Full wwPDB X-ray Structure Validation Report ⓘ

Jan 16, 2023 – 03:21 am GMT

PDB ID	:	7QW5
Title	:	Adenine-specific DNA methyltransferase M.BseCI complexed with AdoHcy and cognate unmethylated DNA duplex
Authors	:	Mitsikas, D.A.; Kouyianou, K.; Kotsifaki, D.; Providaki, M.; Bouriotis, V.; Glykos, N.M.; Kokkinidis, M.
Deposited on	:	2022-01-24
Resolution	:	2.30 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

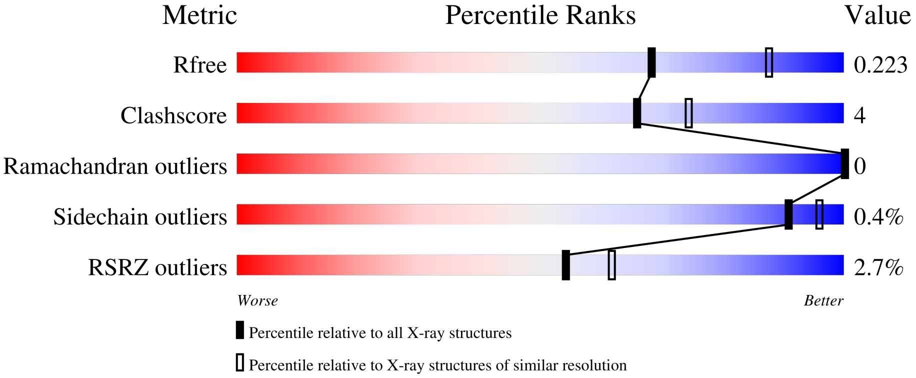
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.3
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

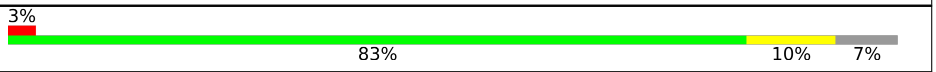
The reported resolution of this entry is 2.30 Å.

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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (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% 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	585	
2	Z	10	
3	Y	10	

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 5049 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 Modification methylase BseCI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	544	4380	2816	736	818	10	0	1	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	195	GLU	GLY	conflict	UNP P43423
A	580	HIS	-	expression tag	UNP P43423
A	581	HIS	-	expression tag	UNP P43423
A	582	HIS	-	expression tag	UNP P43423
A	583	HIS	-	expression tag	UNP P43423
A	584	HIS	-	expression tag	UNP P43423
A	585	HIS	-	expression tag	UNP P43423

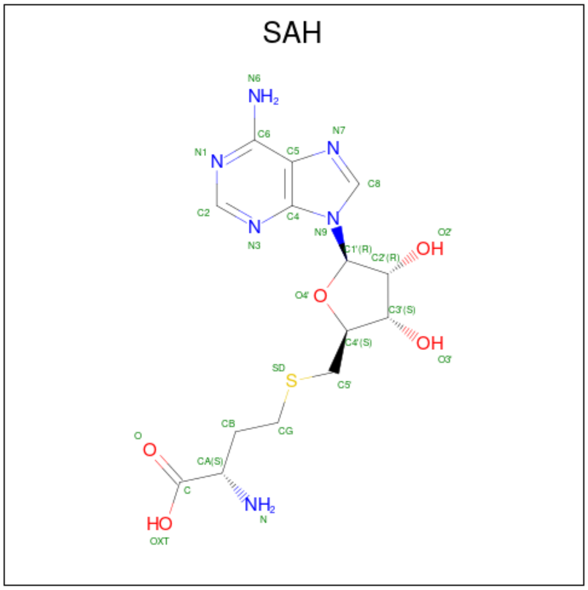
- Molecule 2 is a DNA chain called Unmethylated DNA duplex.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	Z	10	200	96	38	57	9	0	0	0

- Molecule 3 is a DNA chain called Unmethylated DNA duplex.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	Y	10	200	96	38	57	9	0	0	0

- Molecule 4 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$) (labeled as "Ligand of Interest" by depositor).



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

- Molecule 5 is water.

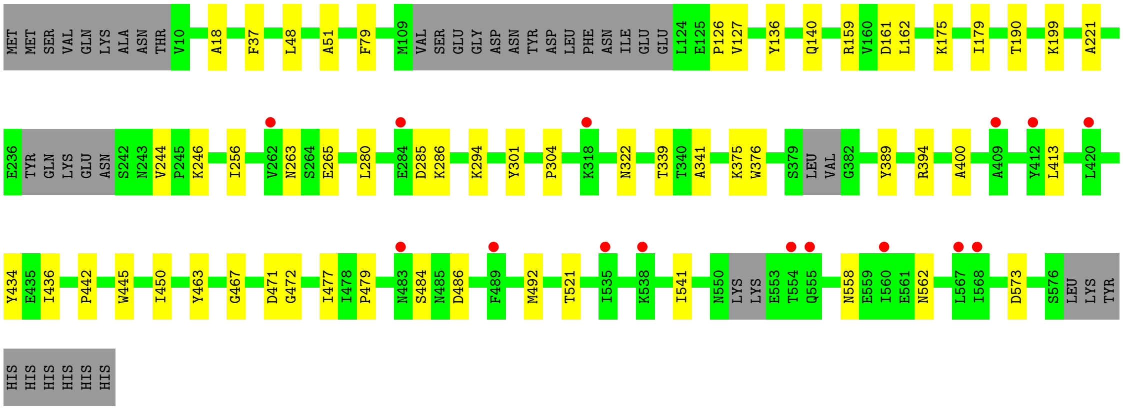
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	210	Total O 210 210	0	0
5	Z	22	Total O 22 22	0	0
5	Y	11	Total O 11 11	0	0

3 Residue-property plots ⓘ

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 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.

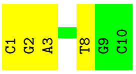
- Molecule 1: Modification methylase BseCI

Chain A:



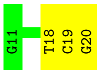
- Molecule 2: Unmethylated DNA duplex

Chain Z:



- Molecule 3: Unmethylated DNA duplex

Chain Y:



4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, α, β, γ	87.04Å 87.04Å 156.35Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	28.49 – 2.30 42.87 – 2.20	Depositor EDS
% Data completeness (in resolution range)	98.7 (28.49-2.30) 98.7 (42.87-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	1.95 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.197 , 0.224 0.197 , 0.223	Depositor DCC
R_{free} test set	1685 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	49.9	Xtriage
Anisotropy	0.179	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 56.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.059 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5049	wwPDB-VP
Average B, all atoms (Å^2)	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.37% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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	0.25	0/4473	0.44	0/6040
2	Z	0.58	0/224	0.99	0/344
3	Y	0.58	0/224	0.98	0/344
All	All	0.30	0/4921	0.52	0/6728

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	4380	0	4360	34	0
2	Z	200	0	110	6	0
3	Y	200	0	110	4	0
4	A	26	0	19	3	0
5	A	210	0	0	1	0
5	Y	11	0	0	0	0
5	Z	22	0	0	0	0
All	All	5049	0	4599	37	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.

All (37) 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:339:THR:HG22	1:A:341:ALA:H	1.57	0.69
2:Z:1:DC:H42	3:Y:20:DG:H1	1.44	0.65
1:A:389:TYR:HE2	1:A:442:PRO:HB3	1.67	0.59
1:A:339:THR:HA	1:A:436:ILE:HG23	1.85	0.59
1:A:199:LYS:NZ	5:A:706:HOH:O	2.35	0.57
1:A:190:THR:HA	1:A:304:PRO:HB3	1.86	0.56
1:A:558:ASN:O	1:A:562:ASN:ND2	2.30	0.56
1:A:413:LEU:HD13	1:A:434:TYR:HB2	1.87	0.55
1:A:37:PHE:HA	1:A:179:ILE:HG13	1.89	0.54
1:A:126:PRO:HB2	1:A:175:LYS:HG3	1.89	0.54
1:A:394:ARG:NH2	1:A:400:ALA:O	2.41	0.53
2:Z:2:DG:H1	3:Y:19:DC:H42	1.57	0.53
1:A:140:GLN:HG3	3:Y:18:DT:H5’	1.94	0.50
1:A:51:ALA:O	4:A:601:SAH:N	2.44	0.49
1:A:375:LYS:HB3	1:A:376:TRP:CE3	2.47	0.49
1:A:322:ASN:HD22	1:A:573:ASP:HB3	1.77	0.49
1:A:161:ASP:OD1	1:A:162:LEU:N	2.45	0.49
1:A:221:ALA:HB1	2:Z:8:DT:OP2	2.13	0.49
1:A:199:LYS:HB2	1:A:301:TYR:CD2	2.49	0.48
2:Z:1:DC:N4	3:Y:20:DG:H1	2.11	0.48
1:A:486:ASP:HB3	1:A:541:ILE:HD13	1.96	0.47
1:A:471:ASP:OD1	1:A:472:GLY:N	2.46	0.47
1:A:244:VAL:O	1:A:244:VAL:HG13	2.15	0.47
1:A:339:THR:HG22	1:A:341:ALA:N	2.26	0.46
1:A:479:PRO:HG3	1:A:484:SER:HA	1.98	0.45
1:A:256:ILE:HD12	1:A:280:LEU:HD13	1.98	0.45
1:A:263:ASN:OD1	1:A:265:GLU:HG2	2.17	0.45
1:A:450:ILE:HB	1:A:477:ILE:HB	1.99	0.45
1:A:445:TRP:O	1:A:467:GLY:HA2	2.18	0.44
1:A:48:LEU:HB2	1:A:127:VAL:HG11	1.99	0.43
1:A:521:THR:HG21	2:Z:3:DA:H5”	2.00	0.43
1:A:51:ALA:HB1	4:A:601:SAH:O4’	2.19	0.43
1:A:463:TYR:HB2	1:A:492:MET:HE1	2.00	0.43
1:A:285:ASP:O	1:A:286:LYS:HD3	2.20	0.42
1:A:246:LYS:HA	1:A:294:LYS:HA	2.03	0.41
1:A:18:ALA:HA	4:A:601:SAH:H3’	2.03	0.40
1:A:159:ARG:HH21	2:Z:3:DA:H2	1.70	0.40

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	535/585 (92%)	517 (97%)	18 (3%)	0	100 100

There are no Ramachandran outliers to report.

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	475/527 (90%)	473 (100%)	2 (0%)	91 96

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	79	PHE
1	A	136	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

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)

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	601	-	24,28,28	0.78	1 (4%)	25,40,40	0.81	1 (4%)

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	601	-	-	0/11/31/31	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	601	SAH	OXT-C	-2.51	1.22	1.30

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	601	SAH	C5-C6-N6	2.01	123.40	120.35

There are no chirality outliers.

There are no torsion outliers.

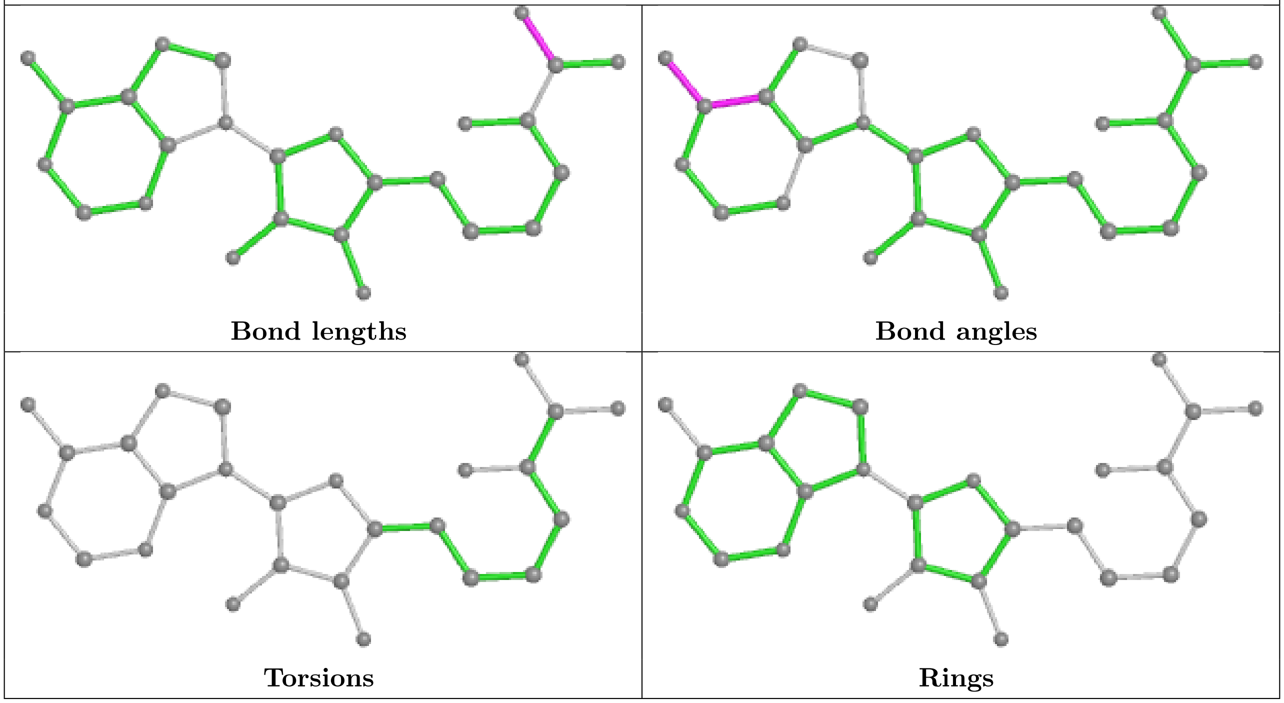
There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	601	SAH	3	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 601



5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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^{th} 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^2)	Q<0.9
1	A	544/585 (92%)	0.06	15 (2%)	53	60	41, 67, 108, 130	0
2	Z	10/10 (100%)	0.42	0	100	100	51, 58, 90, 101	0
3	Y	10/10 (100%)	-0.20	0	100	100	60, 63, 93, 102	0
All	All	564/605 (93%)	0.06	15 (2%)	54	62	41, 67, 108, 130	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	535	ILE	3.5
1	A	412	TYR	3.3
1	A	555	GLN	3.2
1	A	489	PHE	3.2
1	A	409	ALA	3.2
1	A	568	ILE	2.6
1	A	262	VAL	2.6
1	A	420	LEU	2.5
1	A	483	ASN	2.5
1	A	567	LEU	2.4
1	A	538	LYS	2.3
1	A	560	ILE	2.3
1	A	554	THR	2.3
1	A	284	GLU	2.2
1	A	318	LYS	2.1

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

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

6.3 Carbohydrates (i)

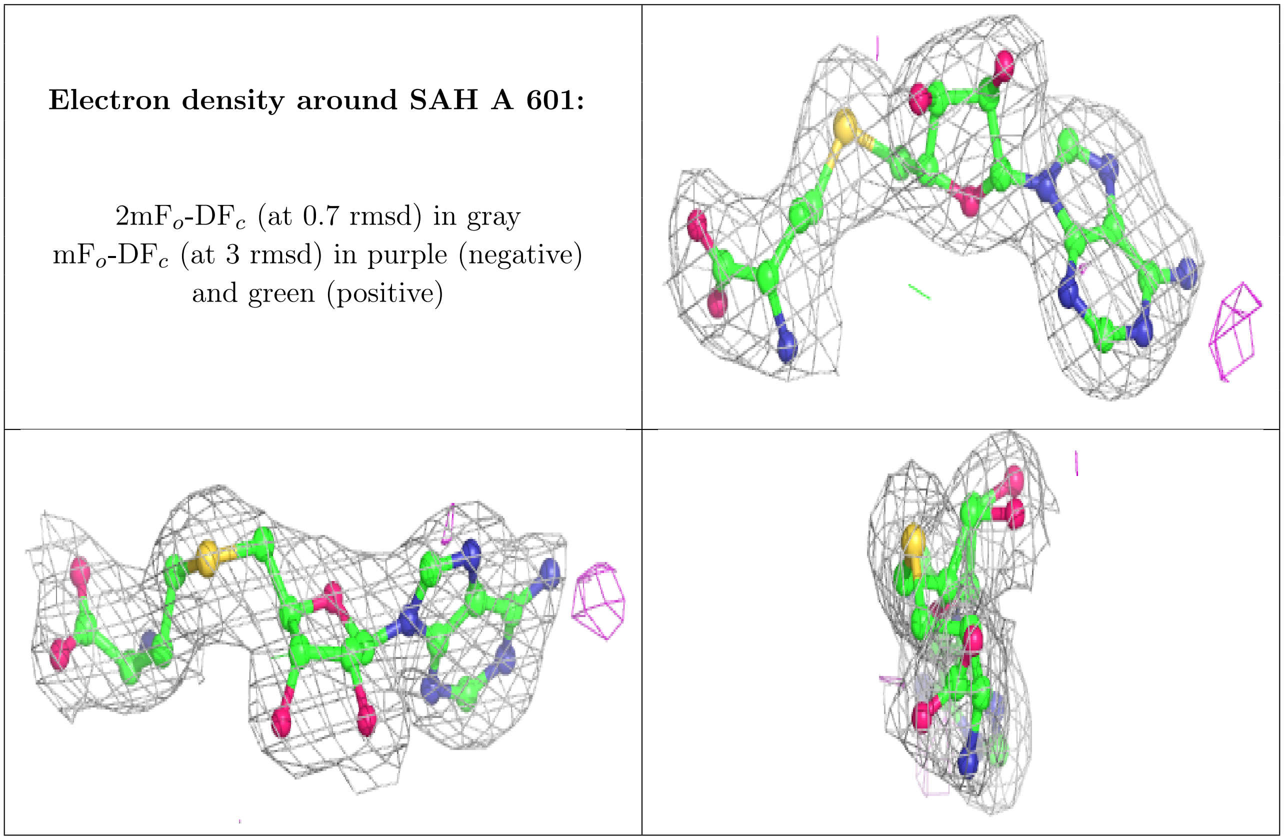
There are no monosaccharides in this entry.

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^{th} 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($Å^2$)	Q<0.9
4	SAH	A	601	26/26	0.95	0.10	47,56,59,63	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.



Electron density around SAH A 601:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

6.5 Other polymers (i)

There are no such residues in this entry.