wwPDB X-ray Structure Validation Summary Report (i)

Aug 7, 2020 – 09:47 PM BST

PDB ID	:	6YMM
Title	:	Crystal structure of the SAM-SAH riboswitch with SAM from space group P312
Authors	:	Huang, L.; Lilley, D.M.J.
Deposited on	:	2020-04-08
Resolution	:	2.20 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 following versions of software and data (see references (1)) were used in the production of this report:

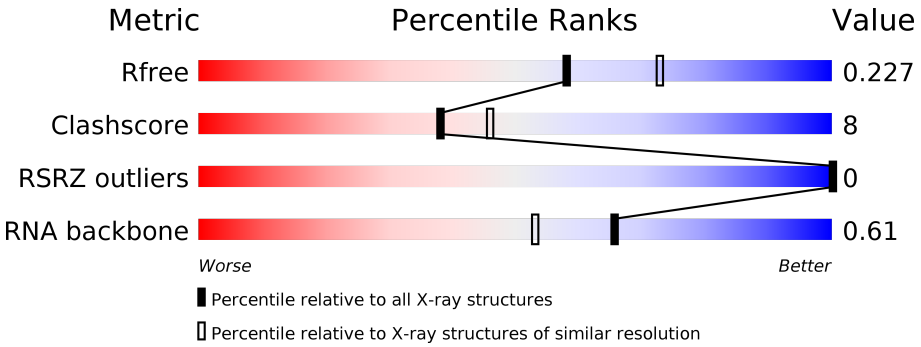
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)
RNA backbone	3102	1032 (2.60-1.80)

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 >=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	26	54% 35% 12%
1	C	26	65% 19% 15%
2	B	9	67% 11% 11% 11%
2	D	9	67% 33%

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
4	NA	A	102	-	-	-	X

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 1613 atoms, of which 51 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.

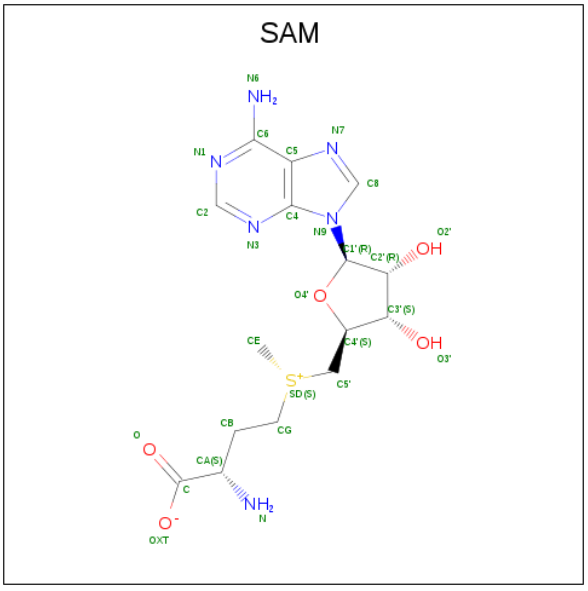
- Molecule 1 is a RNA chain called Chains: A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	26	Total	Br	C	N	O	P	0	0	0
			550	1	246	97	181	25			
1	C	26	Total	Br	C	N	O	P	0	0	0
			550	1	246	97	181	25			

- Molecule 2 is a RNA chain called Chains: B,D.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	9	Total	C	N	O	P	0	0	0
			192	87	37	60	8			
2	D	9	Total	C	N	O	P	0	0	0
			192	87	37	60	8			

- Molecule 3 is S-ADENOSYLMETHIONINE (three-letter code: SAM) (formula: $C_{15}H_{22}N_6O_5S$) (labeled as "Ligand of Interest" by author).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	C	H	N	O	S	0	0
			39	13	17	5	3	1		
3	B	1	Total	C	H	N	O	S	0	0
			39	13	17	5	3	1		
3	C	1	Total	C	H	N	O	S	0	0
			39	13	17	5	3	1		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Na	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	7	Total	O	0	0
			7	7		
5	B	4	Total	O	0	0
			4	4		

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 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: Chains: A

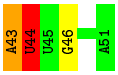
Chain A:

- Molecule 1: Chains: A

Chain C:

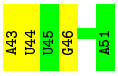
- Molecule 2: Chains: B,D

Chain B:



- Molecule 2: Chains: B,D

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 3 1 2	Depositor
Cell constants a, b, c, α, β, γ	88.17Å 88.17Å 76.08Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.18 – 2.20 76.08 – 2.20	Depositor EDS
% Data completeness (in resolution range)	98.0 (38.18-2.20) 99.5 (76.08-2.20)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.56 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660, PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.192 , 0.221 0.199 , 0.227	Depositor DCC
R_{free} test set	891 reflections (5.16%)	wwPDB-VP
Wilson B-factor (Å^2)	61.5	Xtriage
Anisotropy	0.216	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 73.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.038 for -h,-k,l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	1613	wwPDB-VP
Average B, all atoms (Å^2)	79.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.45% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA, CBV, 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	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	A	0.89	0/589	1.49	12/914 (1.3%)
1	C	0.52	0/589	1.20	2/914 (0.2%)
2	B	0.98	0/215	1.71	3/334 (0.9%)
2	D	0.73	0/215	1.20	0/334
All	All	0.77	0/1608	1.39	17/2496 (0.7%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	24	G	C2-N3-C4	7.43	115.61	111.90
1	C	11	A	OP1-P-OP2	7.09	130.23	119.60
1	A	12	C	C2-N3-C4	6.62	123.21	119.90
2	B	43	A	P-O3'-C3'	6.52	127.53	119.70
2	B	44	U	N3-C4-C5	6.46	118.48	114.60

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	550	0	282	4	0
1	C	550	0	282	10	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	192	0	99	1	0
2	D	192	0	99	1	0
3	A	22	17	17	0	0
3	B	22	17	17	2	0
3	C	22	17	17	2	0
4	A	1	0	0	0	0
5	A	7	0	0	1	0
5	B	4	0	0	2	0
All	All	1562	51	813	18	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 18 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)
2:B:44:U:P	5:B:301:HOH:O	2.49	0.70
3:B:201:SAM:O2'	5:B:301:HOH:O	2.08	0.69
1:C:27:G:O2'	1:C:28:U:H5'	1.98	0.64
1:A:7:G:O2'	1:A:8:G:H5'	1.99	0.62
1:A:24:G:N3	3:B:201:SAM:H2	2.17	0.60

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

There are no protein molecules in this entry.

5.3.2 Protein sidechains

There are no protein molecules in this entry.

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	A	24/26 (92%)	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	C	24/26 (92%)	3 (12%)	1 (4%)
2	B	9/9 (100%)	2 (22%)	1 (11%)
2	D	8/9 (88%)	2 (25%)	0
All	All	65/70 (92%)	7 (10%)	2 (3%)

5 of 7 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	44	U
2	B	46	G
1	C	11	A
1	C	28	U
1	C	29	G

All (2) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	43	A
1	C	28	U

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

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	CBV	A	10	1	15,22,23	1.19	1 (6%)	18,32,35	1.46	3 (16%)
1	CBV	C	10	1	15,22,23	1.09	1 (6%)	18,32,35	1.85	4 (22%)

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	CBV	A	10	1	-	0/5/25/26	0/2/2/2
1	CBV	C	10	1	-	2/5/25/26	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	10	CBV	C4-C5	3.84	1.48	1.41
1	C	10	CBV	C4-C5	3.71	1.48	1.41

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	10	CBV	C5-C4-N4	-4.69	119.06	122.94
1	A	10	CBV	C2-N3-C4	3.95	120.79	116.02
1	C	10	CBV	C2-N3-C4	3.81	120.62	116.02
1	C	10	CBV	N4-C4-N3	3.53	122.03	117.03
1	A	10	CBV	C5-C4-N4	-2.94	120.51	122.94

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	C	10	CBV	O4'-C4'-C5'-O5'
1	C	10	CBV	C3'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	C	10	CBV	1	0

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 for Mogul analysis.

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	SAM	A	101	-	21,24,29	1.12	2 (9%)	20,35,42	1.60	3 (15%)
3	SAM	B	201	-	21,24,29	1.14	1 (4%)	20,35,42	1.72	3 (15%)
3	SAM	C	101	-	21,24,29	1.25	2 (9%)	20,35,42	1.43	2 (10%)

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	SAM	A	101	-	-	1/5/26/33	0/3/3/3
3	SAM	B	201	-	-	4/5/26/33	0/3/3/3
3	SAM	C	101	-	-	2/5/26/33	0/3/3/3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	101	SAM	C2-N3	4.23	1.38	1.32
3	B	201	SAM	C2-N3	3.25	1.37	1.32
3	A	101	SAM	C2-N3	3.14	1.37	1.32
3	C	101	SAM	C2-N1	2.27	1.38	1.33
3	A	101	SAM	O4'-C1'	2.04	1.43	1.41

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	201	SAM	C3'-C2'-C1'	5.10	108.65	100.98
3	A	101	SAM	N3-C2-N1	-4.92	120.98	128.68
3	C	101	SAM	N3-C2-N1	-4.42	121.77	128.68
3	B	201	SAM	N3-C2-N1	-3.78	122.77	128.68
3	C	101	SAM	C1'-N9-C4	-2.84	121.65	126.64

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

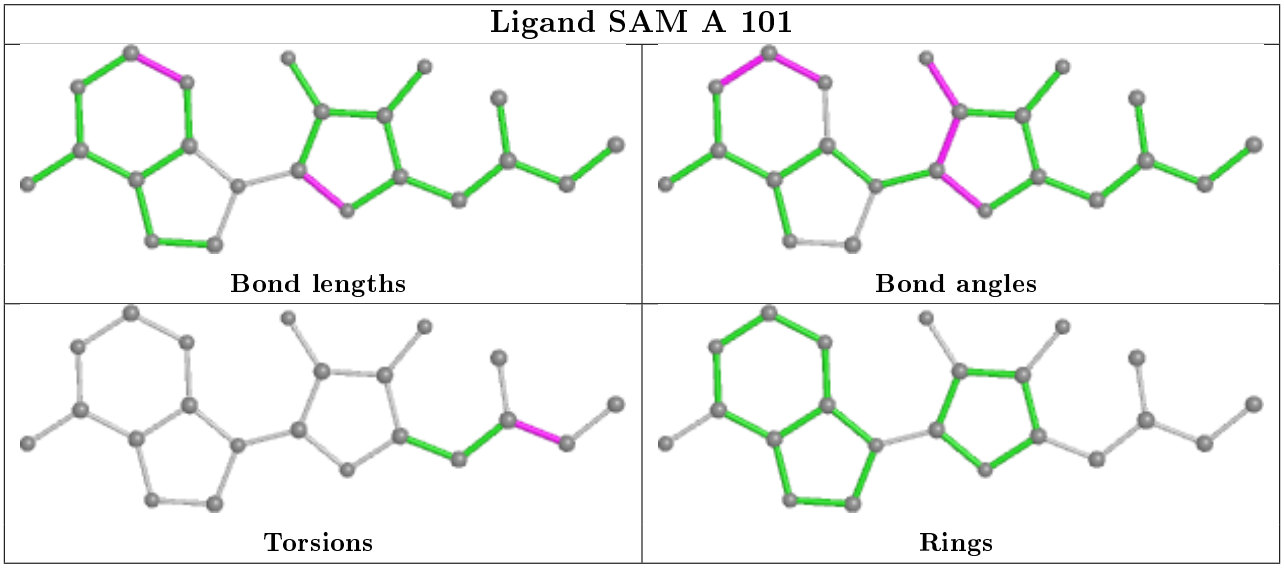
Mol	Chain	Res	Type	Atoms
3	B	201	SAM	O4'-C4'-C5'-SD
3	B	201	SAM	C3'-C4'-C5'-SD
3	C	101	SAM	CB-CG-SD-CE
3	A	101	SAM	CB-CG-SD-CE
3	B	201	SAM	CB-CG-SD-CE

There are no ring outliers.

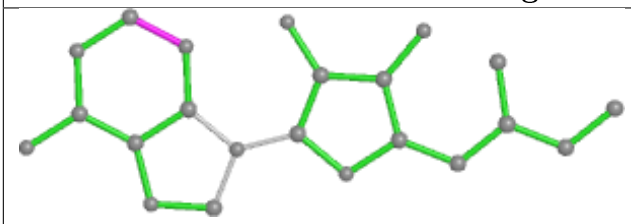
2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	201	SAM	2	0
3	C	101	SAM	2	0

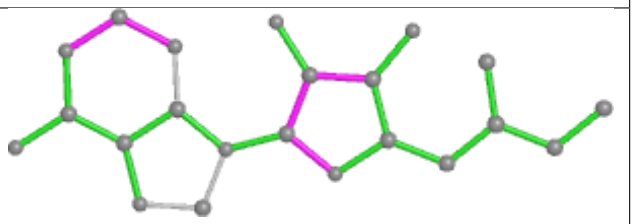
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included.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



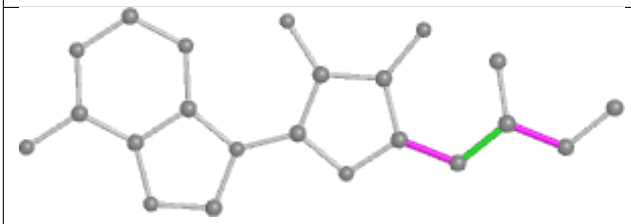
Ligand SAM B 201



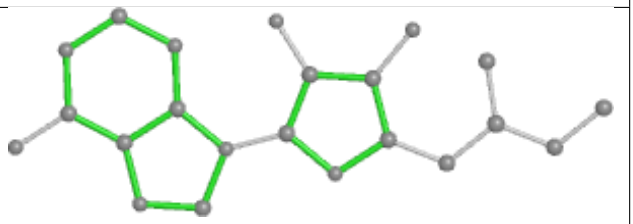
Bond lengths



Bond angles

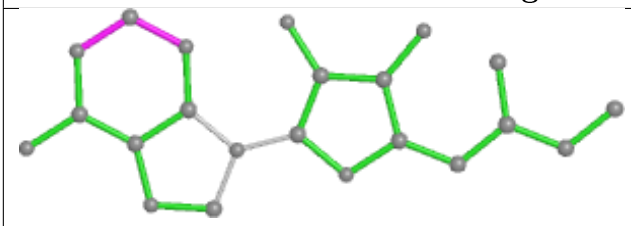


Torsions

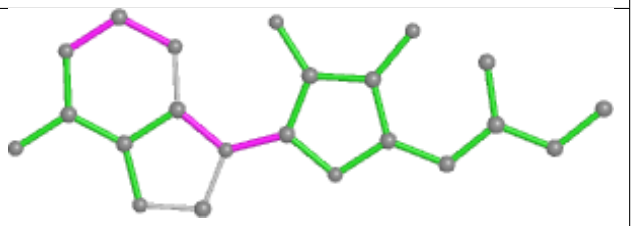


Rings

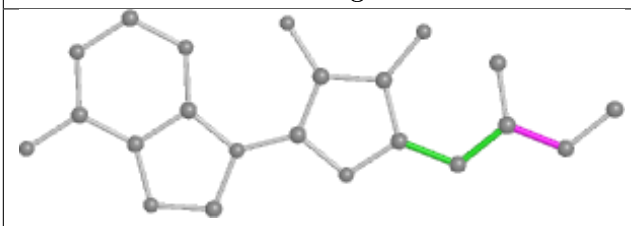
Ligand SAM C 101



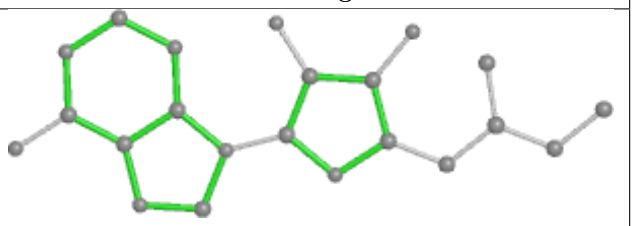
Bond lengths



Bond angles



Torsions



Rings

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

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	25/26 (96%)	0.29	0	100	100	57, 61, 81, 97	0
1	C	25/26 (96%)	0.05	0	100	100	65, 95, 120, 124	0
2	B	9/9 (100%)	0.37	0	100	100	59, 61, 67, 74	0
2	D	9/9 (100%)	0.22	0	100	100	63, 75, 106, 107	0
All	All	68/70 (97%)	0.20	0	100	100	57, 73, 120, 124	0

There are no RSRZ outliers to report.

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

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
1	CBV	A	10	21/22	0.92	0.15	55,65,73,83	0
1	CBV	C	10	21/22	0.93	0.10	85,114,120,124	0

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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

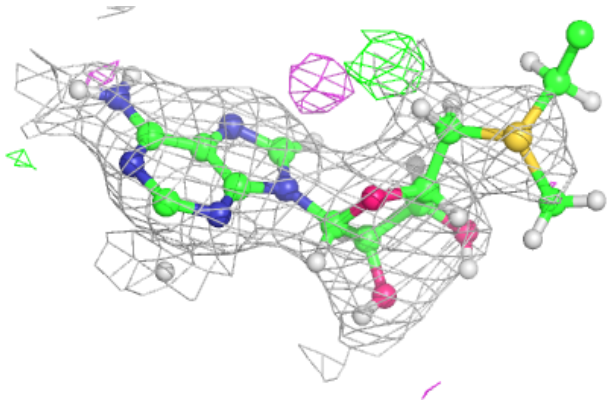
labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

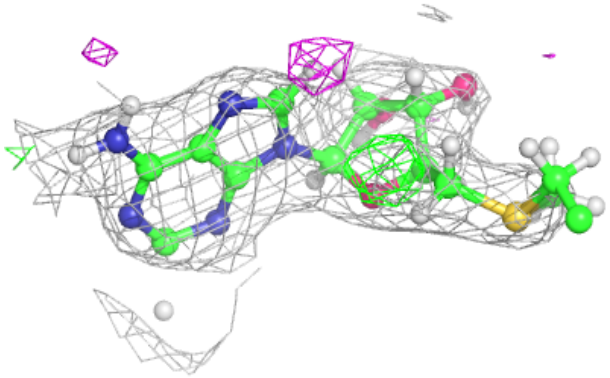
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	NA	A	102	1/1	0.78	0.57	87,87,87,87	0
3	SAM	B	201	22/27	0.95	0.19	0,74,121,135	0
3	SAM	C	101	22/27	0.95	0.20	0,82,114,114	0
3	SAM	A	101	22/27	0.95	0.18	0,58,118,118	0

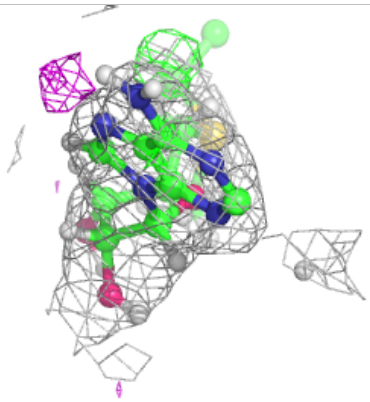
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around SAM B 201:

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

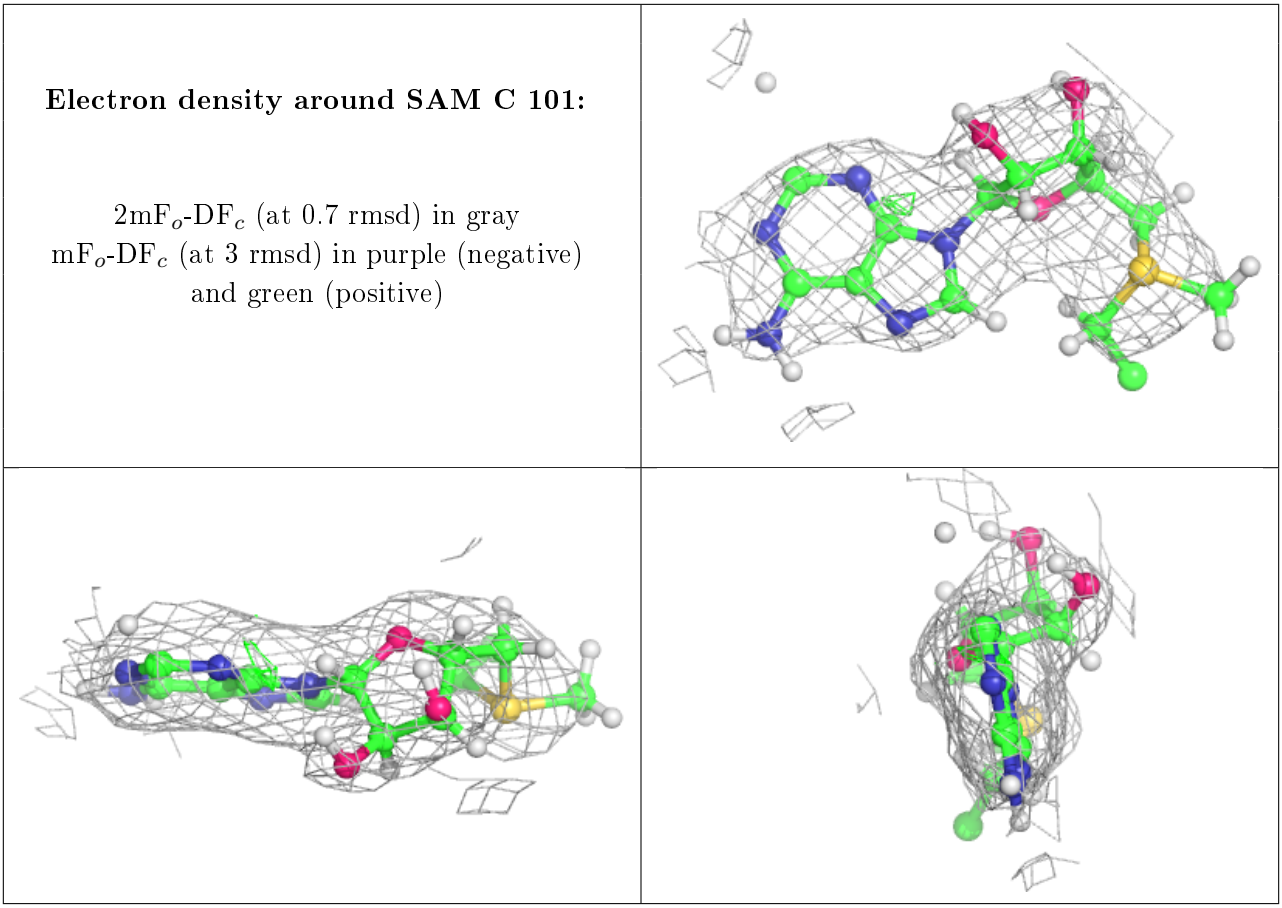






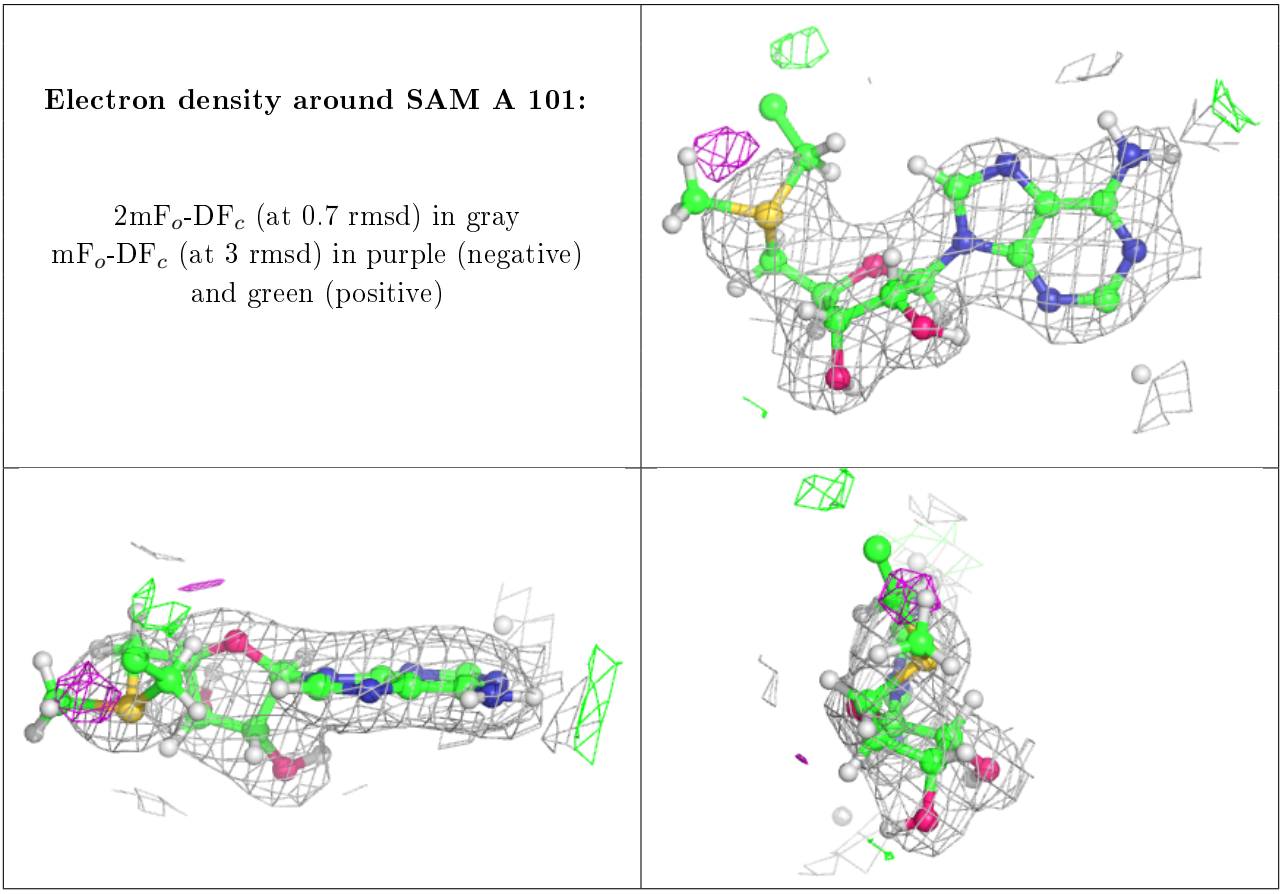
Electron density around SAM C 101:

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



Electron density around SAM A 101:

$2mF_o$-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.