



# Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 05:37 am BST

PDB ID : 1G55  
Title : Structure of human DNMT2, an enigmatic DNA methyltransferase homologue  
Authors : Dong, A.; Yoder, J.A.; Zhang, X.; Zhou, L.; Bestor, T.H.; Cheng, X.  
Deposited on : 2000-10-30  
Resolution : 1.80 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)  
Xtrriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
buster-report : 1.1.7 (2018)  
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) : 2.11

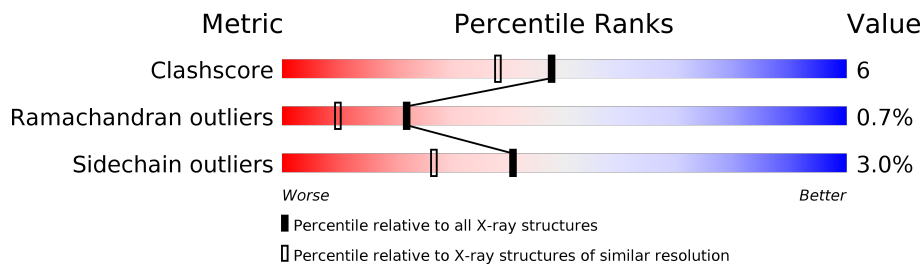
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)

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  $\leq 5\%$

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	343	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 2791 atoms, of which 6 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 DNA CYTOSINE METHYLTRANSFERASE DNMT2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	313	2490	1621	6	405	448	10	0	5	0

There are 47 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	SER	DELETION	UNP O14717
A	?	-	VAL	DELETION	UNP O14717
A	?	-	HIS	DELETION	UNP O14717
A	?	-	PRO	DELETION	UNP O14717
A	?	-	GLN	DELETION	UNP O14717
A	?	-	LYS	DELETION	UNP O14717
A	?	-	TYR	DELETION	UNP O14717
A	?	-	ALA	DELETION	UNP O14717
A	?	-	MET	DELETION	UNP O14717
A	?	-	ASP	DELETION	UNP O14717
A	?	-	VAL	DELETION	UNP O14717
A	?	-	GLU	DELETION	UNP O14717
A	?	-	ASN	DELETION	UNP O14717
A	?	-	LYS	DELETION	UNP O14717
A	?	-	ILE	DELETION	UNP O14717
A	?	-	GLN	DELETION	UNP O14717
A	?	-	GLU	DELETION	UNP O14717
A	?	-	LYS	DELETION	UNP O14717
A	?	-	ASN	DELETION	UNP O14717
A	?	-	VAL	DELETION	UNP O14717
A	?	-	GLU	DELETION	UNP O14717
A	?	-	PRO	DELETION	UNP O14717
A	?	-	ASN	DELETION	UNP O14717
A	?	-	ILE	DELETION	UNP O14717
A	?	-	SER	DELETION	UNP O14717
A	?	-	PHE	DELETION	UNP O14717
A	?	-	ASP	DELETION	UNP O14717

*Continued on next page...*

Continued from previous page...

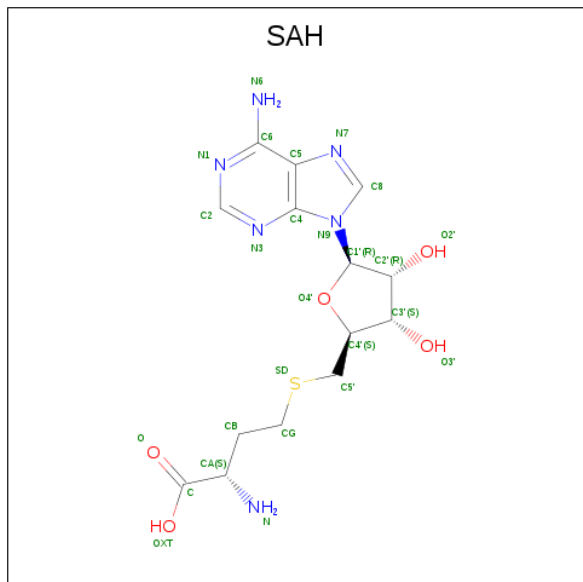
Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	GLY	DELETION	UNP O14717
A	?	-	SER	DELETION	UNP O14717
A	?	-	ILE	DELETION	UNP O14717
A	?	-	GLN	DELETION	UNP O14717
A	?	-	CYS	DELETION	UNP O14717
A	?	-	SER	DELETION	UNP O14717
A	?	-	GLY	DELETION	UNP O14717
A	?	-	LYS	DELETION	UNP O14717
A	?	-	ASP	DELETION	UNP O14717
A	?	-	ALA	DELETION	UNP O14717
A	?	-	ILE	DELETION	UNP O14717
A	?	-	LEU	DELETION	UNP O14717
A	?	-	PHE	DELETION	UNP O14717
A	?	-	LYS	DELETION	UNP O14717
A	?	-	LEU	DELETION	UNP O14717
A	?	-	GLU	DELETION	UNP O14717
A	?	-	THR	DELETION	UNP O14717
A	?	-	ALA	DELETION	UNP O14717
A	?	-	GLU	DELETION	UNP O14717
A	?	-	GLU	DELETION	UNP O14717

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



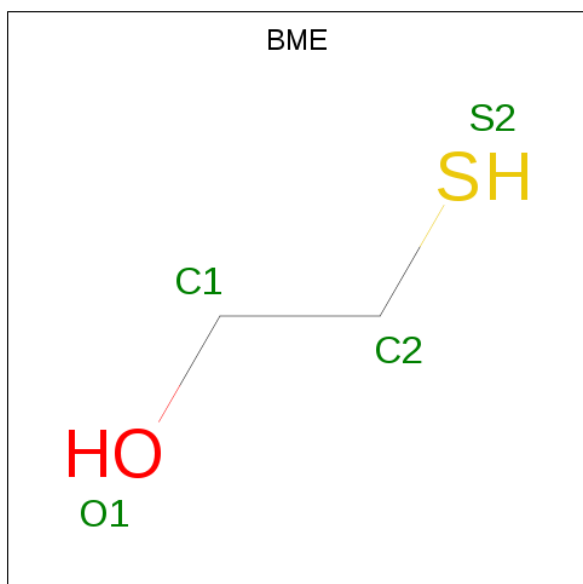
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 3 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula:  $C_{14}H_{20}N_6O_5S$ ).



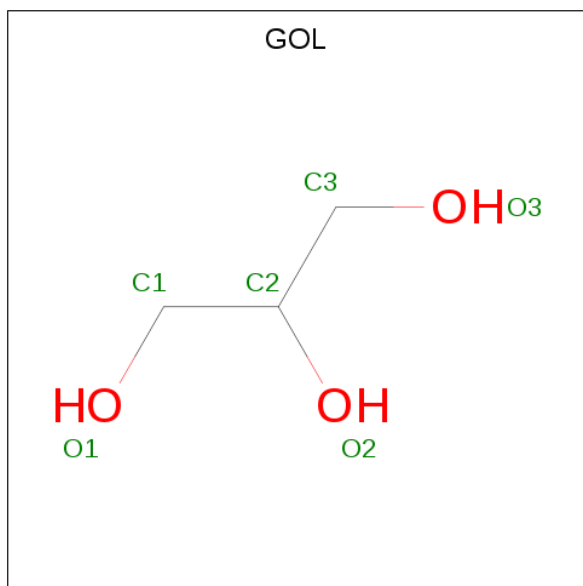
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
3	A	1	26	14	6	5	1	0	0

- Molecule 4 is BETA-MERCAPTOETHANOL (three-letter code: BME) (formula:  $C_2H_6OS$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	O	S		
4	A	1	8	4	2	2	0	1

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	C O	0	0
			6	3 3		

- Molecule 6 is water.

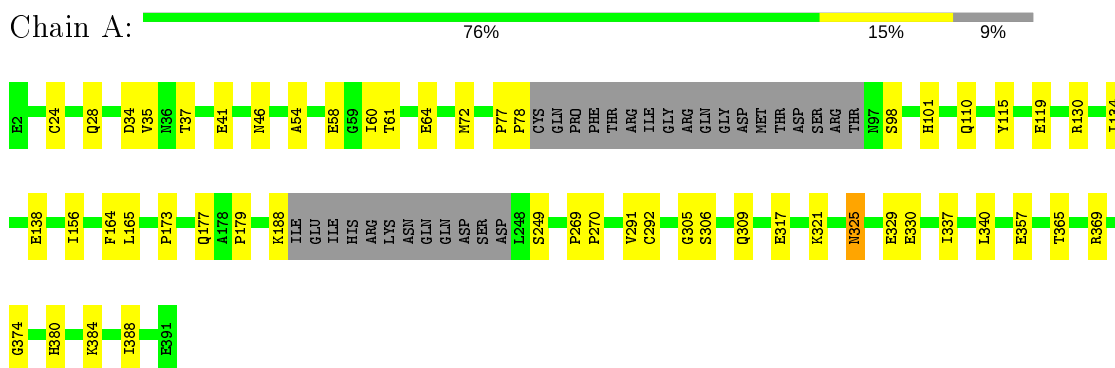
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	256	Total	O	0	0
			256	256		

### 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 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: DNA CYTOSINE METHYLTRANSFERASE DNMT2



## 4 Data and refinement statistics

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

Property	Value	Source
Space group	I 41	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	116.50Å 116.50Å 69.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 1.80	Depositor
% Data completeness (in resolution range)	(Not available) (20.00-1.80)	Depositor
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	X-PLOR 3.851	Depositor
R, $R_{free}$	0.210 , 0.250	Depositor
Estimated twinning fraction	No twinning to report.	Xtrriage
Total number of atoms	2791	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	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: GOL, CSO, SAH, SO4, BME

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	0.49	0/2546	0.68	0/3453

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	2484	6	2525	31	1
2	A	5	0	0	0	0
3	A	26	0	19	1	0
4	A	8	0	10	0	0
5	A	6	0	8	0	0
6	A	256	0	0	6	1
All	All	2785	6	2562	31	1

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.

All (31) 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:177:GLN:HE21	1:A:179:PRO:HG3	1.52	0.75
1:A:291[A]:VAL:HG23	1:A:305:GLY:HA2	1.70	0.74
1:A:60:ILE:O	1:A:101:HIS:HE1	1.71	0.72
1:A:374:GLY:HA2	6:A:476:HOH:O	1.93	0.69
1:A:384:LYS:O	1:A:388:ILE:HG12	1.97	0.63
1:A:173:PRO:HB3	1:A:177:GLN:OE1	1.99	0.63
1:A:337[A]:ILE:HD13	6:A:447:HOH:O	2.00	0.61
1:A:37:THR:O	1:A:41:GLU:HG3	2.06	0.56
1:A:119:GLU:HG3	1:A:164:PHE:CE2	2.42	0.54
1:A:60:ILE:O	1:A:101:HIS:CE1	2.58	0.53
1:A:58:GLU:HG3	1:A:98:SER:HA	1.91	0.52
1:A:61:THR:OG1	1:A:64:GLU:HG3	2.11	0.51
1:A:58:GLU:OE1	3:A:392:SAH:N6	2.31	0.51
1:A:325:ASN:CA	6:A:545:HOH:O	2.59	0.49
1:A:177:GLN:OE1	6:A:444:HOH:O	2.20	0.49
1:A:177:GLN:NE2	1:A:179:PRO:HG3	2.23	0.49
1:A:291[B]:VAL:HG11	6:A:535:HOH:O	2.14	0.48
1:A:34:ASP:O	1:A:54:ALA:HA	2.14	0.47
1:A:291[B]:VAL:HG12	1:A:292:CYS:N	2.30	0.47
1:A:329:GLU:HG3	1:A:330:GLU:N	2.29	0.46
1:A:101:HIS:HD2	6:A:560:HOH:O	1.99	0.46
1:A:309:GLN:HA	1:A:340:LEU:HD23	2.00	0.44
1:A:58:GLU:HG3	1:A:98:SER:CA	2.47	0.44
1:A:317:GLU:O	1:A:321:LYS:HG3	2.18	0.43
1:A:134:ILE:O	1:A:138:GLU:HG3	2.18	0.43
1:A:72:MET:HA	1:A:115:TYR:O	2.19	0.42
1:A:365:THR:O	1:A:369:ARG:HG3	2.19	0.42
1:A:77:PRO:HA	1:A:78:PRO:HD3	1.90	0.41
1:A:156:ILE:HD11	1:A:380:HIS:CD2	2.54	0.41
1:A:269:PRO:HA	1:A:270:PRO:HD3	1.93	0.41
1:A:291[A]:VAL:HG22	1:A:306:SER:N	2.36	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:46:ASN:OD1	6:A:637:HOH:O[6_665]	2.04	0.16

## 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	310/343 (90%)	301 (97%)	7 (2%)	2 (1%)	25 12

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	249	SER
1	A	325	ASN

### 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	276/312 (88%)	267 (97%)	9 (3%)	38 23

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	24[A]	CYS
1	A	24[B]	CYS
1	A	28	GLN
1	A	35	VAL
1	A	110	GLN
1	A	130	ARG
1	A	165	LEU
1	A	188	LYS
1	A	357	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	97	ASN
1	A	101	HIS
1	A	110	GLN
1	A	139	ASN
1	A	158	ASN
1	A	177	GLN
1	A	328	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 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 (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	CSO	A	140	1	3,6,7	0.90	0	0,6,8	0.00	-
1	CSO	A	287	1	3,6,7	0.55	0	0,6,8	0.00	-

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	CSO	A	140	1	-	0/1/5/7	-
1	CSO	A	287	1	-	0/1/5/7	-

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.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

5 ligands 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
4	BME	A	395[A]	1	3,3,3	0.27	0	1,2,2	0.17	0
4	BME	A	395[B]	1	3,3,3	0.27	0	1,2,2	0.33	0
2	SO4	A	399	-	4,4,4	0.18	0	6,6,6	0.18	0
3	SAH	A	392	-	21,28,28	0.58	0	20,40,40	1.22	3 (15%)
5	GOL	A	397	-	5,5,5	0.30	0	5,5,5	0.25	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
4	BME	A	395[A]	1	-	0/1/1/1	-
4	BME	A	395[B]	1	-	0/1/1/1	-
3	SAH	A	392	-	-	0/7/31/31	0/3/3/3
5	GOL	A	397	-	-	0/4/4/4	-

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
3	A	392	SAH	C4'-C5'-SD	-3.76	100.27	113.78
3	A	392	SAH	C5-C6-N6	2.24	123.76	120.35
3	A	392	SAH	CB-CG-SD	-2.01	108.80	113.31

There are no chirality outliers.

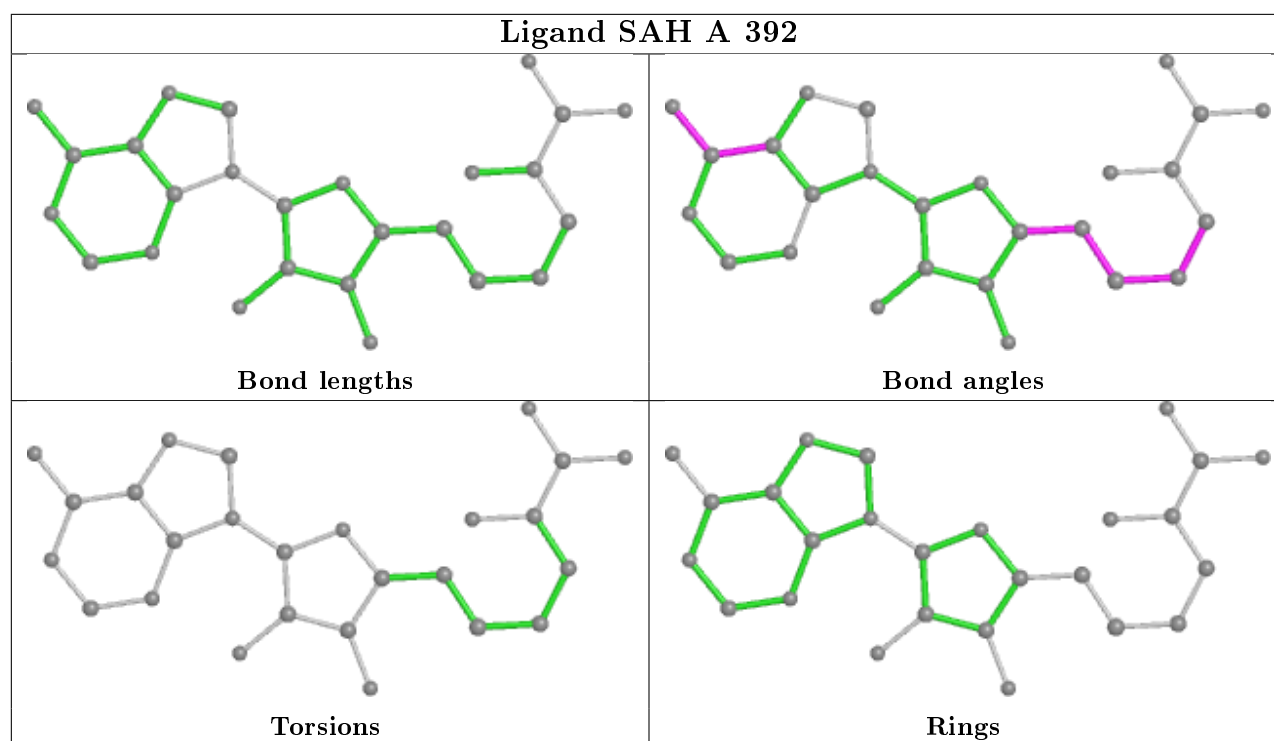
There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	392	SAH	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



## 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.