wwPDB X-ray Structure Validation Summary Report (i)

Apr 22, 2025 – 03:42 AM EDT

PDB ID	:	6C2F / pdb_00006c2f
Title	:	MBD2 in complex with methylated DNA
Authors	:	Liu, K.; Xu, C.; Tempel, W.; Arrowsmith, C.H.; Bountra, C.; Edwards, A.M.; Min, J.; Structural Genomics Consortium; Structural Genomics Consortium (SGC)
Deposited on	:	2018-01-08
Resolution	:	2.65 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 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:

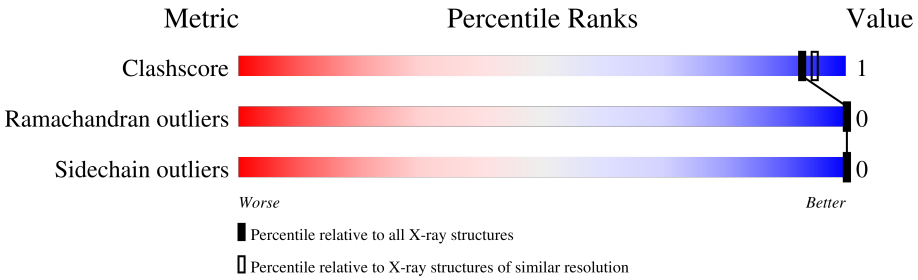
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0rc1
EDS	:	FAILED
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.42

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 2.65 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	180529	1063 (2.66-2.66)
Ramachandran outliers	177936	1052 (2.66-2.66)
Sidechain outliers	177891	1052 (2.66-2.66)

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 failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	79	81% 6% 13%
1	D	79	81% 5% 14%
1	G	79	81% 5% 14%
1	J	79	82% . 15%
1	M	79	84% . 14%
1	P	79	84% . 14%
2	B	12	58% 33% 8%
2	E	12	92% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	12	58% 42%
2	K	12	92% 8%
2	N	12	67% 33%
2	Q	12	92% 8%
3	C	12	92% 8%
3	F	12	100%
3	I	12	83% 17%
3	L	12	83% 17%
3	O	12	83% 17%
3	R	12	83% 17%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5915 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 Methyl-CpG-binding domain protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	69	Total	C	N	O	S	0	0	0
			508	328	85	92	3			
1	D	68	Total	C	N	O	S	0	0	0
			499	324	84	89	2			
1	G	68	Total	C	N	O	S	0	0	0
			502	321	88	91	2			
1	J	67	Total	C	N	O	S	0	0	0
			482	309	82	89	2			
1	M	68	Total	C	N	O	S	0	0	1
			493	316	85	90	2			
1	P	68	Total	C	N	O	S	0	0	0
			498	320	85	91	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	142	GLY	-	expression tag	UNP Q9UBB5
D	142	GLY	-	expression tag	UNP Q9UBB5
G	142	GLY	-	expression tag	UNP Q9UBB5
J	142	GLY	-	expression tag	UNP Q9UBB5
M	142	GLY	-	expression tag	UNP Q9UBB5
P	142	GLY	-	expression tag	UNP Q9UBB5

- Molecule 2 is a DNA chain called 12-mer DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	12	Total	C	N	O	P	0	0	0
			248	118	51	68	11			
2	E	12	Total	C	N	O	P	0	0	0
			248	118	51	68	11			
2	H	12	Total	C	N	O	P	0	0	0
			248	118	51	68	11			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	K	12	Total	C	N	O	P	0	0	0
			248	118	51	68	11			
2	N	12	Total	C	N	O	P	0	0	0
			248	118	51	68	11			
2	Q	12	Total	C	N	O	P	0	0	0
			248	118	51	68	11			

- Molecule 3 is a DNA chain called 12-mer DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	12	Total	C	N	O	P	0	0	0
			239	115	41	72	11			
3	F	12	Total	C	N	O	P	0	0	0
			239	115	41	72	11			
3	I	12	Total	C	N	O	P	0	0	0
			239	115	41	72	11			
3	L	12	Total	C	N	O	P	0	0	0
			239	115	41	72	11			
3	O	12	Total	C	N	O	P	0	0	0
			239	115	41	72	11			
3	R	12	Total	C	N	O	P	0	0	0
			239	115	41	72	11			

- Molecule 4 is UNKNOWN ATOM OR ION (CCD ID: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	X	0	0
			2	2		
4	D	2	Total	X	0	0
			2	2		
4	G	2	Total	X	0	0
			2	2		
4	J	2	Total	X	0	0
			2	2		
4	M	2	Total	X	0	0
			2	2		
4	P	1	Total	X	0	0
			1	1		

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.

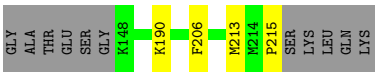
Note EDS failed to run properly.

• Molecule 1: Methyl-CpG-binding domain protein 2

Chain A:

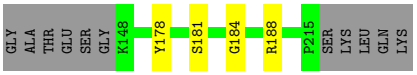
• Molecule 1: Methyl-CpG-binding domain protein 2

Chain D:



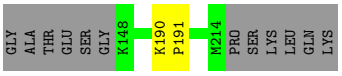
• Molecule 1: Methyl-CpG-binding domain protein 2

Chain G:



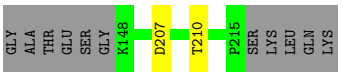
• Molecule 1: Methyl-CpG-binding domain protein 2

Chain J:



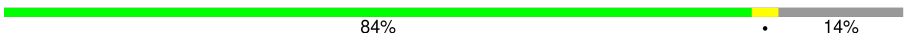
• Molecule 1: Methyl-CpG-binding domain protein 2

Chain M:



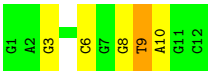
• Molecule 1: Methyl-CpG-binding domain protein 2

Chain P:



• Molecule 2: 12-mer DNA

Chain B:

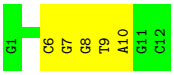


• Molecule 2: 12-mer DNA

Chain E:

• Molecule 2: 12-mer DNA

Chain H:



• Molecule 2: 12-mer DNA

Chain K:

• Molecule 2: 12-mer DNA

Chain N:

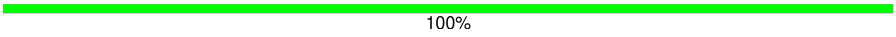
• Molecule 2: 12-mer DNA

Chain Q:

• Molecule 3: 12-mer DNA

Chain C:

• Molecule 3: 12-mer DNA

Chain F: 

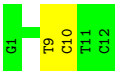
There are no outlier residues recorded for this chain.

• Molecule 3: 12-mer DNA

Chain I: 

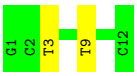
• Molecule 3: 12-mer DNA

Chain L: 



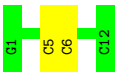
• Molecule 3: 12-mer DNA

Chain O: 



• Molecule 3: 12-mer DNA

Chain R: 



4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	40.51Å 40.60Å 202.38Å 90.07° 93.11° 119.75°	Depositor
Resolution (Å)	35.10 – 2.65	Depositor
% Data completeness (in resolution range)	92.0 (35.10-2.65)	Depositor
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.84 (at 2.54Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.211 , 0.242	Depositor
Wilson B-factor (Å^2)	77.1	Xtriage
Anisotropy	0.661	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.066 for k,-h-k,h+l 0.066 for -h-k,h,h+k+l 0.048 for h,-h-k,-h-l 0.054 for -h-k,k,-l 0.049 for k,h,-h-k-l	Xtriage
Total number of atoms	5915	wwPDB-VP
Average B, all atoms (Å^2)	87.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.66% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

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

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	0.43	0/522	0.63	2/707 (0.3%)
1	D	0.44	0/513	0.61	0/696
1	G	0.46	0/516	0.58	0/700
1	J	0.42	0/495	0.53	0/674
1	M	0.46	0/506	0.58	0/685
1	P	0.45	0/512	0.58	0/695
2	B	1.36	2/256 (0.8%)	1.19	1/392 (0.3%)
2	E	0.95	0/256	0.97	0/392
2	H	1.21	1/256 (0.4%)	1.09	1/392 (0.3%)
2	K	0.92	0/256	0.97	0/392
2	N	1.27	2/256 (0.8%)	1.14	1/392 (0.3%)
2	Q	0.96	0/256	0.95	0/392
3	C	1.33	1/266 (0.4%)	1.14	0/408
3	F	0.93	0/266	1.09	0/408
3	I	1.14	0/266	1.11	0/408
3	L	0.83	0/266	0.96	0/408
3	O	1.28	1/266 (0.4%)	1.20	1/408 (0.2%)
3	R	0.91	0/266	1.05	0/408
All	All	0.85	7/6196 (0.1%)	0.88	6/8957 (0.1%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	H	9	DT	C3'-O3'	-6.54	1.35	1.44
2	B	9	DT	C3'-O3'	-6.08	1.36	1.44
2	B	3	DG	C3'-O3'	-5.97	1.36	1.44
3	O	9	DT	C3'-O3'	-5.96	1.36	1.44
2	N	9	DT	C3'-O3'	-5.80	1.36	1.44

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	N	7	DG	O5'-P-OP2	-7.02	99.39	105.70
1	A	193	LEU	CA-CB-CG	5.88	128.83	115.30
2	B	8	DG	O5'-P-OP2	-5.45	100.80	105.70
2	H	10	DA	O4'-C1'-N9	-5.39	104.22	108.00
3	O	3	DT	O4'-C1'-N1	5.33	111.73	108.00

There are no chirality outliers.

There are no planarity outliers.

5.2 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 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	508	0	464	2	0
1	D	499	0	448	2	0
1	G	502	0	453	2	0
1	J	482	0	416	1	0
1	M	493	0	432	1	0
1	P	498	0	442	1	0
2	B	248	0	137	1	0
2	E	248	0	137	0	0
2	H	248	0	137	1	0
2	K	248	0	137	0	0
2	N	248	0	137	0	0
2	Q	248	0	137	0	0
3	C	239	0	137	0	0
3	F	239	0	137	0	0
3	I	239	0	137	1	0
3	L	239	0	137	1	0
3	O	239	0	137	0	0
3	R	239	0	137	1	0
4	A	2	0	0	0	0
4	D	2	0	0	0	0
4	G	2	0	0	0	0
4	J	2	0	0	0	0
4	M	2	0	0	0	0
4	P	1	0	0	0	0
All	All	5915	0	4299	14	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 1.

The worst 5 of 14 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)
1:A:205:SER:O	1:A:214:MET:N	2.35	0.56
1:D:190:LYS:NZ	1:D:206:PHE:O	2.35	0.53
3:I:9:DT:H2"	3:I:10:DC:C6	2.51	0.45
1:D:213:MET:O	1:D:215:PRO:HD3	2.16	0.45
1:M:207:ASP:OD2	1:M:210:THR:HG23	2.17	0.45

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	67/79 (85%)	67 (100%)	0	0	100	100
1	D	66/79 (84%)	65 (98%)	1 (2%)	0	100	100
1	G	66/79 (84%)	66 (100%)	0	0	100	100
1	J	65/79 (82%)	65 (100%)	0	0	100	100
1	M	66/79 (84%)	66 (100%)	0	0	100	100
1	P	66/79 (84%)	66 (100%)	0	0	100	100
All	All	396/474 (84%)	395 (100%)	1 (0%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 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 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	47/67 (70%)	47 (100%)	0	100	100
1	D	43/67 (64%)	43 (100%)	0	100	100
1	G	46/67 (69%)	46 (100%)	0	100	100
1	J	42/67 (63%)	42 (100%)	0	100	100
1	M	42/67 (63%)	42 (100%)	0	100	100
1	P	44/67 (66%)	44 (100%)	0	100	100
All	All	264/402 (66%)	264 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

6 non-standard protein/DNA/RNA residues are 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	5CM	K	6	3,2	18,21,22	1.68	3 (16%)	24,30,33	1.59	4 (16%)
2	5CM	B	6	3,2	18,21,22	1.55	3 (16%)	24,30,33	1.53	6 (25%)
2	5CM	Q	6	3,2	18,21,22	1.62	3 (16%)	24,30,33	1.61	4 (16%)
2	5CM	E	6	3,2	18,21,22	1.57	3 (16%)	24,30,33	1.71	5 (20%)
2	5CM	H	6	3,2	18,21,22	1.59	3 (16%)	24,30,33	1.61	4 (16%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	5CM	N	6	3,2	18,21,22	1.56	3 (16%)	24,30,33	1.77	6 (25%)

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	5CM	K	6	3,2	-	3/7/21/22	0/2/2/2
2	5CM	B	6	3,2	-	4/7/21/22	0/2/2/2
2	5CM	Q	6	3,2	-	3/7/21/22	0/2/2/2
2	5CM	E	6	3,2	-	2/7/21/22	0/2/2/2
2	5CM	H	6	3,2	-	4/7/21/22	0/2/2/2
2	5CM	N	6	3,2	-	5/7/21/22	0/2/2/2

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	K	6	5CM	C5-C4	5.61	1.48	1.44
2	Q	6	5CM	C5-C4	5.41	1.48	1.44
2	E	6	5CM	C5-C4	5.32	1.48	1.44
2	B	6	5CM	C5-C4	5.20	1.48	1.44
2	H	6	5CM	C5-C4	5.03	1.47	1.44

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	K	6	5CM	C5-C6-N1	-4.71	118.19	123.31
2	E	6	5CM	O4'-C1'-N1	4.44	115.74	107.86
2	N	6	5CM	C5-C6-N1	-4.13	118.83	123.31
2	Q	6	5CM	C5-C6-N1	-4.11	118.84	123.31
2	N	6	5CM	O4'-C1'-N1	3.65	114.34	107.86

There are no chirality outliers.

5 of 21 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	N	6	5CM	O4'-C4'-C5'-O5'
2	B	6	5CM	O4'-C4'-C5'-O5'
2	H	6	5CM	O4'-C4'-C5'-O5'

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	Q	6	5CM	O4'-C4'-C5'-O5'
2	N	6	5CM	O4'-C1'-N1-C6

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 11 ligands modelled in this entry, 11 are unknown - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

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 failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates

EDS failed to run properly - this section is therefore empty.

6.4 Ligands

EDS failed to run properly - this section is therefore empty.

6.5 Other polymers

EDS failed to run properly - this section is therefore empty.