wwPDB NMR Structure Validation Summary Report (i)

Mar 6, 2022 – 11:38 PM EST

PDB ID	:	2KMK
Title	:	Gfi-1 Zinc Fingers 3-5 complexed with DNA
Authors	:	Lee, S.; Wu, Z.
Deposited on	:	2009-07-30

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 (i) symbol.

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

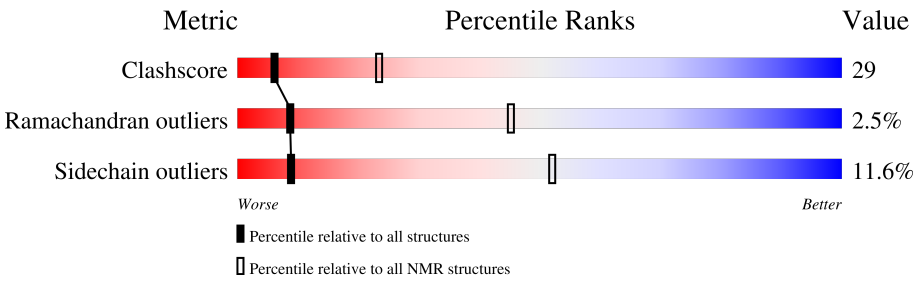
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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	A	82	45% 50%
2	B	16	31% 56% 12%
3	C	16	44% 38% 19%

2 Ensemble composition and analysis (i)

This entry contains 12 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:2-A:82 (81)	0.27	8

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 2 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 5, 7, 8, 9, 10, 11
2	4, 6
Single-model clusters	12

3 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2320 atoms, of which 1013 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Zinc finger protein Gfi-1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	82	1301	407	647	126	114	7	0

- Molecule 2 is a DNA chain called DNA (5'-D(*CP*AP*TP*AP*AP*AP*TP*CP*AP*CP*TP*GP*CP*CP*TP*A)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	16	502	155	182	58	92	15	0

- Molecule 3 is a DNA chain called DNA (5'-D(*TP*AP*GP*GP*CP*AP*GP*TP*GP*AP*TP*TP*TP*AP*TP*G)-3').

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
3	C	16	514	159	184	60	96	15	0

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

Mol	Chain	Residues	Atoms	
			Total	Zn
4	A	3	3	3

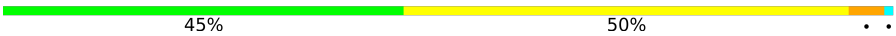
4 Residue-property plots (i)

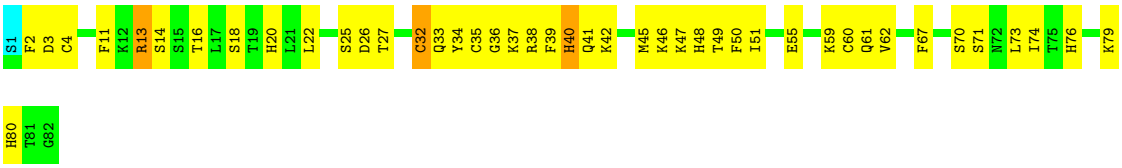
4.1 Average score per residue in the NMR ensemble

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: Zinc finger protein Gfi-1

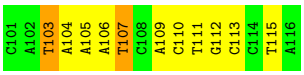
Chain A:





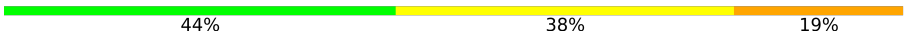
• Molecule 2: DNA (5'-D(*CP*AP*TP*AP*AP*AP*TP*CP*AP*CP*TP*GP*CP*CP*TP*A)-3')

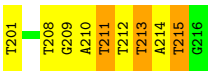
Chain B:



• Molecule 3: DNA (5'-D(*TP*AP*GP*GP*CP*AP*GP*TP*GP*AP*TP*TP*TP*AP*TP*G)-3')

Chain C:



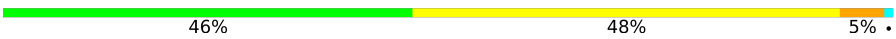


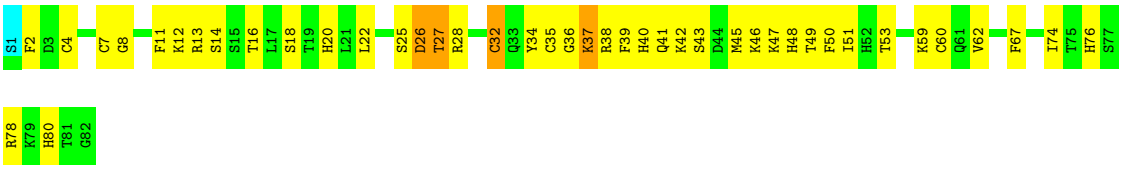
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 8. Colouring as in section 4.1 above.

• Molecule 1: Zinc finger protein Gfi-1

Chain A:



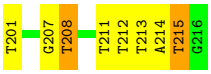


• Molecule 2: DNA (5'-D(*CP*AP*TP*AP*AP*AP*TP*CP*AP*CP*TP*GP*CP*CP*TP*A)-3')

Chain B:

• Molecule 3: DNA (5'-D(*TP*AP*GP*GP*CP*AP*GP*TP*GP*AP*TP*TP*TP*AP*TP*G)-3')

Chain C:



5 Refinement protocol and experimental data overview (i)

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

Of the 50 calculated structures, 12 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
X-PLOR NIH	structure solution	
X-PLOR NIH	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	A	0.99±0.01	0±0/665 (0.0± 0.0%)	0.94±0.02	0±0/887 (0.0± 0.0%)
2	B	1.09±0.01	3±1/358 (0.9± 0.2%)	1.57±0.01	7±0/549 (1.3± 0.1%)
3	C	1.12±0.01	4±1/370 (1.1± 0.4%)	1.62±0.00	12±1/571 (2.1± 0.1%)
All	All	1.05	84/16716 (0.5%)	1.35	230/24084 (1.0%)

5 of 10 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
2	B	103	DT	C5-C7	6.01	1.53	1.50	11	12
2	B	111	DT	C5-C7	5.86	1.53	1.50	8	11
3	C	208	DT	C5-C7	5.36	1.53	1.50	10	11
3	C	211	DT	C5-C7	5.24	1.53	1.50	11	5
3	C	213	DT	C5-C7	5.21	1.53	1.50	8	8

5 of 22 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(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	103	DT	C6-C5-C7	-6.96	118.72	122.90	10	12
2	B	103	DT	C4-C5-C6	6.38	121.83	118.00	8	12
2	B	107	DT	C6-C5-C7	-6.08	119.25	122.90	11	12
3	C	208	DT	C6-C5-C7	-6.07	119.26	122.90	10	12
3	C	213	DT	C6-C5-C7	-6.06	119.27	122.90	3	12

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	A	648	640	640	42±6
2	B	320	182	182	19±4
3	C	330	184	184	14±5
All	All	15612	12072	12072	808

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 29.

5 of 306 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:52:HIS:O	1:A:54:GLY:N	0.91	2.02	1	1
1:A:3:ASP:O	1:A:4:CYS:O	0.90	1.88	6	1
1:A:44:ASP:OD1	1:A:45:MET:N	0.73	2.21	12	1
1:A:62:VAL:HG11	1:A:80:HIS:CG	0.73	2.19	8	7
1:A:48:HIS:O	1:A:51:ILE:HG12	0.72	1.84	5	12

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	80/82 (98%)	69±2 (87±3%)	9±2 (11±3%)	2±1 (2±1%)	9	45
All	All	960/984 (98%)	831 (87%)	105 (11%)	24 (2%)	9	45

5 of 13 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	60	CYS	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	61	GLN	3
1	A	42	LYS	3
1	A	37	LYS	3
1	A	36	GLY	2

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	75/76 (99%)	66±1 (88±2%)	9±1 (12±2%)	9 52
All	All	900/912 (99%)	796 (88%)	104 (12%)	9 52

5 of 24 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	13	ARG	11
1	A	32	CYS	11
1	A	40	HIS	9
1	A	60	CYS	9
1	A	4	CYS	8

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 3 ligands modelled in this entry, 3 are 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