wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 01:49 am BST

PDB ID	:	5XLX
Title	:	Crystal structure of the C-terminal domain of CheR1 containing SAH
Authors	:	Yuan, Z.; Zhu, Y.; Gu, L.
Deposited on	:	2017-05-12
Resolution	:	1.97 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 (i)) were used in the production of this report:

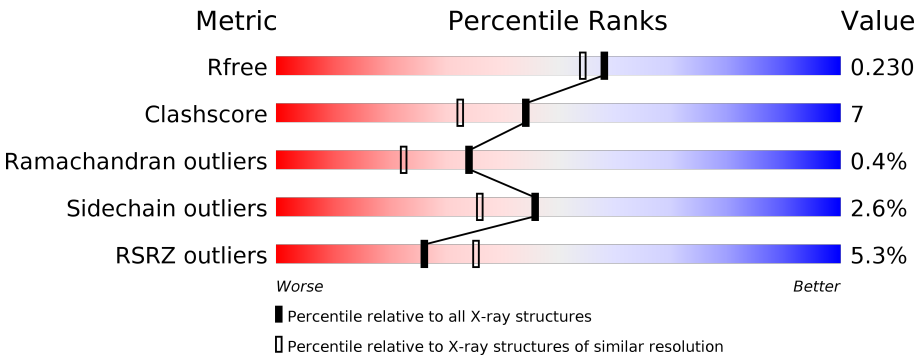
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.97 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	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	282	% 66% 5% 29%
1	B	282	2% 63% 8% 29%
1	C	282	% 63% 7% • 29%
1	D	282	11% 55% 15% • 29%

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	SAH	D	301	-	-	X	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7154 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 Chemotaxis protein methyltransferase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	201	1582	1013	278	283	8	0	0	0
1	B	201	1582	1013	278	283	8	0	0	0
1	C	201	1582	1013	278	283	8	0	0	0
1	D	201	1582	1013	278	283	8	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

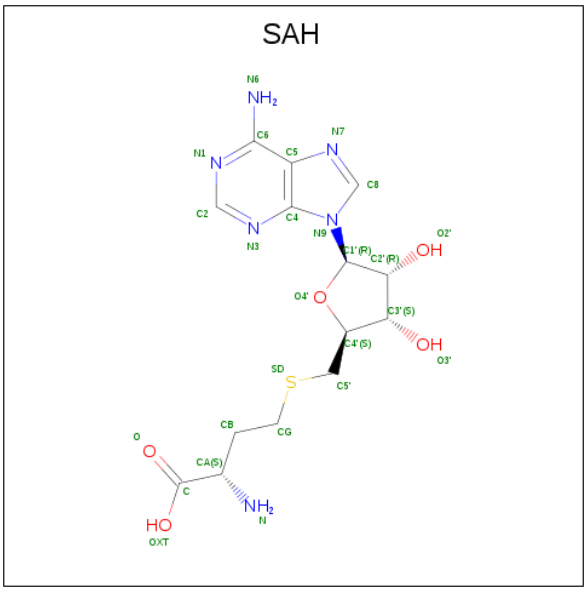
Chain	Residue	Modelled	Actual	Comment	Reference
A	275	LEU	-	expression tag	UNP O87131
A	276	GLU	-	expression tag	UNP O87131
A	277	HIS	-	expression tag	UNP O87131
A	278	HIS	-	expression tag	UNP O87131
A	279	HIS	-	expression tag	UNP O87131
A	280	HIS	-	expression tag	UNP O87131
A	281	HIS	-	expression tag	UNP O87131
A	282	HIS	-	expression tag	UNP O87131
B	275	LEU	-	expression tag	UNP O87131
B	276	GLU	-	expression tag	UNP O87131
B	277	HIS	-	expression tag	UNP O87131
B	278	HIS	-	expression tag	UNP O87131
B	279	HIS	-	expression tag	UNP O87131
B	280	HIS	-	expression tag	UNP O87131
B	281	HIS	-	expression tag	UNP O87131
B	282	HIS	-	expression tag	UNP O87131
C	275	LEU	-	expression tag	UNP O87131
C	276	GLU	-	expression tag	UNP O87131
C	277	HIS	-	expression tag	UNP O87131
C	278	HIS	-	expression tag	UNP O87131
C	279	HIS	-	expression tag	UNP O87131

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	280	HIS	-	expression tag	UNP O87131
C	281	HIS	-	expression tag	UNP O87131
C	282	HIS	-	expression tag	UNP O87131
D	275	LEU	-	expression tag	UNP O87131
D	276	GLU	-	expression tag	UNP O87131
D	277	HIS	-	expression tag	UNP O87131
D	278	HIS	-	expression tag	UNP O87131
D	279	HIS	-	expression tag	UNP O87131
D	280	HIS	-	expression tag	UNP O87131
D	281	HIS	-	expression tag	UNP O87131
D	282	HIS	-	expression tag	UNP O87131

- Molecule 2 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$).



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

- Molecule 3 is water.

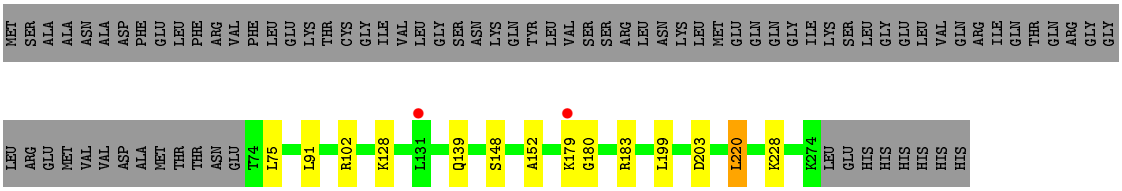
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	256	256	256	0	0
			Total	O		
3	B	172	172	172	0	0
			Total	O		
3	C	192	192	192	0	0
			Total	O		
3	D	102	102	102	0	0

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 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.

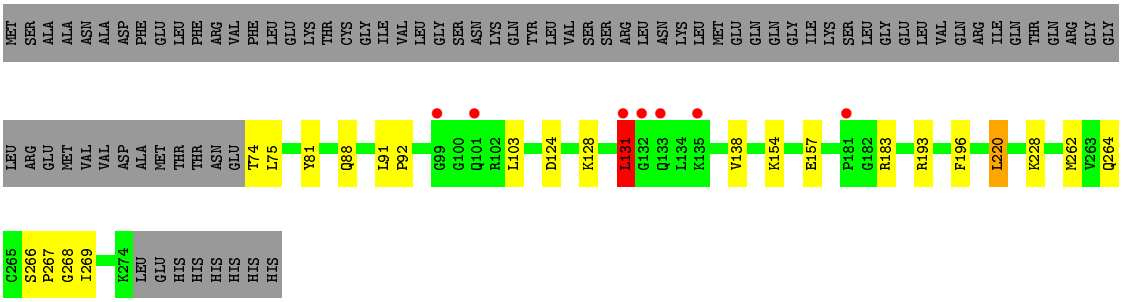
- Molecule 1: Chemotaxis protein methyltransferase 1

Chain A:



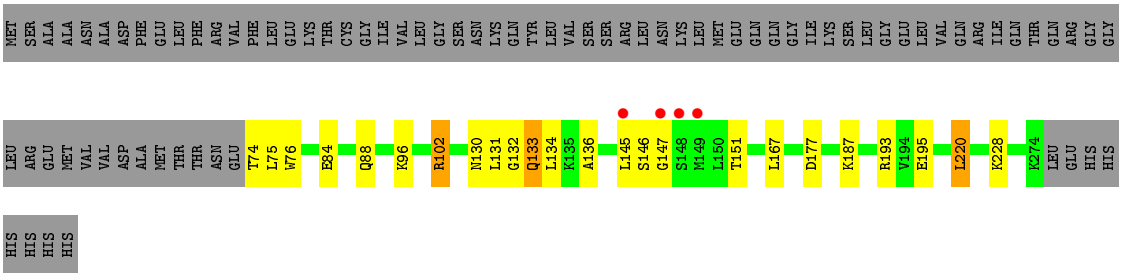
- Molecule 1: Chemotaxis protein methyltransferase 1

Chain B:



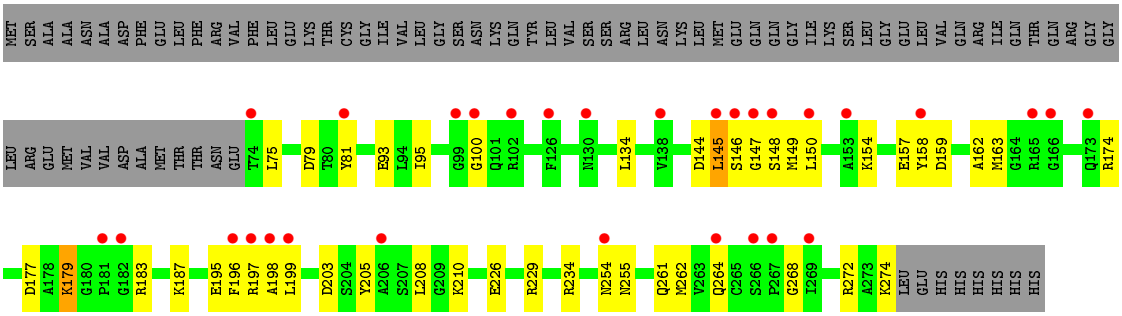
- Molecule 1: Chemotaxis protein methyltransferase 1

Chain C:



- Molecule 1: Chemotaxis protein methyltransferase 1

Chain D:


MET SER ALA ALA ASN ALA ASP PHE GLU LEU PHE ARG VAL PHE LEU GLU LYS THR CYS GLY ILE VAL LEU GLY SER ASN LYS GLN TYR LEU VAL SER SER ARG LEU ASN LYS LEU MET GLU GLN GLN GLY ILE LYS SER LEU GLY GLU LEU VAL GLN ARG ILE GLN THR GLN ARG GLY GLY
LEU ARG GLU MET VAL VAL ASP ALA MET THR THR ASN GLU T74 L75 D79 T80 Y81 E93 L94 I95 G99 G100 Q101 R102 F126 N130 L134 V138 D144 L145 S146 G147 S148 M149 L150 A153 K154 E157 Y158 D159 A162 M163 G164 R165 G166 Q173 R174
D177 A178 K179 G180 P181 G182 R183 K187 E195 F196 R197 A198 L199 D203 S204 Y205 A206 S207 L208 G209 K210 E226 R229 R234 N254 N255 Q261 M262 V263 Q264 C265 S266 P267 G268 I269 R272 A273 K274 LEU GLU HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	36.24Å 78.39Å 140.81Å 90.00° 95.93° 90.00°	Depositor
Resolution (Å)	40.11 – 1.97 40.11 – 1.97	Depositor EDS
% Data completeness (in resolution range)	93.8 (40.11-1.97) 93.7 (40.11-1.97)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.22 (at 1.97Å)	Xtriage
Refinement program	PHENIX dev_2719	Depositor
R, R_{free}	0.191 , 0.230 0.191 , 0.230	Depositor DCC
R_{free} test set	2000 reflections (3.76%)	wwPDB-VP
Wilson B-factor (Å^2)	21.2	Xtriage
Anisotropy	0.057	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 56.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.028 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7154	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.79% 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	0/1616	0.47	0/2181
1	B	0.25	0/1616	0.48	1/2181 (0.0%)
1	C	0.25	0/1616	0.45	0/2181
1	D	0.26	0/1616	0.45	0/2181
All	All	0.25	0/6464	0.46	1/8724 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	131	LEU	CA-CB-CG	7.14	131.72	115.30

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	1582	0	1602	10	0
1	B	1582	0	1602	14	0
1	C	1582	0	1602	18	0
1	D	1582	0	1602	47	0
2	A	26	0	19	0	0
2	B	26	0	19	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	26	0	19	0	0
2	D	26	0	19	11	0
3	A	256	0	0	5	0
3	B	172	0	0	7	2
3	C	192	0	0	6	1
3	D	102	0	0	7	1
All	All	7154	0	6484	85	3

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 7.

The worst 5 of 85 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:D:145:LEU:HD21	2:D:301:SAH:C2	1.72	1.18
1:D:145:LEU:CD2	2:D:301:SAH:C2	2.25	1.14
1:C:102:ARG:NH1	3:C:401:HOH:O	1.81	1.14
1:D:145:LEU:HD21	2:D:301:SAH:N1	1.66	1.11
1:C:146:SER:OG	1:D:146:SER:O	1.76	1.01

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:B:551:HOH:O	3:D:488:HOH:O[1_565]	1.96	0.24
3:C:469:HOH:O	3:C:543:HOH:O[1_655]	2.14	0.06
3:B:421:HOH:O	3:B:463:HOH:O[1_455]	2.15	0.05

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	199/282 (71%)	189 (95%)	10 (5%)	0	100	100
1	B	199/282 (71%)	189 (95%)	9 (4%)	1 (0%)	29	17
1	C	199/282 (71%)	189 (95%)	10 (5%)	0	100	100
1	D	199/282 (71%)	188 (94%)	9 (4%)	2 (1%)	15	6
All	All	796/1128 (71%)	755 (95%)	38 (5%)	3 (0%)	34	22

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	131	LEU
1	D	148	SER
1	D	179	LYS

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	
1	A	165/236 (70%)	161 (98%)	4 (2%)	49	40
1	B	165/236 (70%)	160 (97%)	5 (3%)	41	30
1	C	165/236 (70%)	160 (97%)	5 (3%)	41	30
1	D	165/236 (70%)	162 (98%)	3 (2%)	59	53
All	All	660/944 (70%)	643 (97%)	17 (3%)	46	36

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	220	LEU
1	B	264	GLN
1	C	220	LEU
1	B	131	LEU
1	D	75	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	254	ASN
1	D	139	GLN

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)

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	$\#\|Z\| > 2$
2	SAH	B	301	-	21,28,28	1.19	2 (9%)	20,40,40	1.65	2 (10%)
2	SAH	D	301	-	21,28,28	1.18	2 (9%)	20,40,40	1.74	3 (15%)
2	SAH	A	301	-	21,28,28	1.19	2 (9%)	20,40,40	1.69	2 (10%)
2	SAH	C	301	-	21,28,28	1.19	2 (9%)	20,40,40	1.69	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
2	SAH	B	301	-	-	1/7/31/31	0/3/3/3
2	SAH	D	301	-	-	1/7/31/31	0/3/3/3
2	SAH	A	301	-	-	1/7/31/31	0/3/3/3
2	SAH	C	301	-	-	1/7/31/31	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	301	SAH	C2-N3	3.99	1.38	1.32
2	A	301	SAH	C2-N3	3.95	1.38	1.32
2	C	301	SAH	C2-N3	3.94	1.38	1.32
2	D	301	SAH	C2-N3	3.93	1.38	1.32
2	C	301	SAH	C2-N1	2.50	1.38	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	301	SAH	N3-C2-N1	-5.52	120.05	128.68
2	D	301	SAH	N3-C2-N1	-5.48	120.11	128.68
2	C	301	SAH	N3-C2-N1	-5.47	120.13	128.68
2	B	301	SAH	N3-C2-N1	-5.40	120.23	128.68
2	D	301	SAH	C5'-SD-CG	-3.93	90.47	102.27

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	301	SAH	CB-CG-SD-C5'
2	A	301	SAH	CB-CG-SD-C5'
2	C	301	SAH	CB-CG-SD-C5'
2	D	301	SAH	CB-CG-SD-C5'

There are no ring outliers.

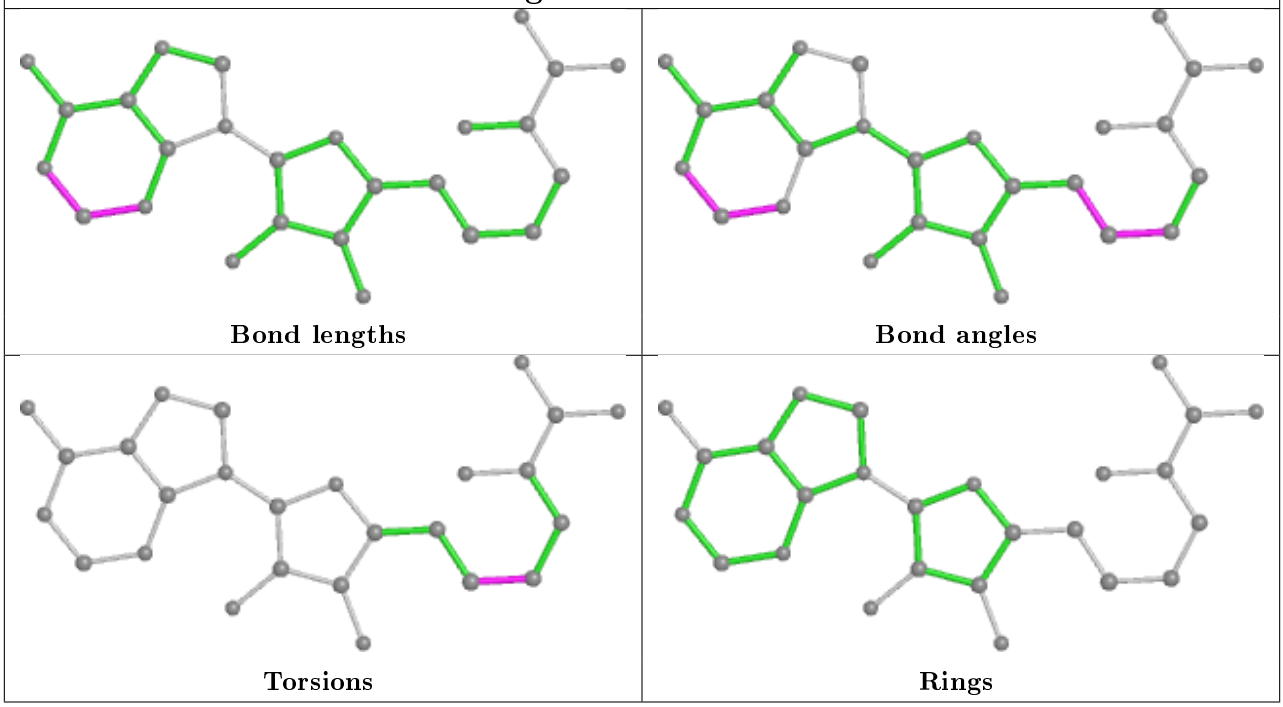
1 monomer is involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	301	SAH	11	0

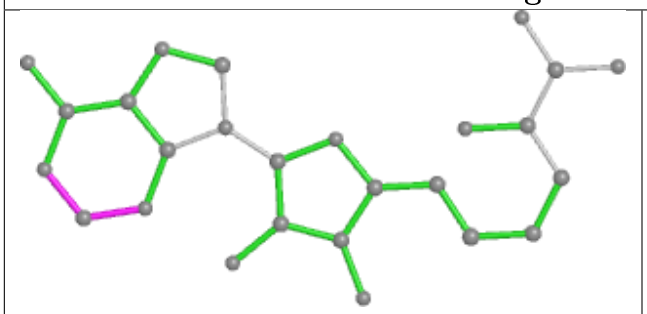
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included.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

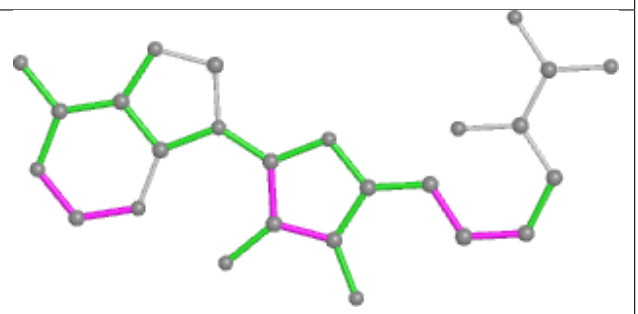
Ligand SAH B 301



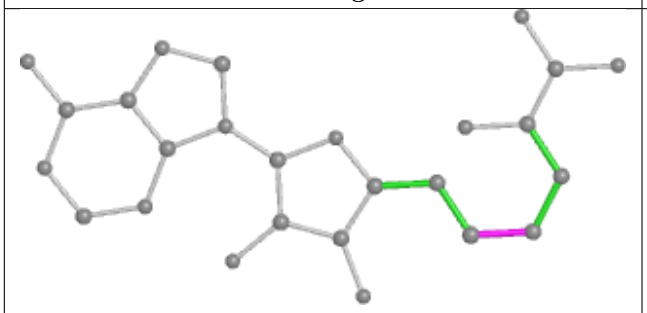
Ligand SAH D 301



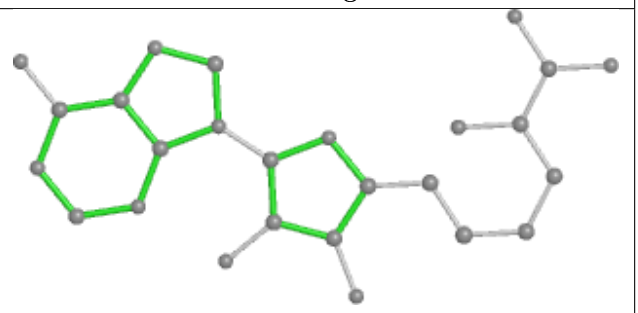
Bond lengths



Bond angles

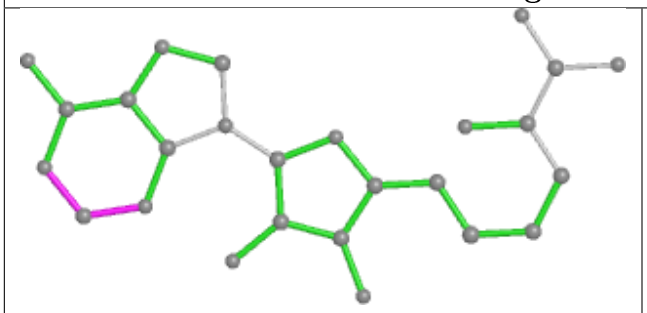


Torsions

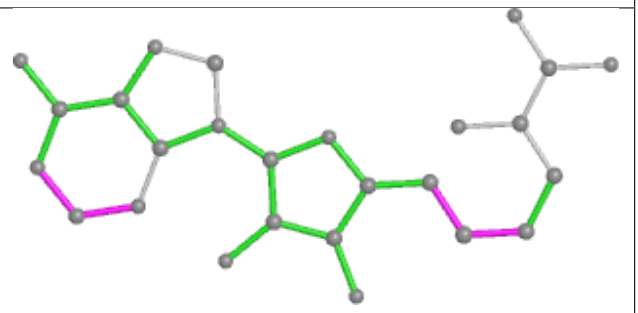


Rings

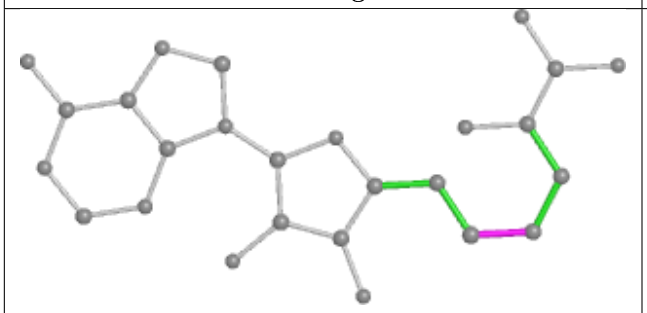
Ligand SAH A 301



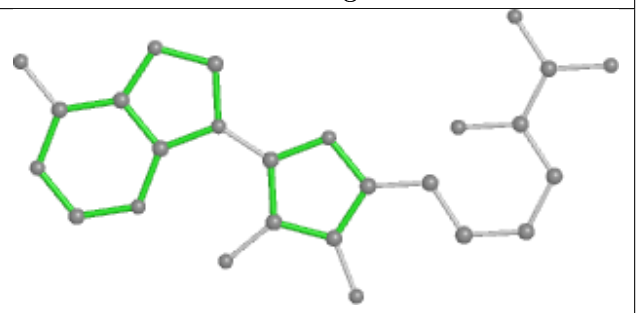
Bond lengths



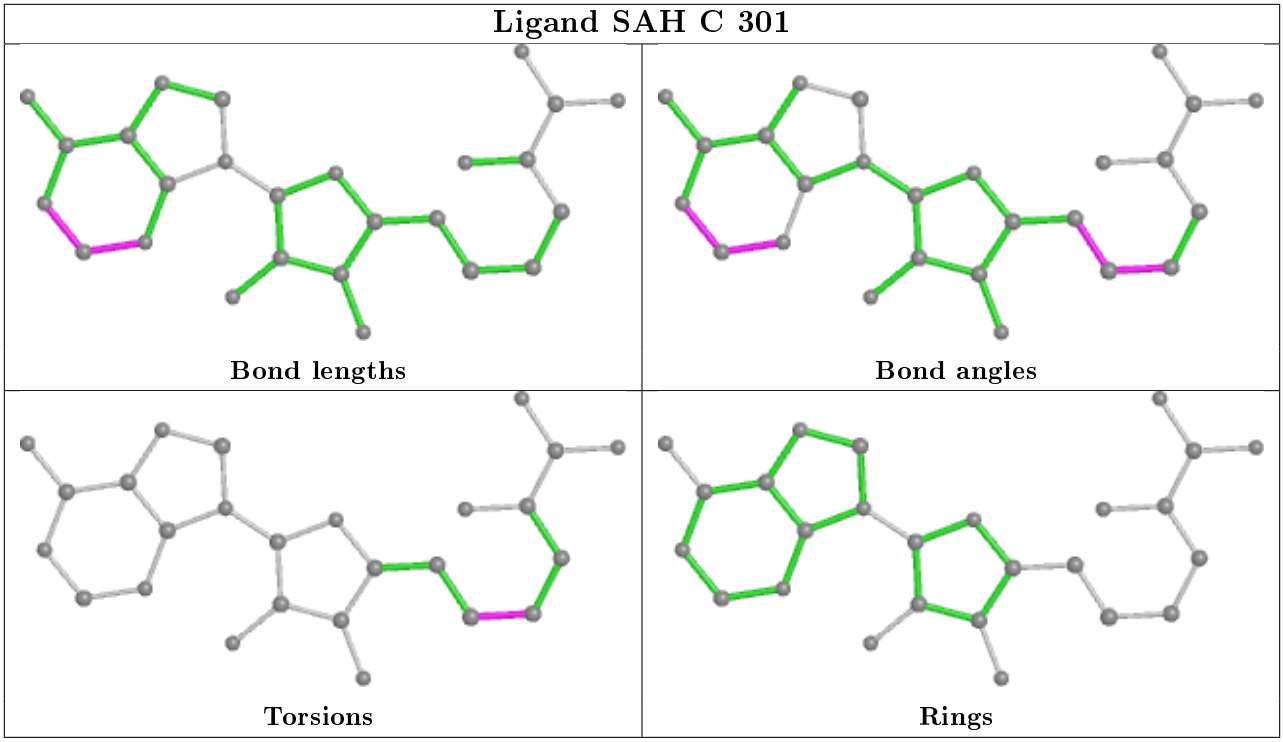
Bond angles



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 (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, 95th 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	201/282 (71%)	-0.08	2 (0%)	82	87	7, 18, 35, 44	0
1	B	201/282 (71%)	0.15	7 (3%)	44	53	10, 23, 47, 64	0
1	C	201/282 (71%)	0.15	4 (1%)	65	73	11, 22, 43, 66	0
1	D	201/282 (71%)	0.88	30 (14%)	2	3	19, 37, 60, 80	1 (0%)
All	All	804/1128 (71%)	0.28	43 (5%)	26	35	7, 25, 50, 80	1 (0%)

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	148	SER	6.8
1	D	148	SER	6.7
1	A	131	LEU	5.2
1	D	165	ARG	4.6
1	D	266	SER	4.2

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

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

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

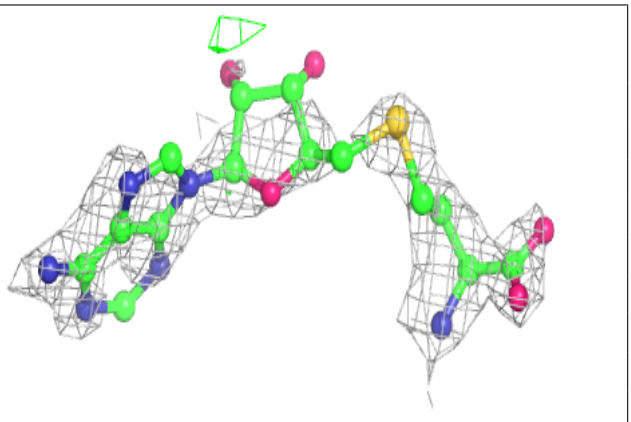
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

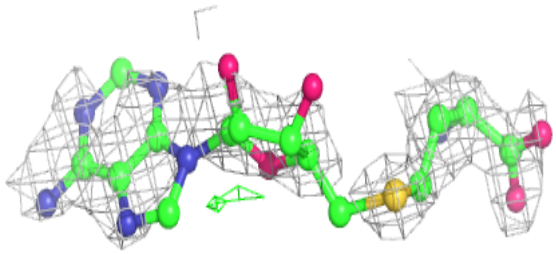
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	SAH	D	301	26/26	0.73	0.28	25,41,49,50	26
2	SAH	A	301	26/26	0.87	0.15	14,28,33,35	26
2	SAH	C	301	26/26	0.89	0.19	13,26,36,40	26
2	SAH	B	301	26/26	0.90	0.15	12,18,22,26	26

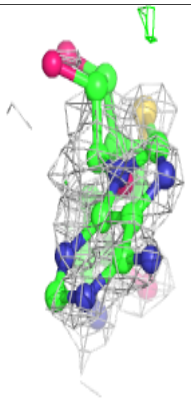
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around SAH D 301:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative) and green (positive)

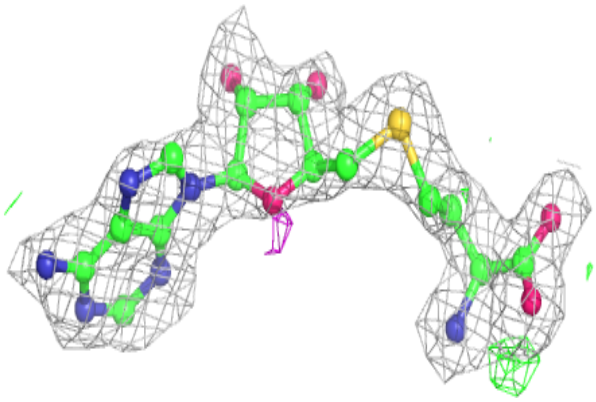


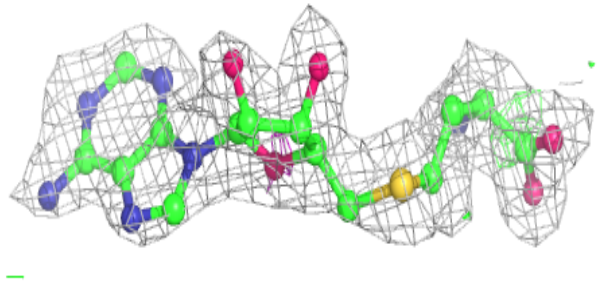


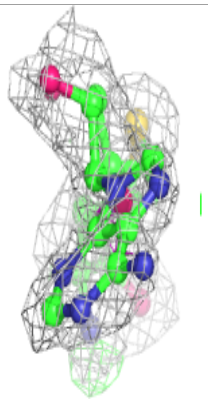


Electron density around SAH A 301:

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

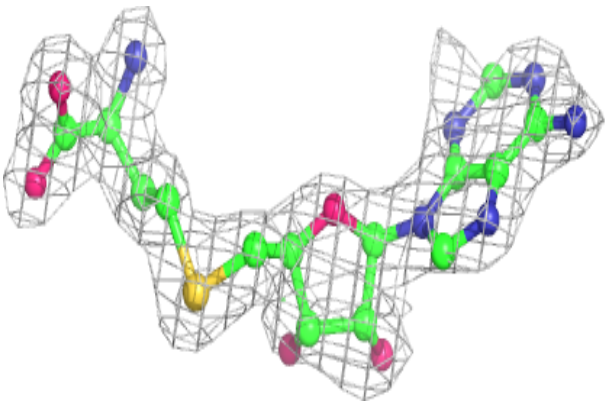


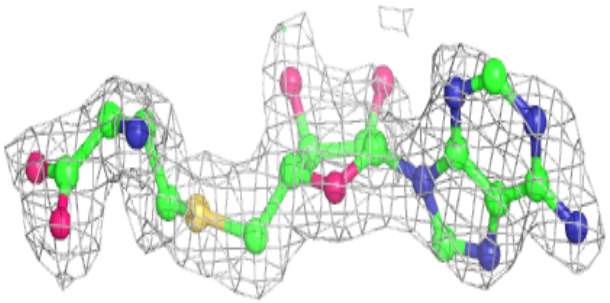


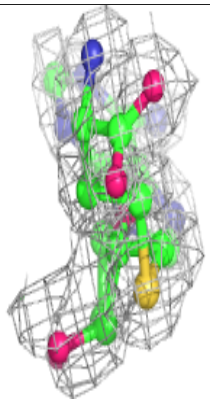


Electron density around SAH C 301:

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

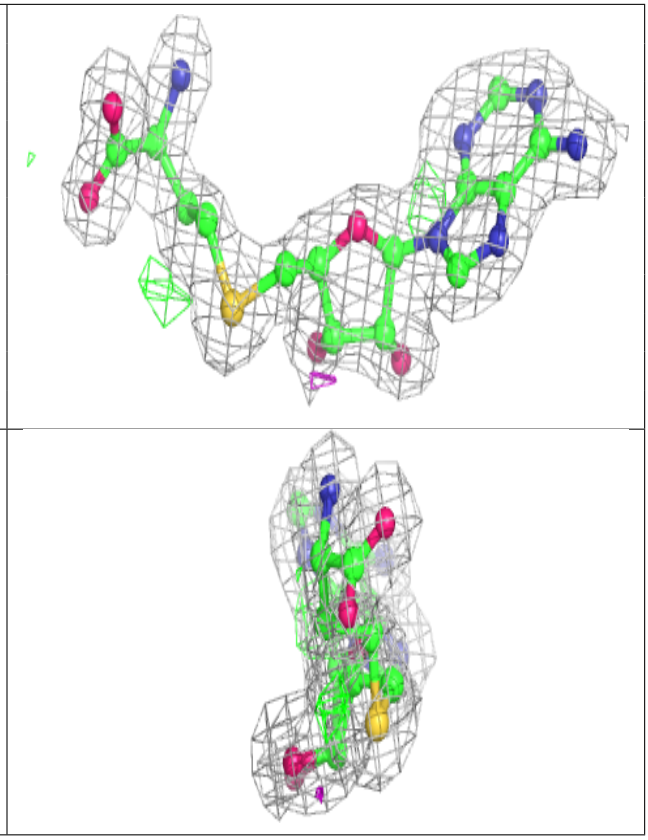






Electron density around SAH B 301:

$2mF_o$-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.