wwPDB X-ray Structure Validation Summary Report (i)

Sep 18, 2021 – 08:02 am BST

PDB ID	:	6YWW
Title	:	MeCP2 is a microsatellite binding protein that protects CA repeats from nucleosome invasion
Authors	:	Ibrahim, A.; Papin, C.; Mohideen-Abdul, K.; Gras, S.L.; Stoll, I.; Bronner, C.; Dimitrov, S.; Klaholz, B.P.; Hamiche, A.
Deposited on	:	2020-04-30
Resolution	:	2.10 Å(reported)

This is a wwPDB X-ray 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/XrayValidationReportHelp
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:

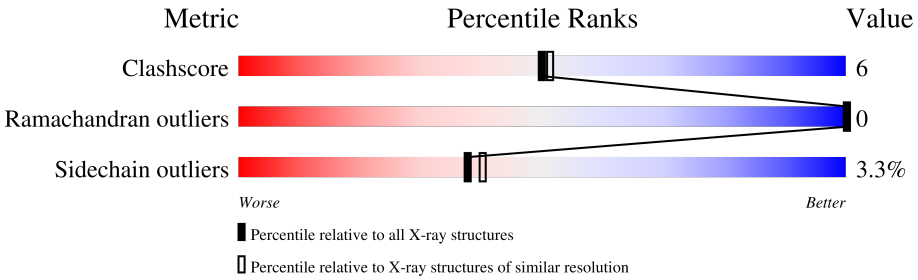
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 Å.

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)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	101	58% 11% • 30%
2	B	20	60% 35% 5%
3	C	20	85% 15%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 1468 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Truncated methyl CpG binding protein 2 transcript 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	71	Total	C	N	O	S	0	0	0
			570	363	98	108	1			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	76	MET	-	initiating methionine	UNP D5L9I4
A	162	ARG	-	expression tag	UNP D5L9I4
A	163	GLY	-	expression tag	UNP D5L9I4
A	164	ARG	-	expression tag	UNP D5L9I4
A	165	GLY	-	expression tag	UNP D5L9I4
A	166	SER	-	expression tag	UNP D5L9I4
A	167	PRO	-	expression tag	UNP D5L9I4
A	168	ALA	-	expression tag	UNP D5L9I4
A	169	ALA	-	expression tag	UNP D5L9I4
A	170	ALA	-	expression tag	UNP D5L9I4
A	171	HIS	-	expression tag	UNP D5L9I4
A	172	HIS	-	expression tag	UNP D5L9I4
A	173	HIS	-	expression tag	UNP D5L9I4
A	174	HIS	-	expression tag	UNP D5L9I4
A	175	HIS	-	expression tag	UNP D5L9I4
A	176	HIS	-	expression tag	UNP D5L9I4

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	20	Total	C	N	O	P	0	0	0
			403	195	73	116	19			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	20	Total	C	N	O	P	0	0	0
			413	199	74	121	19			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	24	Total	O	0	0
			24	24		
4	B	31	Total	O	0	0
			31	31		
4	C	27	Total	O	0	0
			27	27		

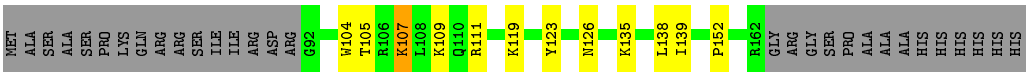
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-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 outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

Note EDS was not executed.

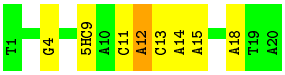
• Molecule 1: Truncated methyl CpG binding protein 2 transcript 1

Chain A:



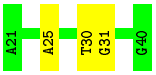
• Molecule 2: DNA/RNA (5'-D(*TP*CP*TP*GP*CP*AP*CP*A)-R(P*(5HC))-D(P*AP*CP*AP*CP*AP*AP*TP*TP*AP*TP*A)-3')

Chain B:



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

Chain C:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	64.38Å 48.99Å 66.19Å 90.00° 90.53° 90.00°	Depositor
Resolution (Å)	66.18 – 2.10	Depositor
% Data completeness (in resolution range)	50.9 (66.18-2.10)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.221 , 0.259	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1468	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 5HC

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 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	0.44	0/584	0.59	0/787
2	B	1.21	4/427 (0.9%)	1.08	1/653 (0.2%)
3	C	0.97	0/463	1.00	0/715
All	All	0.89	4/1474 (0.3%)	0.90	1/2155 (0.0%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	15	DA	O3'-P	-7.31	1.52	1.61
2	B	13	DC	O3'-P	-7.08	1.52	1.61
2	B	14	DA	O3'-P	-6.65	1.53	1.61
2	B	12	DA	O3'-P	-5.64	1.54	1.61

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	13	DC	C1'-O4'-C4'	-5.42	104.69	110.10

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the 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 within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	570	0	559	9	0
2	B	403	0	229	3	0
3	C	413	0	230	2	0
4	A	24	0	0	0	0
4	B	31	0	0	1	0
4	C	27	0	0	0	0
All	All	1468	0	1018	13	0

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 6.

The worst 5 of 13 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:30:DT:H2"	3:C:31:DG:O5'	1.86	0.76
1:A:104:TRP:CZ3	1:A:126:ASN:HB3	2.29	0.68
1:A:104:TRP:CZ3	1:A:126:ASN:CB	2.81	0.64
1:A:104:TRP:CE3	1:A:126:ASN:HA	2.34	0.63
1:A:111:ARG:HG3	1:A:119:LYS:O	2.10	0.51

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 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 X-ray entries followed by that with respect to entries of similar resolution.

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	69/101 (68%)	63 (91%)	6 (9%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 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 X-ray entries followed by that with respect to entries of similar resolution.

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	61/84 (73%)	59 (97%)	2 (3%)	38 40

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	105	THR
1	A	107	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or 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 length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	5HC	B	9	3,2	16,22,23	0.74	0	20,31,34	1.21	2 (10%)

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
2	5HC	B	9	3,2	-	5/6/23/24	0/2/2/2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	9	5HC	C2-N3-C4	3.99	120.83	116.02
2	B	9	5HC	C5-C4-N4	-2.31	118.91	122.19

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	9	5HC	C3'-C4'-C5'-O5'
2	B	9	5HC	O4'-C4'-C5'-O5'
2	B	9	5HC	C6-C5-C5M-O5
2	B	9	5HC	C4-C5-C5M-O5
2	B	9	5HC	C4'-C5'-O5'-P

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

6.4 Ligands

EDS was not executed - this section is therefore empty.

6.5 Other polymers

EDS was not executed - this section is therefore empty.