wwPDB X-ray Structure Validation Summary Report (i)

Aug 16, 2023 – 03:34 AM EDT

PDB ID	:	1YF3
Title	:	T4Dam in Complex with AdoHcy and 13-mer Oligonucleotide Making Non- and Semi-specific (1/4) Contact
Authors	:	Horton, J.R.; Liebert, K.; Hattman, S.; Jeltsch, A.; Cheng, X.
Deposited on	:	2004-12-30
Resolution	:	2.29 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 (1)) were used in the production of this report:

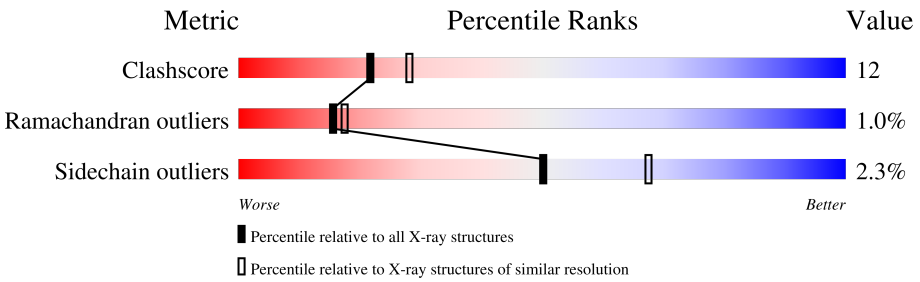
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.29 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	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)

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	C	13	69% 31%
2	D	13	69% 31%
3	A	259	71% 28% .
3	B	259	73% 23% . .

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	GOL	A	498	-	X	-	-
5	GOL	B	497	-	X	-	-

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 4812 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(*AP*CP*CP*AP*TP*GP*AP*TP*CP*TP*GP*AP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	13	243	117	40	74	12	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	13	266	128	49	77	12	0	0	0

- Molecule 3 is a protein called DNA adenine methylase.

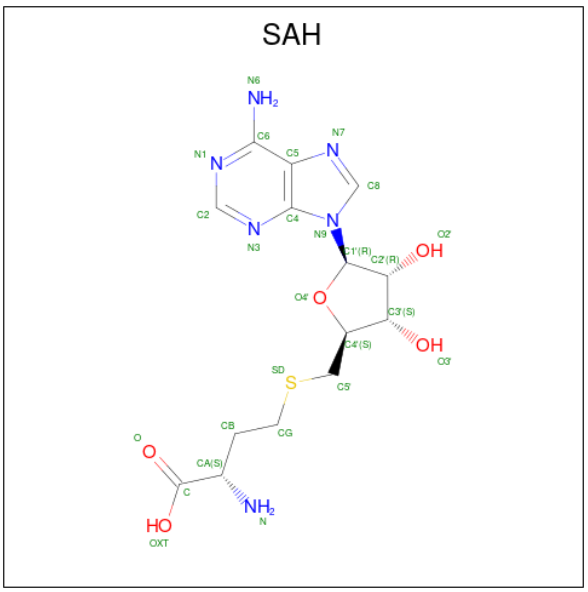
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	259	2050	1327	348	369	6	0	1	0
3	B	249	1985	1285	338	357	5	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	119	TYR	ASP	see remark 999	UNP P04392
A	139	ARG	GLN	see remark 999	UNP P04392
A	140	PHE	TYR	see remark 999	UNP P04392
A	209	LEU	GLN	see remark 999	UNP P04392
B	119	TYR	ASP	see remark 999	UNP P04392
B	139	ARG	GLN	see remark 999	UNP P04392
B	140	PHE	TYR	see remark 999	UNP P04392
B	209	LEU	GLN	see remark 999	UNP P04392

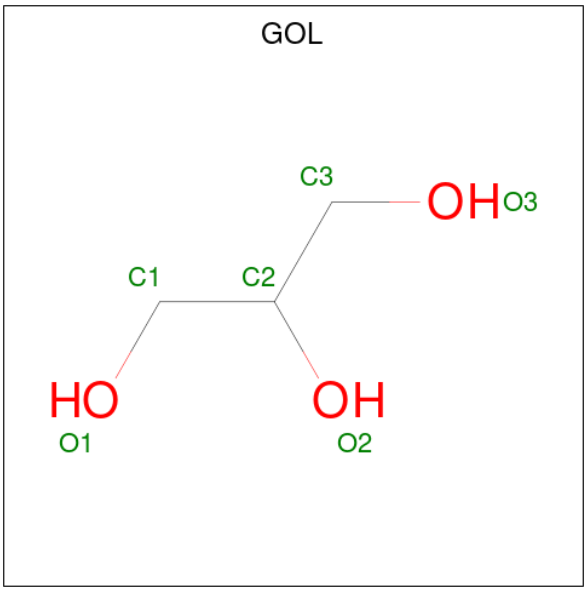
- Molecule 4 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula:

$C_{14}H_{20}N_6O_5S$).



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

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



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

Continued on next page...

Continued from previous page...

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

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	C	16	16	16	0	0
			Total	O		
6	D	18	18	18	0	0
			Total	O		
6	A	59	59	59	0	0
			Total	O		
6	B	111	111	111	0	0

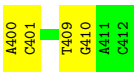
3 Residue-property plots (i)

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.

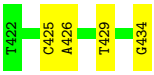
• Molecule 1: 5'-D(*AP*CP*CP*AP*TP*GP*AP*TP*CP*TP*GP*AP*C)-3'

Chain C:



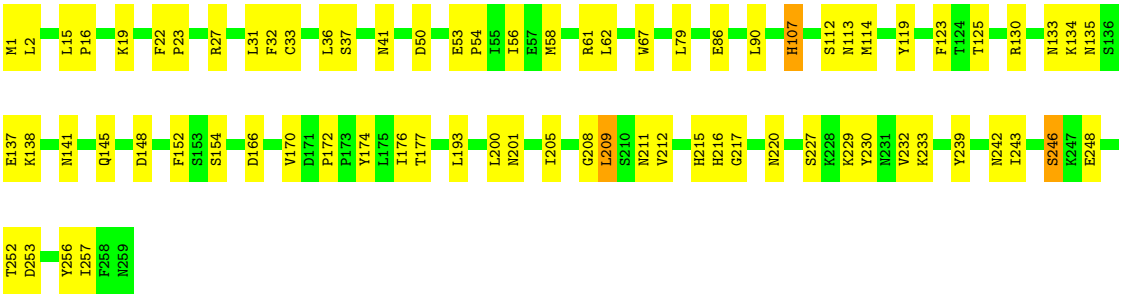
• Molecule 2: 5'-D(*TP*GP*TP*CP*AP*GP*AP*TP*CP*AP*TP*GP*G)-3'

Chain D:



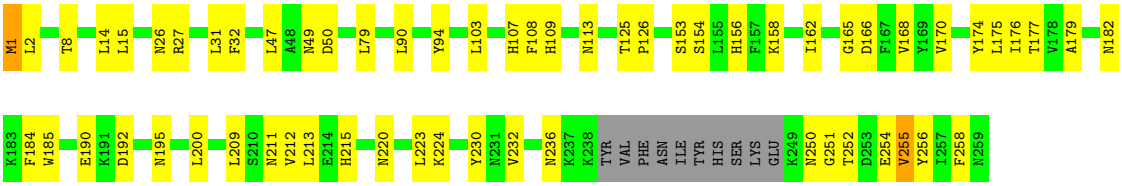
• Molecule 3: DNA adenine methylase

Chain A:



• Molecule 3: DNA adenine methylase

Chain B:



4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	38.90Å 125.80Å 73.20Å 90.00° 104.70° 90.00°	Depositor
Resolution (Å)	28.00 – 2.29	Depositor
% Data completeness (in resolution range)	93.8 (28.00-2.29)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
Refinement program	CNS	Depositor
R, R_{free}	0.214 , 0.270	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4812	wwPDB-VP
Average B, all atoms (Å^2)	46.0	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, GOL

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	0.30	0/271	0.71	0/415
2	D	0.31	0/298	0.70	0/459
3	A	0.39	0/2102	0.60	0/2839
3	B	0.39	0/2033	0.61	0/2744
All	All	0.38	0/4704	0.62	0/6457

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	C	243	0	138	3	0
2	D	266	0	149	3	0
3	A	2050	0	1965	52	0
3	B	1985	0	1927	47	0
4	A	26	0	19	1	0
4	B	26	0	19	0	0
5	A	6	0	4	0	0
5	B	6	0	4	1	0
6	A	59	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	111	0	0	1	0
6	C	16	0	0	0	0
6	D	18	0	0	1	0
All	All	4812	0	4225	105	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 12.

The worst 5 of 105 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:400:DA:H4'	1:C:401:DC:H5'	1.31	1.13
3:A:15:LEU:HD21	3:A:41:ASN:HB2	1.48	0.94
3:A:170:VAL:HG12	3:A:172:PRO:HD3	1.48	0.94
3:A:174:TYR:HB2	3:A:177:THR:HG22	1.62	0.81
3:B:156:HIS:ND1	3:B:158:LYS:HG2	1.97	0.80

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	257/259 (99%)	238 (93%)	16 (6%)	3 (1%)	13 14
3	B	245/259 (95%)	230 (94%)	13 (5%)	2 (1%)	19 23
All	All	502/518 (97%)	468 (93%)	29 (6%)	5 (1%)	15 17

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	248	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	B	251	GLY
3	A	216	HIS
3	A	134	LYS
3	B	236	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	
3	A	216/242 (89%)	211 (98%)	5 (2%)	50	67
3	B	212/242 (88%)	207 (98%)	5 (2%)	49	66
All	All	428/484 (88%)	418 (98%)	10 (2%)	50	67

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	107	HIS
3	B	185	TRP
3	B	255	VAL
3	A	209	LEU
3	A	246	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	113	ASN
3	B	141	ASN
3	B	195	ASN
3	B	182	ASN
3	B	26	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

4 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	$\#\|Z\| > 2$
5	GOL	B	497	-	5,5,5	4.56	5 (100%)	5,5,5	5.79	3 (60%)
5	GOL	A	498	-	5,5,5	4.51	5 (100%)	5,5,5	5.75	3 (60%)
4	SAH	A	400	-	24,28,28	1.58	4 (16%)	25,40,40	2.27	9 (36%)
4	SAH	B	401	-	24,28,28	1.54	3 (12%)	25,40,40	2.23	9 (36%)

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
5	GOL	B	497	-	-	2/4/4/4	-
5	GOL	A	498	-	-	2/4/4/4	-
4	SAH	A	400	-	-	0/11/31/31	0/3/3/3
4	SAH	B	401	-	-	0/11/31/31	0/3/3/3

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	497	GOL	C3-C2	-7.54	1.20	1.51

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	498	GOL	C3-C2	-7.42	1.21	1.51
5	A	498	GOL	O1-C1	4.50	1.61	1.42
5	B	497	GOL	O1-C1	4.49	1.61	1.42
4	A	400	SAH	O4'-C1'	4.03	1.46	1.41

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	B	497	GOL	O3-C3-C2	10.52	160.65	110.20
5	A	498	GOL	O3-C3-C2	10.46	160.35	110.20
4	A	400	SAH	O4'-C1'-C2'	-7.14	96.50	106.93
4	B	401	SAH	O4'-C1'-C2'	-7.11	96.53	106.93
5	A	498	GOL	O2-C2-C3	6.79	139.05	109.12

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	498	GOL	C1-C2-C3-O3
5	B	497	GOL	C1-C2-C3-O3
5	A	498	GOL	O1-C1-C2-C3
5	B	497	GOL	O1-C1-C2-C3

There are no ring outliers.

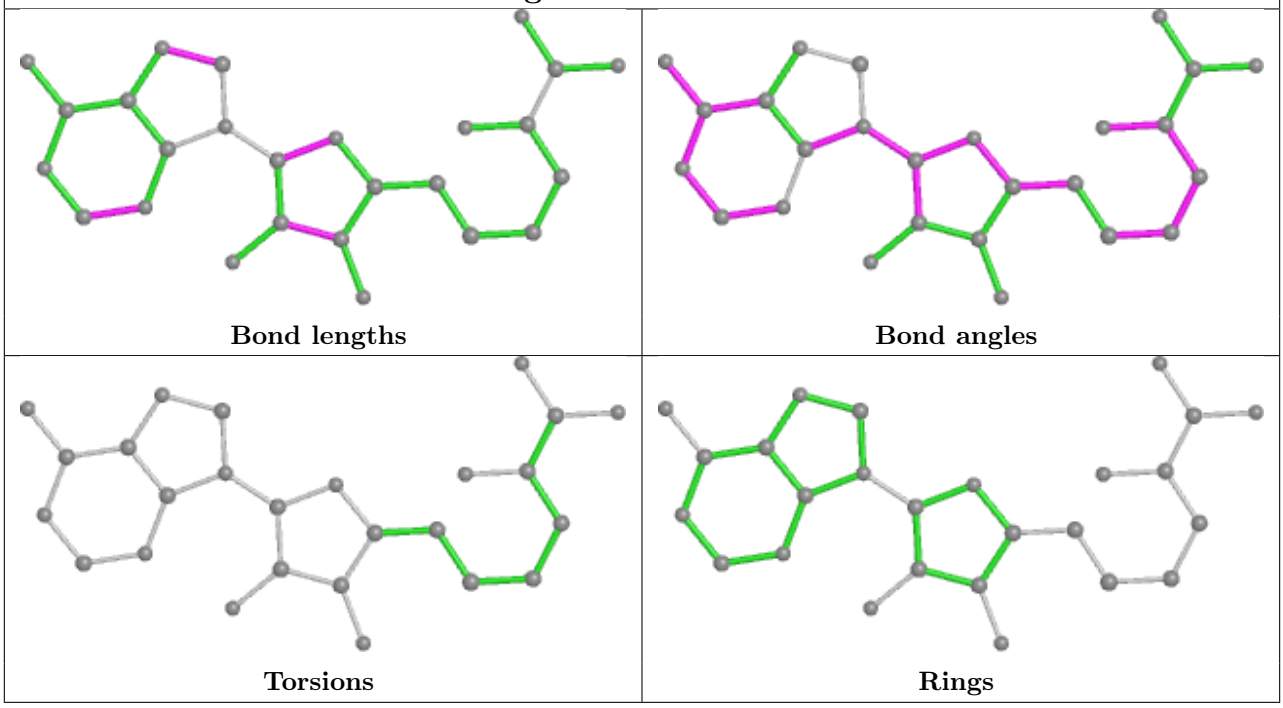
2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	497	GOL	1	0
4	A	400	SAH	1	0

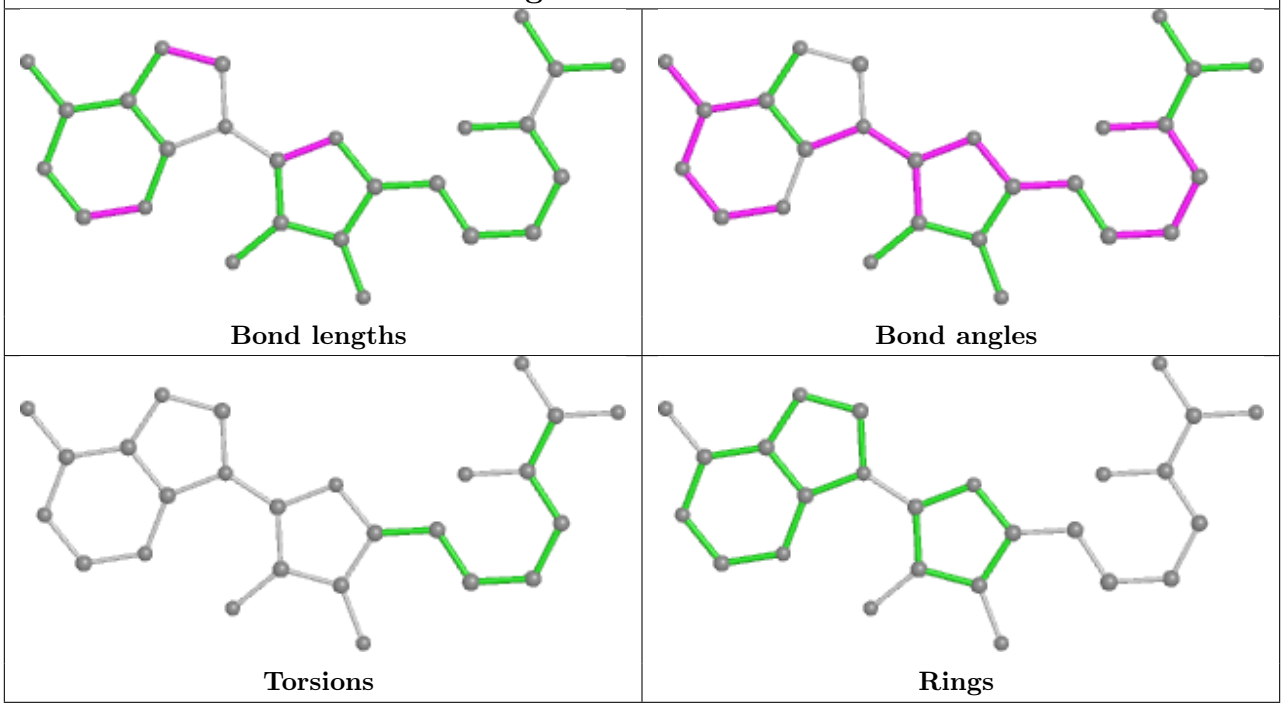
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 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.

Ligand SAH A 400



Ligand SAH B 401



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)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.