



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 12, 2024 – 05:33 PM EDT

PDB ID : 7UBU  
Title : Crystal structure of ZMET2 in complex with hemimethylated CAG DNA and a histone H3Kc9me2 peptide  
Authors : Fang, J.; Song, J.  
Deposited on : 2022-03-15  
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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

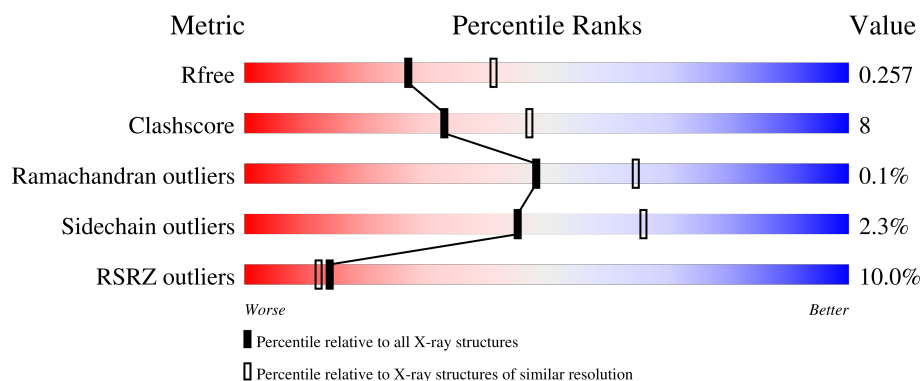
# 1 Overall quality at a glance

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))
$R_{free}$	164625	4642 (2.40-2.40)
Clashscore	180529	5218 (2.40-2.40)
Ramachandran outliers	177936	5158 (2.40-2.40)
Sidechain outliers	177891	5159 (2.40-2.40)
RSRZ outliers	164620	4642 (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 of 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	783	<div> <div>8%</div> <div>73%</div> <div>17%</div> <div>10%</div> </div>
2	P	33	<div> <div>27%</div> <div>52%</div> <div>9%</div> <div>39%</div> </div>
2	Q	33	<div> <div>21%</div> <div>24%</div> <div>9%</div> <div>64%</div> </div>
3	B	18	<div> <div>67%</div> <div>28%</div> <div>6%</div> </div>
4	C	18	<div> <div>56%</div> <div>33%</div> <div>11%</div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 6690 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 protein called DNA (cytosine-5)-methyltransferase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	708	Total	C	N	O	S	0	1	0
			5571	3554	944	1039	34			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	20	Total	C	N	O	S	0	0	0
			148	90	31	26	1			
2	Q	12	Total	C	N	O	S	0	0	0
			92	53	20	18	1			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	0	SER	-	expression tag	UNP P69246
Q	0	SER	-	expression tag	UNP P69246

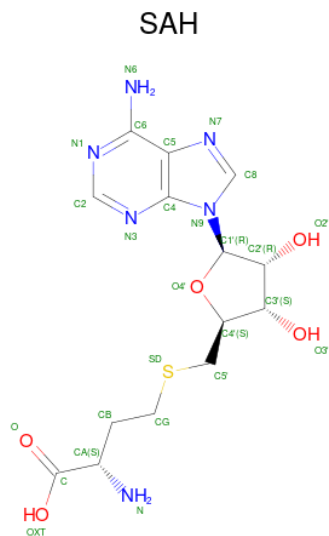
- Molecule 3 is a DNA chain called 5MC SSDNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	18	Total	C	N	O	P	0	0	0
			369	180	67	105	17			

- Molecule 4 is a DNA chain called C49 SSDNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	C	18	Total	C	F	N	O	P	0	0
			364	178	1	63	105	17		

- Molecule 5 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) (labeled as "Ligand of Interest" by depositor).

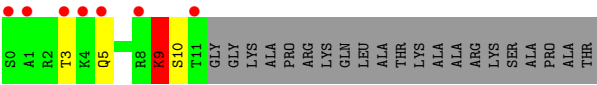


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total 26	C 14	N 6	O 5	S 1	0	0

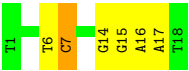
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	109	Total O 109 109	0	0
6	P	4	Total O 4 4	0	0
6	B	4	Total O 4 4	0	0
6	C	3	Total O 3 3	0	0

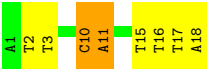




● Molecule 3: 5MC SSDNA



● Molecule 4: C49 SSDNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.57Å 121.39Å 160.45Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	96.81 – 2.39 96.81 – 2.39	Depositor EDS
% Data completeness (in resolution range)	98.9 (96.81-2.39) 98.9 (96.81-2.39)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.20 (at 2.40Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, $R_{free}$	0.211 , 0.256 0.212 , 0.257	Depositor DCC
$R_{free}$ test set	48772 reflections (3.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	62.9	Xtriage
Anisotropy	0.377	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 51.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6690	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.21% 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: SAH, C49, M2L, 5MC

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	0.33	0/5707	0.56	0/7742
2	P	0.30	0/135	0.59	0/176
2	Q	0.35	0/79	0.65	0/103
3	B	0.67	0/392	1.06	0/605
4	C	0.71	0/383	1.11	1/587 (0.2%)
All	All	0.39	0/6696	0.65	1/9213 (0.0%)

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	Q	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	C	11	DA	O4'-C4'-C3'	-5.54	102.28	104.50

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	Q	9	M2L	Mainchain



## 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	5571	0	5386	87	0
2	P	148	0	159	4	0
2	Q	92	0	96	4	0
3	B	369	0	204	5	0
4	C	364	0	204	5	0
5	A	26	0	19	2	0
6	A	109	0	0	2	0
6	B	4	0	0	0	0
6	C	3	0	0	0	0
6	P	4	0	0	0	0
All	All	6690	0	6068	94	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 8.

The worst 5 of 94 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:275:ASP:HB3	1:A:278:MET:HG3	1.46	0.98
1:A:487:GLN:HB3	2:Q:5:GLN:HE22	1.42	0.83
1:A:296:MET:HE2	6:A:1116:HOH:O	1.79	0.83
1:A:466:TRP:HE1	1:A:475:THR:HG22	1.54	0.70
1:A:596:CYS:HB3	1:A:623:PRO:HB3	1.76	0.68

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	
1	A	701/783 (90%)	670 (96%)	30 (4%)	1 (0%)	48	65
2	P	15/33 (46%)	15 (100%)	0	0	100	100
2	Q	9/33 (27%)	9 (100%)	0	0	100	100
All	All	725/849 (85%)	694 (96%)	30 (4%)	1 (0%)	48	65

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	888	SER

### 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 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	589/667 (88%)	576 (98%)	13 (2%)	47	67
2	P	13/22 (59%)	13 (100%)	0	100	100
2	Q	8/22 (36%)	7 (88%)	1 (12%)	3	5
All	All	610/711 (86%)	596 (98%)	14 (2%)	45	66

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	433	ASP
1	A	525	ARG
2	Q	10	SER
1	A	775	ASP
1	A	873	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	487	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	672	GLN
1	A	752	ASN
2	P	5	GLN
2	P	19	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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$
3	5MC	B	7	4,3	18,21,23	4.54	12 (66%)	24,30,35	1.40	3 (12%)
2	M2L	Q	9	2	9,10,11	1.51	1 (11%)	7,11,13	1.40	2 (28%)
2	M2L	P	9	2	9,10,11	1.59	1 (11%)	7,11,13	1.17	0
4	C49	C	10	4,1	17,22,24	4.89	10 (58%)	17,33,38	2.11	4 (23%)

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
3	5MC	B	7	4,3	-	2/7/21/26	0/2/2/2
2	M2L	Q	9	2	-	3/7/9/11	-
2	M2L	P	9	2	-	1/7/9/11	-
4	C49	C	10	4,1	-	7/7/40/46	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	10	C49	C2-N1	9.86	1.49	1.35
4	C	10	C49	C3'-C4'	-8.69	1.30	1.53
3	B	7	5MC	C2'-C3'	-8.25	1.31	1.52
4	C	10	C49	O4'-C4'	7.79	1.62	1.45
3	B	7	5MC	C6-C5	7.63	1.47	1.34

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	C	10	C49	F-C5-CM5	-5.57	98.85	106.49
4	C	10	C49	N3-C2-N1	4.68	121.35	116.65
4	C	10	C49	O2-C2-N1	-3.43	118.97	123.10
3	B	7	5MC	C2'-C3'-C4'	3.19	109.28	102.80
3	B	7	5MC	C5-C6-N1	-3.10	119.95	123.31

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	10	C49	O4'-C1'-N1-C6
3	B	7	5MC	C3'-C4'-C5'-O5'
4	C	10	C49	C3'-C4'-C5'-O5'
2	Q	9	M2L	SG-CD-CE-NZ
3	B	7	5MC	O4'-C4'-C5'-O5'

There are no ring outliers.

3 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	7	5MC	3	0
2	Q	9	M2L	1	0
4	C	10	C49	1	0

## 5.5 Carbohydrates [i](#)

There are no oligosaccharides 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	$\# Z  > 2$
5	SAH	A	1000	-	23,28,28	1.16	2 (8%)	22,40,40	1.80	3 (13%)

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	SAH	A	1000	-	-	3/11/31/31	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1000	SAH	C2-N3	3.74	1.37	1.32
5	A	1000	SAH	C2-N1	2.22	1.37	1.33

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1000	SAH	N3-C2-N1	-6.34	120.07	128.67
5	A	1000	SAH	OXT-C-O	-2.66	118.06	124.08
5	A	1000	SAH	C5'-SD-CG	-2.57	94.64	102.26

There are no chirality outliers.

All (3) torsion outliers are listed below:

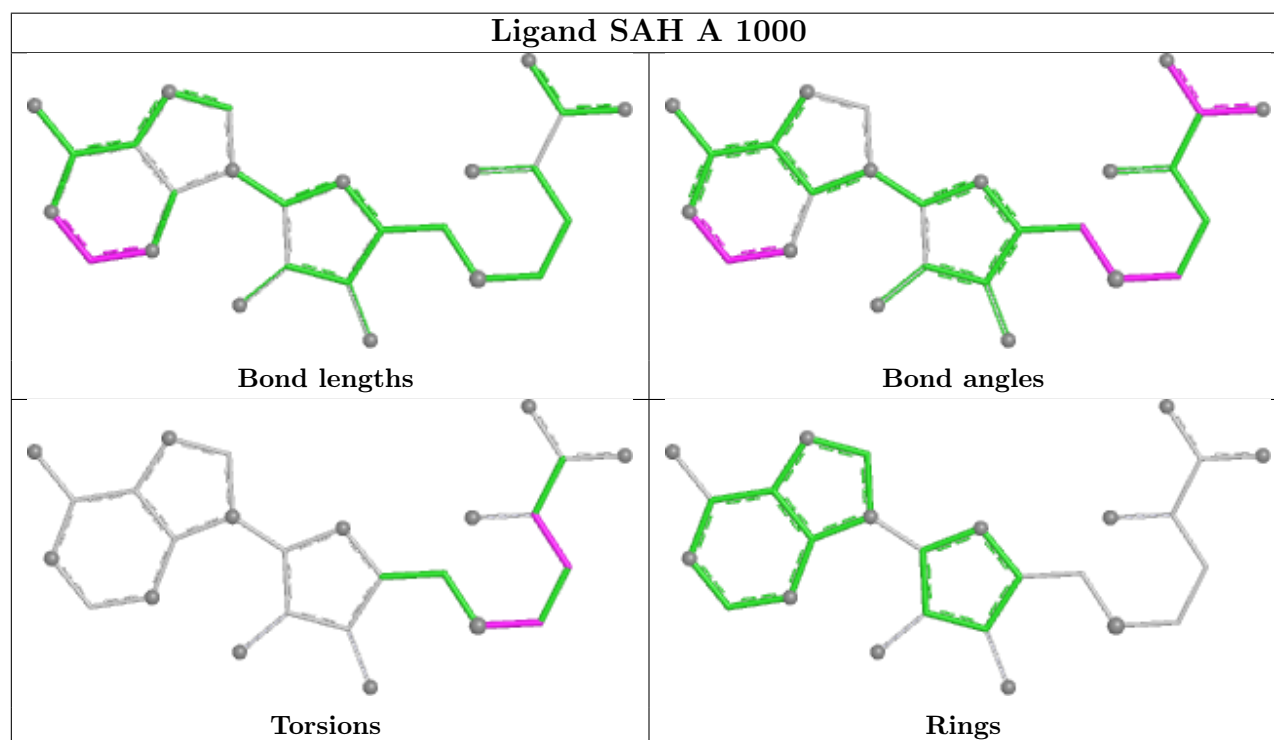
Mol	Chain	Res	Type	Atoms
5	A	1000	SAH	N-CA-CB-CG
5	A	1000	SAH	C-CA-CB-CG
5	A	1000	SAH	CB-CG-SD-C5'

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	1000	SAH	2	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 [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	708/783 (90%)	0.95	61 (8%) 18 16	34, 69, 109, 160	1 (0%)
2	P	19/33 (57%)	2.27	9 (47%) 0 0	75, 95, 131, 142	0
2	Q	11/33 (33%)	2.19	7 (63%) 0 0	81, 108, 127, 135	0
3	B	17/18 (94%)	0.39	0 100 100	77, 90, 136, 138	0
4	C	17/18 (94%)	0.39	0 100 100	73, 95, 131, 137	0
All	All	772/885 (87%)	0.98	77 (9%) 14 12	34, 71, 116, 160	1 (0%)

The worst 5 of 77 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	889	VAL	7.3
1	A	309	SER	5.8
2	P	24	ALA	5.3
2	P	2	ARG	5.2
1	A	416	VAL	4.7

### 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
2	M2L	Q	9	11/12	0.89	0.17	77,87,92,95	0
3	5MC	B	7	20/22	0.91	0.12	75,83,89,91	0
4	C49	C	10	21/23	0.92	0.14	61,71,80,83	0
2	M2L	P	9	11/12	0.97	0.11	57,67,82,85	0

### 6.3 Carbohydrates [i](#)

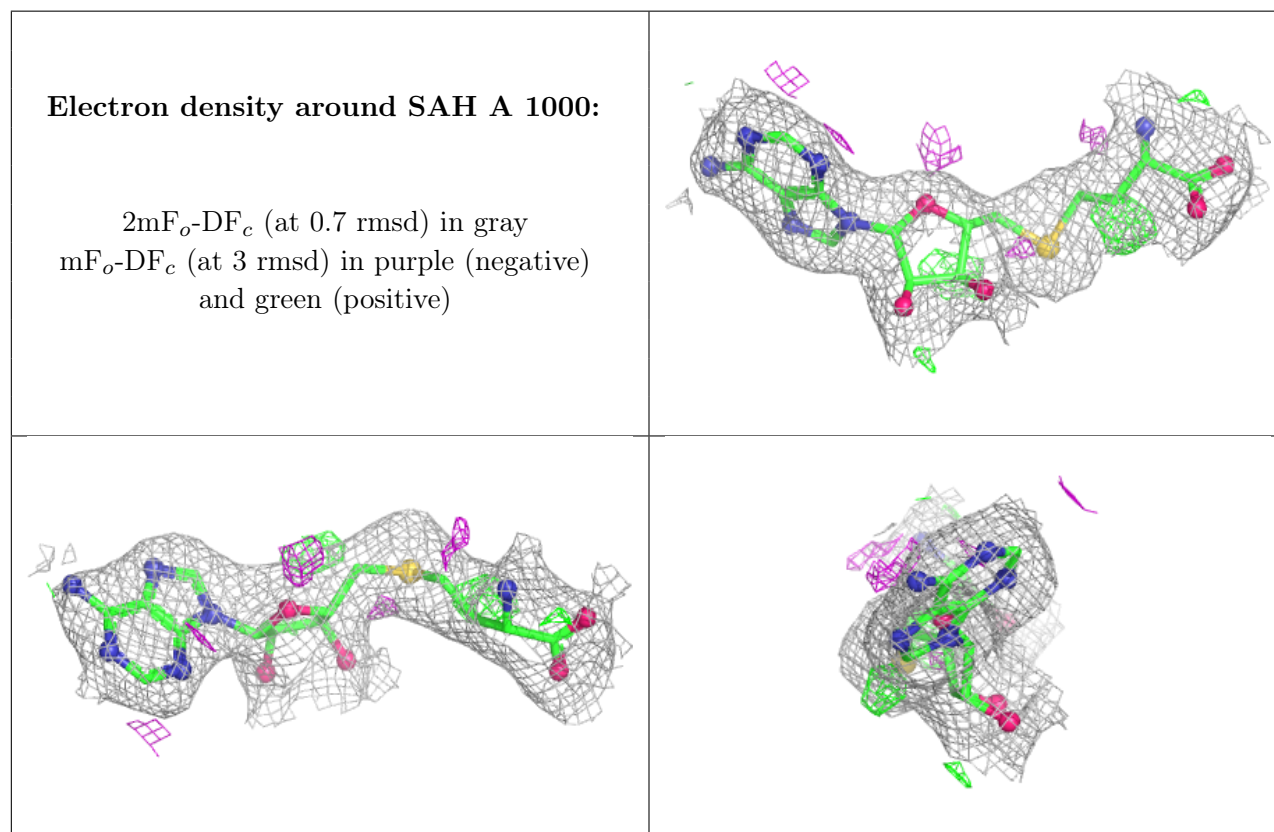
There are no monosaccharides in this entry.

### 6.4 Ligands [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( $\text{\AA}^2$ )	Q<0.9
5	SAH	A	1000	26/26	0.93	0.11	55,64,69,75	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



### 6.5 Other polymers [i](#)

There are no such residues in this entry.