



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 2, 2016 – 12:04 AM GMT

PDB ID : 9MHT  
Title : CYTOSINE-SPECIFIC METHYLTRANSFERASE HHAI/DNA COMPLEX  
Authors : O'Gara, M.; Horton, J.R.; Roberts, R.J.; Cheng, X.  
Deposited on : 1998-08-07  
Resolution : 2.39 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

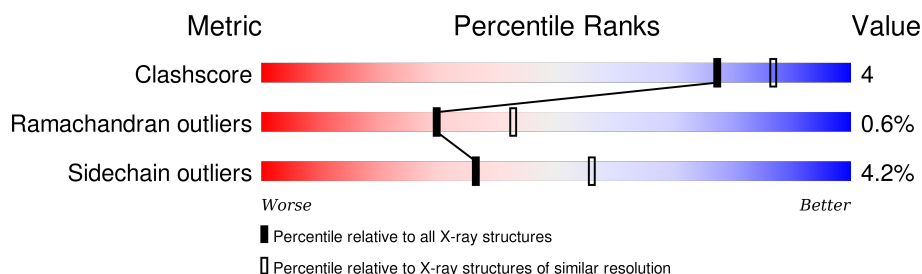
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.39 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	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)

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. 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	12	
2	D	12	
3	A	327	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3358 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 5'-D(P\*CP\*CP\*AP\*TP\*GP\*CP\*GP\*CP\*TP\*GP\*AP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	12	Total	C	N	O	P	0	0	0
			243	115	44	72	12			

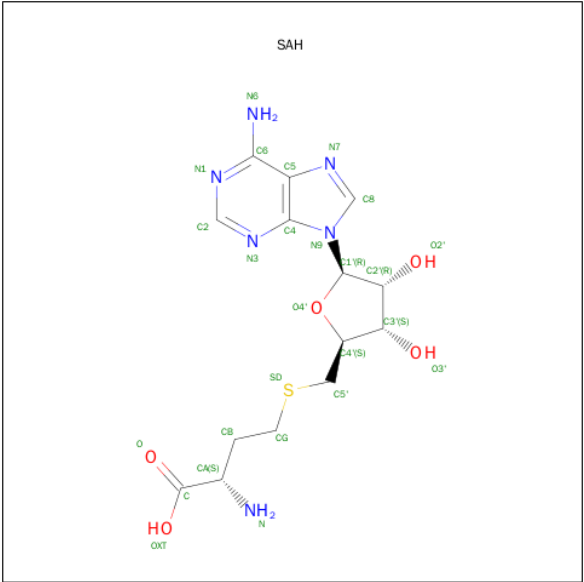
- Molecule 2 is a DNA chain called 5'-D(P\*GP\*TP\*CP\*AP\*GP\*(3DR)P\*GP\*CP\*AP\*TP\*GP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	12	Total	C	N	O	P	0	0	0
			241	113	45	71	12			

- Molecule 3 is a protein called CYTOSINE-SPECIFIC METHYLTRANSFERASE HHAI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	327	Total	C	N	O	S	0	0	0
			2606	1662	444	487	13			

- Molecule 4 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: C<sub>14</sub>H<sub>20</sub>N<sub>6</sub>O<sub>5</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	S	0	0
			26	14	6	5	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	213	Total	O	0	0
			213	213		
5	C	16	Total	O	0	0
			16	16		
5	D	13	Total	O	0	0
			13	13		

### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.


Note EDS was not executed.

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

Chain C: 




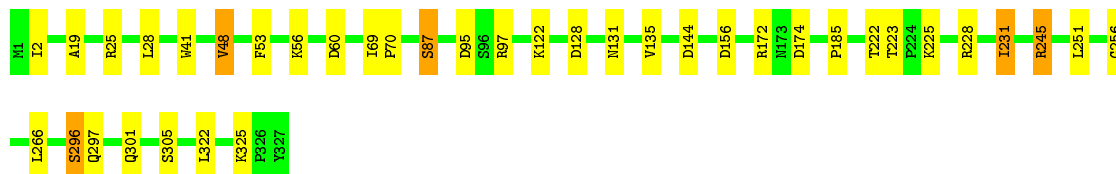
- Molecule 2: 5'-D(P\*GP\*TP\*CP\*AP\*GP\*(3DR)P\*GP\*CP\*AP\*TP\*GP\*G)-3'

Chain D: 



- Molecule 3: CYTOSINE-SPECIFIC METHYLTRANSFERASE HHAI

Chain A: 



## 4 Data and refinement statistics

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

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	99.86 Å 99.86 Å 325.20 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	6.00 – 2.39	Depositor
% Data completeness (in resolution range)	90.0 (6.00-2.39)	Depositor
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	X-PLOR 3.1	Depositor
R, $R_{free}$	0.186 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3358	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

## 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: SAH, 3DR

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	C	2.79	24/271 (8.9%)	3.18	47/415 (11.3%)
2	D	2.52	17/257 (6.6%)	3.41	49/393 (12.5%)
3	A	0.68	0/2661	0.75	2/3586 (0.1%)
All	All	1.25	41/3189 (1.3%)	1.57	98/4394 (2.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	C	0	1

The worst 5 of 41 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	405	DT	C5-C7	10.16	1.56	1.50
1	C	407	DC	O3'-P	7.96	1.70	1.61
1	C	404	DA	N3-C4	7.45	1.39	1.34
1	C	412	DA	O3'-P	7.31	1.70	1.61
1	C	407	DC	C5'-C4'	7.19	1.59	1.51

The worst 5 of 98 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	403	DC	O4'-C4'-C3'	-12.03	98.78	106.00
2	D	423	DT	C6-C5-C7	-11.85	115.79	122.90
2	D	432	DG	O4'-C4'-C3'	-11.62	99.03	106.00
1	C	402	DC	O4'-C4'-C3'	-10.67	99.60	106.00
2	D	424	DC	O4'-C1'-C2'	-10.59	97.43	105.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	407	DC	Sidechain

## 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	C	243	0	135	4	0
2	D	241	0	132	2	0
3	A	2606	0	2587	17	0
4	A	26	0	19	0	0
5	A	213	0	0	4	0
5	C	16	0	0	0	0
5	D	13	0	0	1	0
All	All	3358	0	2873	21	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 4.

The worst 5 of 21 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:A:296:SER:HB2	5:A:497:HOH:O	1.93	0.67
3:A:172:ARG:HD2	3:A:174:ASP:OD1	1.99	0.62
3:A:156:ASP:HB3	3:A:185:PRO:HB3	1.83	0.60
3:A:222:THR:HG22	3:A:245:ARG:HG3	1.88	0.56
3:A:322:LEU:O	3:A:325:LYS:HE2	2.07	0.54

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	
3	A	325/327 (99%)	308 (95%)	15 (5%)	2 (1%)	30	43

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	60	ASP
3	A	144	ASP

### 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	
3	A	283/283 (100%)	271 (96%)	12 (4%)	36	56

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	223	THR
3	A	231	ILE
3	A	266	LEU
3	A	87	SER
3	A	251	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	141	ASN

### 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	3DR	D	427	2	7,11,12	2.63	3 (42%)	8,14,17	2.42	5 (62%)

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	3DR	D	427	2	-	0/3/15/16	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	427	3DR	O4'-C1'	2.38	1.50	1.42
2	D	427	3DR	O3'-C3'	3.68	1.51	1.43
2	D	427	3DR	C2'-C3'	5.16	1.61	1.52

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	427	3DR	O4'-C1'-C2'	-2.27	101.97	106.64
2	D	427	3DR	O3'-C3'-C4'	2.16	118.75	110.05

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	427	3DR	O4'-C4'-C5'	2.27	114.49	109.53
2	D	427	3DR	O3'-C3'-C2'	3.36	119.40	111.71
2	D	427	3DR	C1'-O4'-C4'	4.12	114.81	108.58

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

1 ligand 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
4	SAH	A	328	-	20,28,28	1.41	4 (20%)	19,40,40	1.13	1 (5%)

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
4	SAH	A	328	-	-	0/7/31/31	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	328	SAH	C5'-C4'	2.17	1.58	1.52
4	A	328	SAH	C2'-C3'	2.25	1.59	1.53
4	A	328	SAH	C2-N3	2.96	1.37	1.32
4	A	328	SAH	C4-N3	2.97	1.40	1.35

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	328	SAH	CB-CG-SD	3.08	119.52	113.57

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 was not executed - this section will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

### 6.3 Carbohydrates ⓘ

EDS was not executed - this section will therefore be empty.

### 6.4 Ligands ⓘ

EDS was not executed - this section will therefore be empty.

### 6.5 Other polymers ⓘ

EDS was not executed - this section will therefore be empty.