WORLDWIDE PDB PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report ⓘ

May 29, 2020 – 04:47 am BST

PDB ID	:	2MOE
Title	:	Solution structure of MBD4 methyl-cytosine binding domain bound to methylated DNA
Authors	:	Williams, D.C.; Walavalkar, N.M.
Deposited on	:	2014-04-24

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

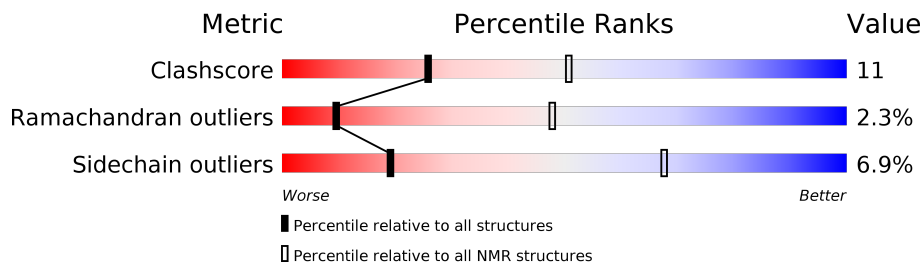|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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	Percentile Ranks	Value
Clashscore		11
Ramachandran outliers		2.3%
Sidechain outliers		6.9%

Worse .... *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all NMR structures

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	71	58% 23% 20%
2	B	10	70% 30%
3	C	10	90% 10%

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 4 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:87-A:143 (57)	0.50	4

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. No single-model clusters were found.

Cluster number	Models
1	2, 3, 4, 7, 8, 9, 12, 13, 15
2	1, 6, 10, 14, 18, 19
3	5, 11, 20
4	16, 17

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1779 atoms, of which 805 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Methyl-CpG-binding domain protein 4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	71	1142	363	574	101	102	2	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	78	GLY	-	EXPRESSION TAG	UNP O95243
A	79	SER	-	EXPRESSION TAG	UNP O95243

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	10	320	98	116	38	59	9	0

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

| Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
3	C	10	317	97	115	39	57	9	0

4   Residue-property plots (i)

4.1   Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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.
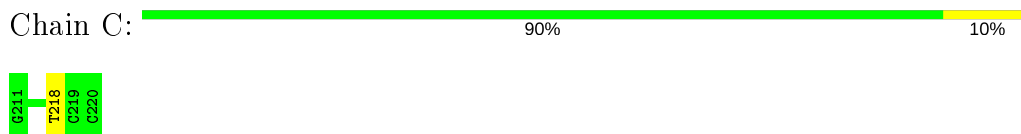
- Molecule 1: Methyl-CpG-binding domain protein 4
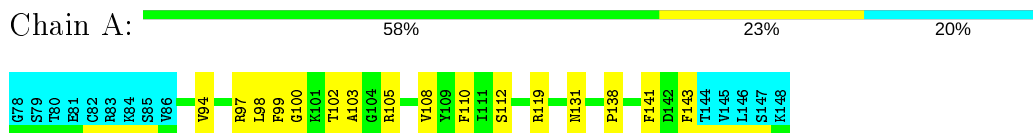
- Molecule 2: DNA (5'-D(*GP*GP*AP*TP*(5CM)P*GP*GP*CP*TP*C)-3')
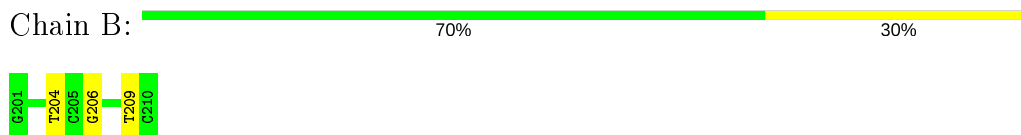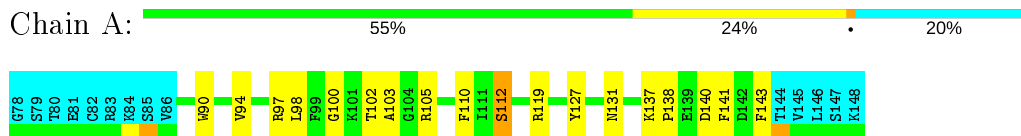

- Molecule 3: DNA (5'-D(*GP*AP*GP*CP*(5CM)P*GP*AP*TP*CP*C)-3')

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

The representative model is number 4. Colouring as in section 4.1 above.

- Molecule 1: Methyl-CpG-binding domain protein 4

- Molecule 2: DNA (5'-D(*GP*GP*AP*TP*(5CM)P*GP*GP*CP*TP*C)-3')

Chain B:

70%   30%

G201 T204 C205 G206 T209 C210

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

Chain C:

90%   10%

G211 T218 C219 C220

5 Refinement protocol and experimental data overview (i)

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 lowest energy.*

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

Software name	Classification	Version
X-PLOR NIH	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	3
Total number of shifts	892
Number of shifts mapped to atoms	892
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%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	467	461	460	18±4
2	B	204	116	116	1±1
3	C	202	115	115	1±1
All	All	17460	13840	13820	355

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 11.

5 of 156 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:98:LEU:HD22	1:A:98:LEU:N	0.80	1.91	19	7
1:A:102:THR:HG21	2:B:206:DG:N7	0.74	1.97	7	8
1:A:97:ARG:C	1:A:98:LEU:HD12	0.69	2.08	13	3
1:A:102:THR:HG21	2:B:206:DG:C8	0.67	2.25	9	5
1:A:97:ARG:NH1	1:A:102:THR:OG1	0.66	2.29	1	3

5.2 Torsion angles ⓘ

5.2.1 Protein backbone ⓘ

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	57/71 (80%)	54±1 (95±2%)	2±1 (3±2%)	1±1 (2±1%)	9	48
All	All	1140/1420 (80%)	1080 (95%)	34 (3%)	26 (2%)	9	48

5 of 7 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	103	ALA	8
1	A	100	GLY	7
1	A	98	LEU	4
1	A	143	PHE	4
1	A	134	THR	1

5.2.2 Protein sidechains ⓘ

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	50/63 (79%)	47±1 (93±3%)	3±1 (7±3%)	19	68
All	All	1000/1260 (79%)	931 (93%)	69 (7%)	19	68

5 of 18 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	112	SER	15
1	A	119	ARG	12
1	A	105	ARG	9
1	A	101	LYS	7
1	A	110	PHE	5

5.2.3   RNA (i)

There are no RNA molecules in this entry.

5.3   Non-standard residues in protein, DNA, RNA chains (i)

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
3	5CM	C	215	3,2	15,21,22	0.76±0.01	0±0 (0±0%)
2	5CM	B	205	3,2	15,21,22	0.76±0.01	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
3	5CM	C	215	3,2	19,30,33	1.17±0.02	0±0 (0±0%)
2	5CM	B	205	3,2	19,30,33	1.17±0.02	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	5CM	C	215	3,2	-	0±0,4,21,22	0±0,2,2,2
2	5CM	B	205	3,2	-	0±0,4,21,22	0±0,2,2,2

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 Ligand geometry (i)

There are no ligands in this entry.

5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 62% for the well-defined parts and 63% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *DNAfor_ chemshift*

6.1.1 Bookkeeping (i)

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	50
Number of shifts mapped to atoms	50
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

6.1.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

6.1.3 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 3%, i.e. 36 atoms were assigned a chemical shift out of a possible 1106. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	0/279 (0%)	0/111 (0%)	0/114 (0%)	0/54 (0%)
Sidechain	0/376 (0%)	0/225 (0%)	0/129 (0%)	0/22 (0%)
Aromatic	0/90 (0%)	0/48 (0%)	0/39 (0%)	0/3 (0%)
Overall	36/1106 (3%)	36/601 (6%)	0/401 (0%)	0/104 (0%)

6.1.4   Statistically unusual chemical shifts ⓘ

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
2	B	206	DG	H8	6.63	$8.85 - 6.75$	-5.6

6.1.5   Random Coil Index (RCI) plots ⓘ

No *random coil index* (RCI) plot could be generated from the current chemical shift list (DNAfor_chemshift). RCI is only applicable to proteins.

6.2   Chemical shift list 2

File name: input_cs.cif

Chemical shift list name: *DNArev_ chemshift*

6.2.1   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	44
Number of shifts mapped to atoms	44
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

6.2.2   Chemical shift referencing ⓘ

No chemical shift referencing corrections were calculated (not enough data).

6.2.3   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 3%, i.e. 34 atoms were assigned a chemical shift out of a possible 1106. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	0/279 (0%)	0/111 (0%)	0/114 (0%)	0/54 (0%)
Sidechain	0/376 (0%)	0/225 (0%)	0/129 (0%)	0/22 (0%)
Aromatic	0/90 (0%)	0/48 (0%)	0/39 (0%)	0/3 (0%)
Overall	34/1106 (3%)	34/601 (6%)	0/401 (0%)	0/104 (0%)

6.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

6.2.5 Random Coil Index (RCI) plots (i)

No *random coil index* (RCI) plot could be generated from the current chemical shift list (DNArev_chemshift). RCI is only applicable to proteins.

6.3 Chemical shift list 3

File name: input_cs.cif

Chemical shift list name: *MBD4mbd_ chemshift*

6.3.1 Bookkeeping (i)

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	798
Number of shifts mapped to atoms	798
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

6.3.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
^{13}C$_\alpha$	70	-0.42 ± 0.15	None needed (< 0.5 ppm)
^{13}C$_\beta$	65	0.21 ± 0.15	None needed (< 0.5 ppm)
^{13}C′	65	-0.08 ± 0.16	None needed (< 0.5 ppm)
^{15}N	65	-0.44 ± 0.55	None needed (< 0.5 ppm)

6.3.3 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 56%, i.e. 619 atoms were assigned a chemical shift out of a possible 1106. 1 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	275/279 (99%)	109/111 (98%)	112/114 (98%)	54/54 (100%)
Sidechain	300/376 (80%)	183/225 (81%)	109/129 (84%)	8/22 (36%)
Aromatic	44/90 (49%)	25/48 (52%)	18/39 (46%)	1/3 (33%)
Overall	619/1106 (56%)	317/601 (53%)	239/401 (60%)	63/104 (61%)

6.3.4 Statistically unusual chemical shifts (i)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	113	PRO	HD2	1.61	$5.45 - 1.85$	-5.7

6.3.5 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: