



# Full wwPDB X-ray Structure Validation Report ⓘ

May 16, 2020 – 07:00 pm BST

PDB ID : 6JV3  
Title : Crystal structure of 5-hydroxymethylcytosine containing decamer dsDNA  
Authors : Zhang, L.; Wang, Y.X.  
Deposited on : 2019-04-15  
Resolution : 2.85 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

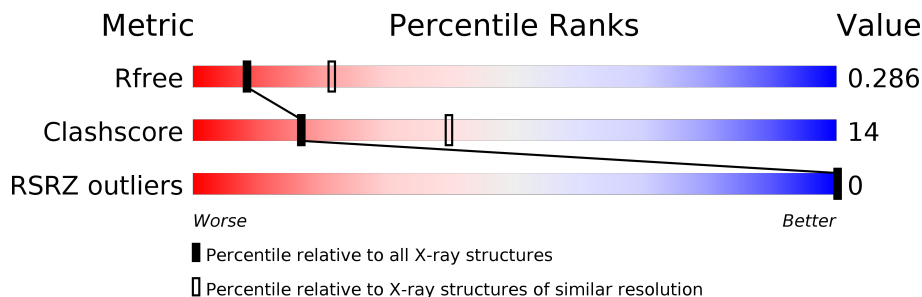
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.85 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))
$R_{free}$	130704	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	10	
1	B	10	
1	C	10	
1	D	10	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 816 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 DNA chain called DNA (5'-D(\*CP\*CP\*AP\*GP\*(5HC)P\*GP\*CP\*TP\*GP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	10	Total	C	N	O	P	0	0	0
			204	97	39	59	9			
1	B	10	Total	C	N	O	P	0	0	0
			204	97	39	59	9			
1	C	10	Total	C	N	O	P	0	0	0
			204	97	39	59	9			
1	D	10	Total	C	N	O	P	0	0	0
			204	97	39	59	9			

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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 and electron density. 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. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). 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.

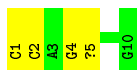
- Molecule 1: DNA (5'-D(\*CP\*CP\*AP\*GP\*(5HC)P\*GP\*CP\*TP\*GP\*G)-3')

Chain A: 




- Molecule 1: DNA (5'-D(\*CP\*CP\*AP\*GP\*(5HC)P\*GP\*CP\*TP\*GP\*G)-3')

Chain B: 



- Molecule 1: DNA (5'-D(\*CP\*CP\*AP\*GP\*(5HC)P\*GP\*CP\*TP\*GP\*G)-3')

Chain C: 



- Molecule 1: DNA (5'-D(\*CP\*CP\*AP\*GP\*(5HC)P\*GP\*CP\*TP\*GP\*G)-3')

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	63.91Å 23.82Å 85.56Å 90.00° 111.12° 90.00°	Depositor
Resolution (Å)	26.60 – 2.85 26.60 – 2.85	Depositor EDS
% Data completeness (in resolution range)	87.4 (26.60-2.85) 87.7 (26.60-2.85)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.66 (at 2.85Å)	Xtrriage
Refinement program	PHENIX (1.13_2998: ???)	Depositor
R, $R_{free}$	0.228 , 0.282 0.230 , 0.286	Depositor DCC
$R_{free}$ test set	130 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	57.5	Xtrriage
Anisotropy	0.316	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.23 , 34.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	0.019 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	816	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.87% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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: 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	0.52	0/204	0.85	0/311
1	B	0.55	0/204	0.79	0/311
1	C	0.53	0/204	0.86	0/311
1	D	0.60	0/204	0.95	0/311
All	All	0.55	0/816	0.86	0/1244

There are no bond length outliers.

There are no bond angle outliers.

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	204	0	115	5	0
1	B	204	0	115	4	0
1	C	204	0	115	3	0
1	D	204	0	114	5	0
All	All	816	0	459	16	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 14.

All (16) 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:C:5:5HC:N3	1:D:6:DG:N1	2.27	0.81
1:D:5:5HC:H5'	1:D:5:5HC:H6	1.63	0.78
1:D:5:5HC:O5	1:D:5:5HC:N4	2.08	0.71
1:B:5:5HC:N4	1:B:5:5HC:O5	2.11	0.58
1:A:5:5HC:HN42	1:A:5:5HC:HO5	1.52	0.57
1:C:5:5HC:O5	1:C:5:5HC:N4	2.16	0.57
1:C:5:5HC:H2''	1:C:6:DG:C8	2.42	0.55
1:D:9:DG:H2''	1:D:10:DG:C8	2.43	0.54
1:A:5:5HC:O5	1:A:5:5HC:N4	2.22	0.49
1:A:2:DC:H1'	1:A:3:DA:C5	2.49	0.47
1:A:5:5HC:H2''	1:A:6:DG:C8	2.50	0.46
1:B:4:DG:H2''	1:B:5:5HC:H5''	1.98	0.45
1:D:5:5HC:H2'	1:D:6:DG:C8	2.50	0.45
1:B:1:DC:H2'	1:B:2:DC:C6	2.52	0.44
1:A:8:DT:H2''	1:A:9:DG:N7	2.34	0.43
1:B:5:5HC:H2''	1:B:5:5HC:H5'	1.66	0.41

There are no symmetry-related clashes.

### 5.3 Torsion angles [i](#)

#### 5.3.1 Protein backbone [i](#)

There are no protein molecules in this entry.

#### 5.3.2 Protein sidechains [i](#)

There are no protein molecules in this entry.

#### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

4 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
1	5HC	C	5	1	16,22,23	1.55	2 (12%)	20,31,34	2.11	6 (30%)
1	5HC	B	5	1	16,22,23	1.53	2 (12%)	20,31,34	1.86	5 (25%)
1	5HC	D	5	1	16,22,23	1.61	3 (18%)	20,31,34	1.97	6 (30%)
1	5HC	A	5	1	16,22,23	1.56	2 (12%)	20,31,34	1.79	5 (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
1	5HC	C	5	1	-	2/6/23/24	0/2/2/2
1	5HC	B	5	1	-	5/6/23/24	0/2/2/2
1	5HC	D	5	1	-	4/6/23/24	0/2/2/2
1	5HC	A	5	1	-	3/6/23/24	0/2/2/2

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	5	5HC	C5M-C5	-4.90	1.39	1.51
1	D	5	5HC	C5M-C5	-4.87	1.39	1.51
1	A	5	5HC	C5M-C5	-4.87	1.39	1.51
1	C	5	5HC	C5M-C5	-4.67	1.39	1.51
1	C	5	5HC	C5-C4	-2.56	1.38	1.42
1	D	5	5HC	C5-C4	-2.45	1.38	1.42
1	A	5	5HC	C5-C4	-2.30	1.39	1.42
1	D	5	5HC	O4'-C1'	2.24	1.47	1.42
1	B	5	5HC	C5-C4	-2.16	1.39	1.42

All (22) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	5	5HC	C2-N3-C4	5.87	123.10	116.02
1	A	5	5HC	C2-N3-C4	4.24	121.14	116.02

*Continued on next page...*



Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	5	5HC	C2-N3-C4	4.11	120.98	116.02
1	D	5	5HC	C5-C4-N4	-3.96	116.57	122.19
1	B	5	5HC	C2-N3-C4	3.91	120.73	116.02
1	D	5	5HC	N4-C4-N3	3.76	122.35	117.03
1	B	5	5HC	N4-C4-N3	3.74	122.32	117.03
1	B	5	5HC	C5-C4-N4	-3.62	117.05	122.19
1	C	5	5HC	C5-C4-N4	-3.59	117.09	122.19
1	C	5	5HC	N4-C4-N3	3.58	122.10	117.03
1	A	5	5HC	C5-C4-N4	-3.57	117.12	122.19
1	A	5	5HC	N4-C4-N3	3.37	121.80	117.03
1	B	5	5HC	C2'-C3'-C4'	-3.11	96.29	102.76
1	A	5	5HC	O4'-C4'-C3'	-2.84	99.04	105.67
1	C	5	5HC	O4'-C4'-C3'	-2.80	99.13	105.67
1	D	5	5HC	C2'-C1'-N1	-2.73	107.97	114.27
1	C	5	5HC	C4'-O4'-C1'	-2.64	103.07	109.45
1	A	5	5HC	C4'-O4'-C1'	-2.33	103.83	109.45
1	D	5	5HC	C5M-C5-C6	2.15	121.25	119.05
1	D	5	5HC	C2'-C3'-C4'	-2.06	98.47	102.76
1	C	5	5HC	C6-N1-C1'	-2.05	114.64	119.24
1	B	5	5HC	C5-C6-N1	-2.04	118.70	121.51

There are no chirality outliers.

All (14) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	C	5	5HC	C4-C5-C5M-O5
1	C	5	5HC	C6-C5-C5M-O5
1	B	5	5HC	O4'-C1'-N1-C6
1	B	5	5HC	C4-C5-C5M-O5
1	B	5	5HC	C6-C5-C5M-O5
1	D	5	5HC	O4'-C1'-N1-C6
1	D	5	5HC	C4-C5-C5M-O5
1	D	5	5HC	C6-C5-C5M-O5
1	A	5	5HC	C4-C5-C5M-O5
1	A	5	5HC	C6-C5-C5M-O5
1	B	5	5HC	O4'-C4'-C5'-O5'
1	D	5	5HC	O4'-C4'-C5'-O5'
1	B	5	5HC	C3'-C4'-C5'-O5'
1	A	5	5HC	O4'-C4'-C5'-O5'

There are no ring outliers.

4 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	C	5	5HC	3	0
1	B	5	5HC	3	0
1	D	5	5HC	3	0
1	A	5	5HC	3	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å <sup>2</sup> )	Q < 0.9
1	A	9/10 (90%)	-0.75	0 100 100	24, 27, 40, 51	0
1	B	9/10 (90%)	-0.44	0 100 100	27, 29, 36, 37	0
1	C	9/10 (90%)	0.21	0 100 100	48, 56, 67, 67	0
1	D	9/10 (90%)	-0.07	0 100 100	19, 33, 65, 66	0
All	All	36/40 (90%)	-0.26	0 100 100	19, 34, 66, 67	0

There are no RSRZ outliers to report.

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q < 0.9
1	5HC	C	5	21/22	0.87	0.20	49,66,89,93	0
1	5HC	B	5	21/22	0.92	0.17	18,36,58,71	0
1	5HC	D	5	21/22	0.92	0.18	21,30,37,42	0
1	5HC	A	5	21/22	0.97	0.13	12,18,26,50	0

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.