A:34:TYR:CZ	3:A:38:HIS:CD2	0.47	3.02
1:B:108:DA:H2"	1:B:109:DT:O5'	0.47	2.10
2:C:125:DT:H72	3:A:16:THR:HB	0.46	1.88
1:B:106:DA:N7	3:A:17:LEU:HD22	0.45	2.27
3:A:7:CYS:SG	3:A:27:VAL:HA	0.44	2.52
3:A:5:THR:HG23	3:A:26:PRO:CB	0.44	2.43
3:A:2:ARG:CG	3:A:18:TRP:CD1	0.44	3.01
3:A:42:ARG:NE	3:A:42:ARG:N	0.43	2.65
3:A:2:ARG:HG2	3:A:18:TRP:CD1	0.43	2.49
1:B:107:DG:N7	3:A:17:LEU:HD13	0.43	2.28
3:A:42:ARG:NE	3:A:42:ARG:CA	0.43	2.81
1:B:107:DG:O6	3:A:17:LEU:CD1	0.43	2.59
3:A:2:ARG:N	3:A:2:ARG:CD	0.42	2.82
3:A:10:CYS:SG	3:A:12:THR:CG2	0.42	3.07
3:A:35:TYR:O	3:A:39:GLN:CA	0.41	2.68
2:C:125:DT:H2"	2:C:126:DC:O5'	0.41	2.15
3:A:5:THR:O	3:A:18:TRP:CH2	0.41	2.74
3:A:5:THR:HG23	3:A:26:PRO:HB2	0.40	1.93
3:A:10:CYS:SG	3:A:47:ARG:CD	0.40	3.04

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	58/60 (97%)	45 (78%)	13 (22%)	0 (0%)	100	100
All	All	58/60 (97%)	45 (78%)	13 (22%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	54/54 (100%)	42 (78%)	12 (22%)	3	30
All	All	54/54 (100%)	42 (78%)	12 (22%)	3	30

All 12 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
3	A	9	ASN
3	A	19	ARG
3	A	20	ARG
3	A	36	LYS
3	A	39	GLN
3	A	42	ARG
3	A	44	LEU
3	A	46	MET
3	A	47	ARG
3	A	48	LYS
3	A	53	THR
3	A	56	ARG

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 (i)

No chemical shift data were provided