wwPDB/EMDataBank EM Map/Model Validation Summary Report (i)

Oct 29, 2019 – 03:42 PM EDT

PDB ID	:	6JM9
EMDB ID:	:	EMD-9843
Title	:	cryo-EM structure of DOT1L bound to unmodified nucleosome
Authors	:	Jang, S.; Song, J.J.
Deposited on	:	2019-03-07
Resolution	:	7.30 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/EMValidationReportHelp
with specific help available everywhere you see the (i) symbol.

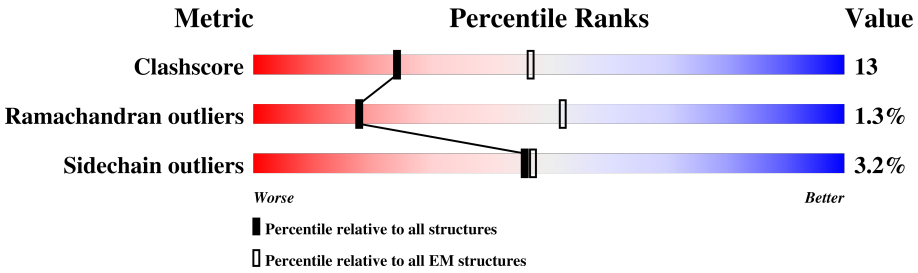
MolProbity	:	4.02b-467
Mogul	:	1.8.0 (224370), CSD as540be (2019)
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.4

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 7.30 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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for >=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 <=5%

Mol	Chain	Length	Quality of chain
1	I	123	57% 43%
2	J	123	53% 47%
3	A	98	80% 18% •
3	E	98	74% 23% •
4	B	87	72% 20% •• 6%
4	F	87	79% 18% •
5	C	107	79% 17% •
5	G	107	81% 17% ••
6	D	94	80% 18% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	H	94	
7	X	328	

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
8	SAM	X	500	-	-	X	-

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 13827 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 strand I.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
1	I	123	2517	1203	453	738	123	0	0

- Molecule 2 is a DNA chain called DNA strand J.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
2	J	123	2526	1206	459	738	123	0	0

- Molecule 3 is a protein called Histone H3.2.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	A	98	807	508	156	140	3	0	0
3	E	98	Total	C	N	O	S	0	0
			807	508	156	140	3		

- Molecule 4 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	B	82	653	412	127	113	1	0	0
4	F	87	Total	C	N	O	S	0	0
			703	442	142	118	1		

- Molecule 5 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				AltConf	Trace
			Total	C	N	O		
5	C	107	825	520	161	144	0	0
5	G	106	Total	C	N	O	0	0
			818	516	160	142		

- Molecule 6 is a protein called Histone H2B 1.1.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	D	94	Total	C	N	O	S	0	0
			736	463	132	139	2		
6	H	94	Total	C	N	O	S	0	0
			736	463	132	139	2		

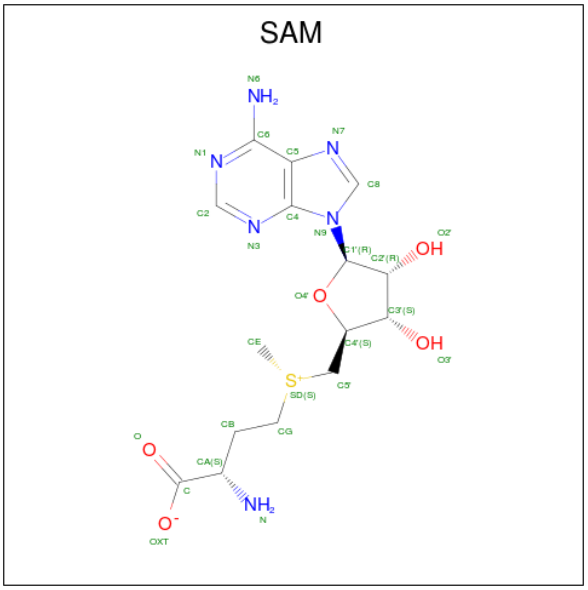
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	29	THR	-	expression tag	UNP P02281
H	29	THR	-	expression tag	UNP P02281

- Molecule 7 is a protein called Histone-lysine N-methyltransferase, H3 lysine-79 specific.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	X	328	Total	C	N	O	S	0	0
			2672	1706	455	499	12		

- Molecule 8 is S-ADENOSYLMETHIONINE (three-letter code: SAM) (formula: $C_{15}H_{22}N_6O_5S$).



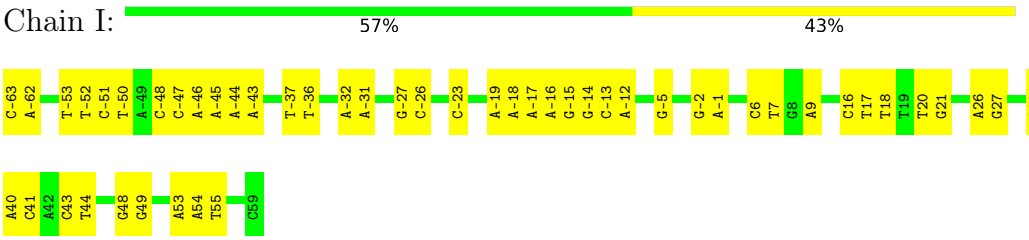
Mol	Chain	Residues	Atoms					AltConf
8	X	1	Total	C	N	O	S	0
			27	15	6	5	1	

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.

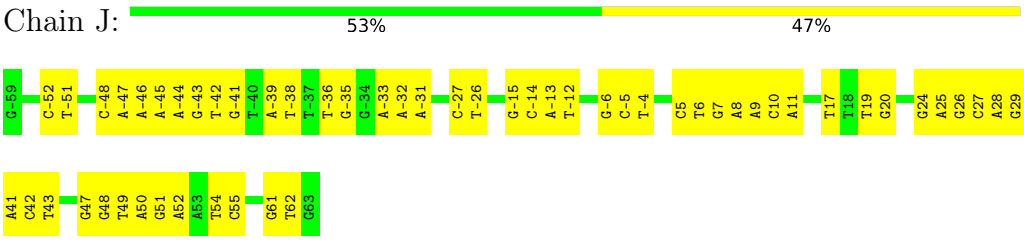
- Molecule 1: DNA strand I

Chain I:



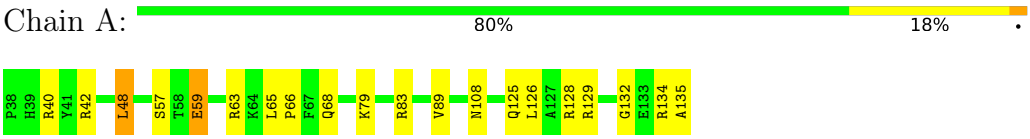
- Molecule 2: DNA strand J

Chain J:



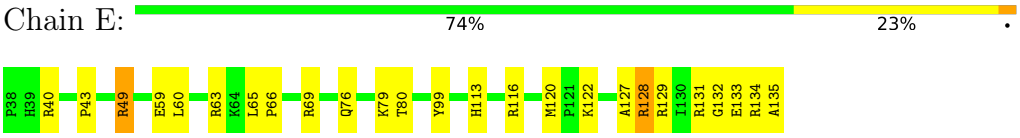
- Molecule 3: Histone H3.2

Chain A:



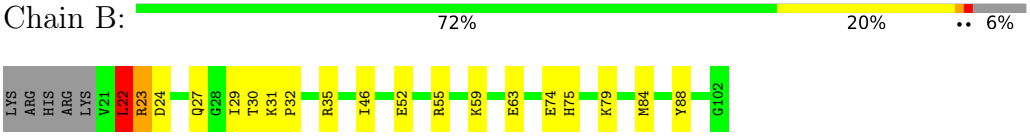
- Molecule 3: Histone H3.2

Chain E:



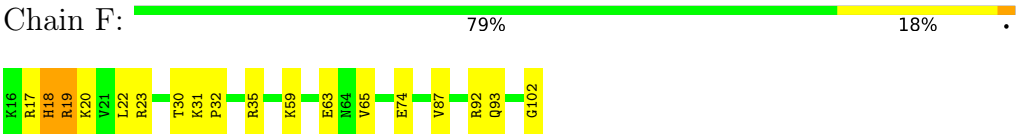
- Molecule 4: Histone H4

Chain B:



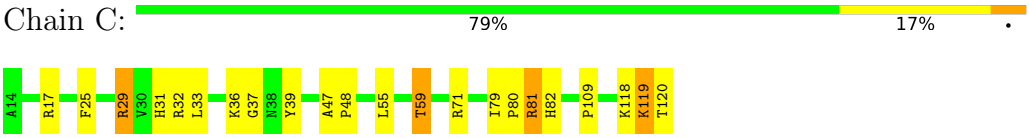
- Molecule 4: Histone H4

Chain F:



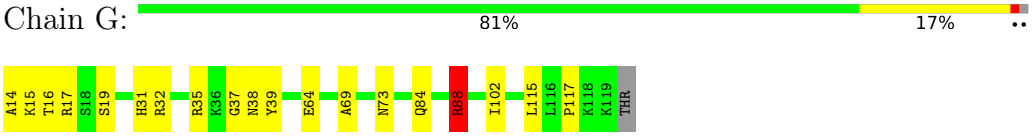
• Molecule 5: Histone H2A

Chain C:



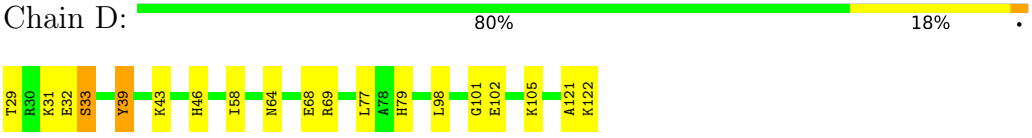
• Molecule 5: Histone H2A

Chain G:



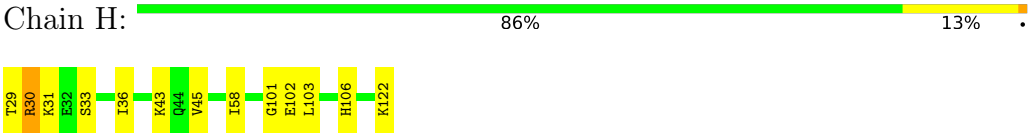
• Molecule 6: Histone H2B 1.1

Chain D:



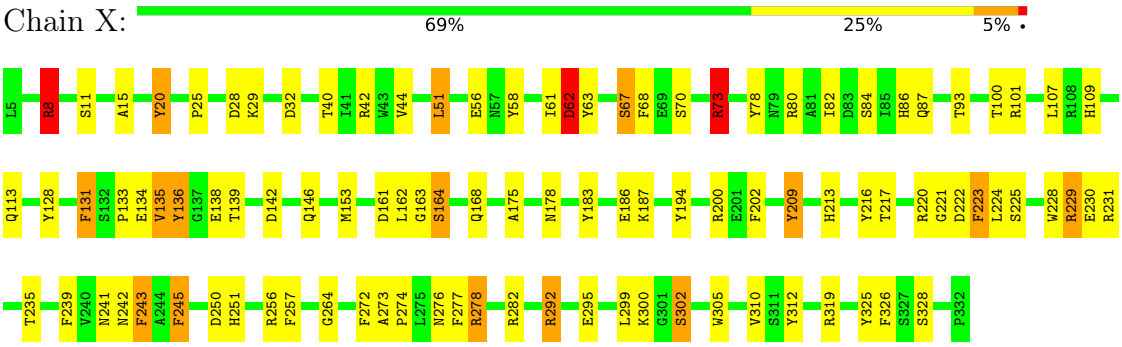
• Molecule 6: Histone H2B 1.1

Chain H:



• Molecule 7: Histone-lysine N-methyltransferase, H3 lysine-79 specific

Chain X:



4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	21229	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	FEI TITAN	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	37.28	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

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: SAM

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	#$\|Z\|$ >2	RMSZ	#$\|Z\|$ >2
1	I	0.36	0/2822	0.71	0/4352
2	J	0.40	0/2834	0.71	0/4373
3	A	0.52	0/819	0.67	0/1097
3	E	0.69	0/819	0.80	1/1097 (0.1%)
4	B	0.56	0/660	0.72	1/883 (0.1%)
4	F	0.69	0/711	0.83	1/948 (0.1%)
5	C	0.67	0/835	0.83	2/1127 (0.2%)
5	G	0.51	0/828	0.69	2/1117 (0.2%)
6	D	0.66	0/747	0.73	0/1004
6	H	0.55	0/747	0.67	0/1004
7	X	1.77	30/2742 (1.1%)	2.03	68/3718 (1.8%)
All	All	0.90	30/14564 (0.2%)	1.08	75/20720 (0.4%)

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
2	J	0	2
6	D	0	1
7	X	0	9
All	All	0	12

The worst 5 of 30 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	X	216	TYR	CE2-CZ	8.25	1.49	1.38
7	X	20	TYR	CE2-CZ	7.45	1.48	1.38
7	X	200	ARG	CD-NE	7.23	1.58	1.46
7	X	136	TYR	CG-CD2	7.14	1.48	1.39

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	X	134	GLU	CD-OE2	6.58	1.32	1.25

The worst 5 of 75 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	X	278	ARG	NE-CZ-NH2	-17.51	111.55	120.30
7	X	101	ARG	NE-CZ-NH1	16.73	128.67	120.30
7	X	216	TYR	CB-CG-CD1	-15.43	111.74	121.00
7	X	278	ARG	NE-CZ-NH1	15.23	127.92	120.30
7	X	216	TYR	CB-CG-CD2	13.43	129.06	121.00

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
6	D	39	TYR	Sidechain
2	J	-12	DT	Sidechain
2	J	-6	DG	Sidechain
7	X	20	TYR	Sidechain
7	X	8	ARG	Sidechain

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	I	2517	0	1390	55	0
2	J	2526	0	1390	64	0
3	A	807	0	844	37	0
3	E	807	0	844	35	0
4	B	653	0	696	18	0
4	F	703	0	755	20	0
5	C	825	0	884	24	0
5	G	818	0	877	23	0
6	D	736	0	760	23	0
6	H	736	0	760	12	0
7	X	2672	0	2615	91	0
8	X	27	0	16	63	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	13827	0	11831	330	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 13.

The worst 5 of 330 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)
7:X:245:PHE:CZ	8:X:500:SAM:H5'1	1.16	1.66
7:X:245:PHE:HZ	8:X:500:SAM:C5'	1.03	1.63
7:X:223:PHE:CD1	8:X:500:SAM:N3	1.74	1.53
7:X:223:PHE:CD2	8:X:500:SAM:C5	1.97	1.45
7:X:241:ASN:ND2	8:X:500:SAM:HE1	1.03	1.33

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 PDB entries followed by that with respect to all EM entries.

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	96/98 (98%)	96 (100%)	0	0	100 100
3	E	96/98 (98%)	95 (99%)	1 (1%)	0	100 100
4	B	80/87 (92%)	78 (98%)	0	2 (2%)	6 39
4	F	85/87 (98%)	81 (95%)	1 (1%)	3 (4%)	4 32
5	C	105/107 (98%)	101 (96%)	4 (4%)	0	100 100
5	G	104/107 (97%)	102 (98%)	2 (2%)	0	100 100
6	D	92/94 (98%)	90 (98%)	1 (1%)	1 (1%)	16 58
6	H	92/94 (98%)	88 (96%)	2 (2%)	2 (2%)	7 42
7	X	326/328 (99%)	303 (93%)	17 (5%)	6 (2%)	9 47

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1076/1100 (98%)	1034 (96%)	28 (3%)	14 (1%)	18 54

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	F	19	ARG
7	X	67	SER
4	B	22	LEU
6	D	101	GLY
6	H	101	GLY

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

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	85/85 (100%)	83 (98%)	2 (2%)	52 75
3	E	85/85 (100%)	83 (98%)	2 (2%)	52 75
4	B	67/72 (93%)	66 (98%)	1 (2%)	67 84
4	F	72/72 (100%)	71 (99%)	1 (1%)	69 85
5	C	85/85 (100%)	80 (94%)	5 (6%)	21 52
5	G	84/85 (99%)	83 (99%)	1 (1%)	74 87
6	D	80/80 (100%)	79 (99%)	1 (1%)	71 86
6	H	80/80 (100%)	77 (96%)	3 (4%)	36 64
7	X	294/294 (100%)	280 (95%)	14 (5%)	28 58
All	All	932/938 (99%)	902 (97%)	30 (3%)	46 68

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
6	H	31	LYS
7	X	8	ARG
7	X	235	THR
6	H	103	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
7	X	51	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
6	H	79	HIS
6	H	106	HIS
7	X	178	ASN
5	G	31	HIS
7	X	241	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 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	$\#\|Z\| > 2$
8	SAM	X	500	-	21,29,29	0.99	2 (9%)	16,42,42	1.62	2 (12%)

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
8	SAM	X	500	-	-	2/8/33/33	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	X	500	SAM	C2-N3	2.21	1.35	1.32
8	X	500	SAM	C8-N7	-2.03	1.31	1.34

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
8	X	500	SAM	C4'-O4'-C1'	-3.96	105.70	109.83
8	X	500	SAM	O2'-C2'-C3'	3.10	121.84	111.80

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
8	X	500	SAM	N-CA-CB-CG
8	X	500	SAM	C-CA-CB-CG

There are no ring outliers.

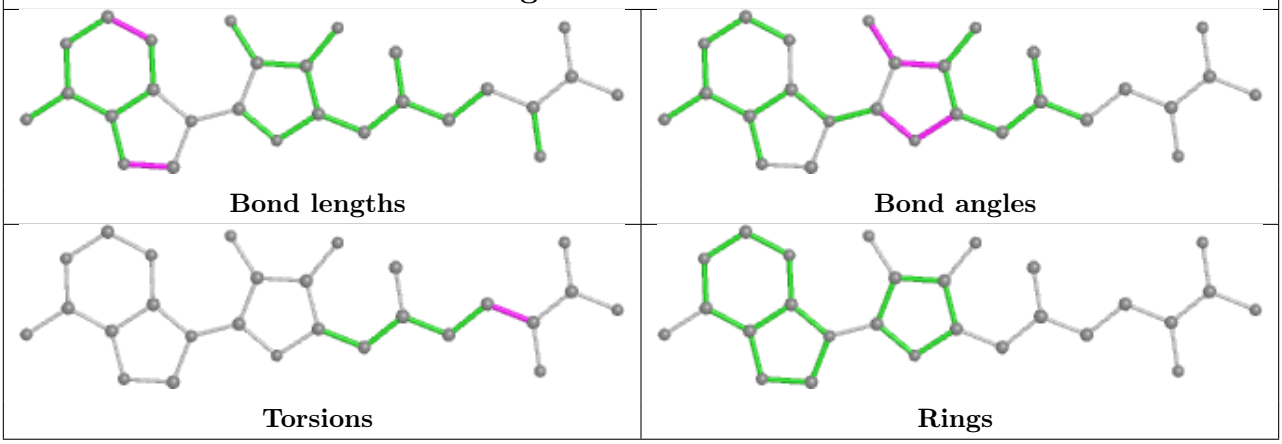
1 monomer is involved in 63 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	X	500	SAM	63	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 SAM X 500



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.