Full wwPDB X-ray Structure Validation Report (i)

Aug 22, 2023 – 02:47 AM EDT

PDB ID	:	2PX2
Title	:	Crystal structure of the Murray Valley Encephalitis Virus NS5 2'-O Methyltransferase domain in complex with SAH (Monoclinic form 1)
Authors	:	Assenberg, R.; Ren, J.; Verma, A.; Walter, T.S.; Alderton, D.; Hurrelbrink, R.J.; Fuller, S.D.; Owens, R.J.; Stuart, D.I.; Grimes, J.M.; Oxford Protein Production Facility (OPPF)
Deposited on	:	2007-05-14
Resolution	:	2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

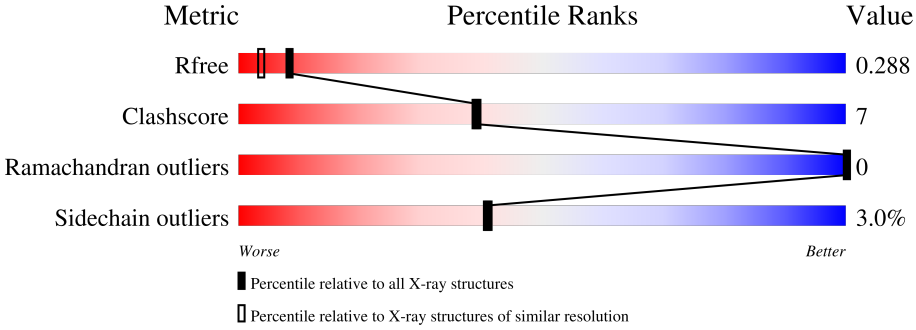
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)

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%

Mol	Chain	Length	Quality of chain
1	A	269	78% 16% 5%
1	B	269	77% 16% 7%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4409 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.

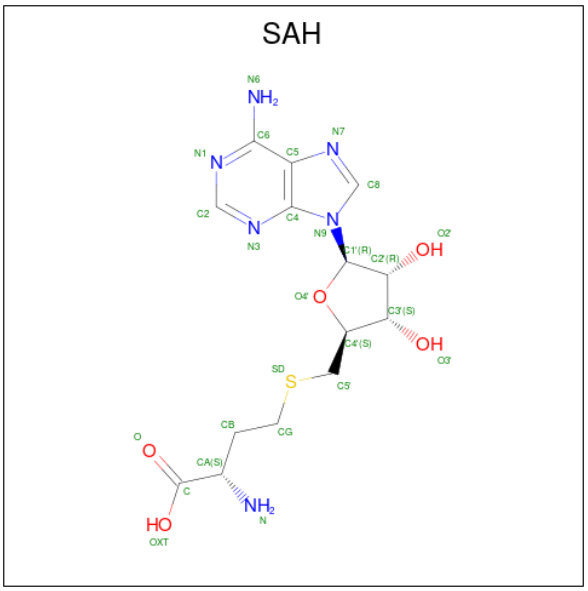
- Molecule 1 is a protein called Genome polyprotein [Contains: Capsid protein C (Core protein); Envelope protein M (Matrix protein); Major envelope protein E; Non-structural protein 1 (NS1); Non-structural protein 2A (NS2A); Flavivirin protease NS2B regulatory subunit; Flavivirin protease NS3 catalytic subunit; Non-structural protein 4A (NS4A); Non-structural protein 4B (NS4B); RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)].

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	255	2025	1278	363	370	14	0	0	0
1	B	251	1992	1259	353	366	14	0	0	0

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
2	A	3	3	3	0	0
2	B	4	4	4	0	0

- Molecule 3 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$).



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

- Molecule 4 is water.

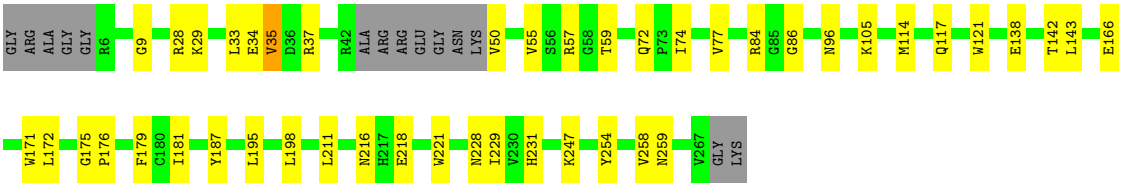
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	187	187	187	0	0
			Total	O		
4	B	146	146	146	0	0

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. 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.

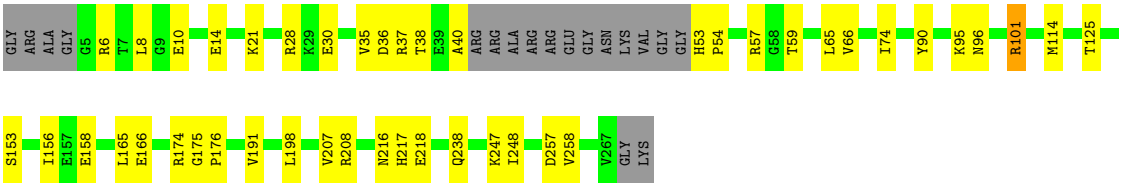
• Molecule 1: Genome polyprotein [Contains: Capsid protein C (Core protein); Envelope protein M (Matrix protein); Major envelope protein E; Non-structural protein 1 (NS1); Non-structural protein 2A (NS2A); Flavivirin protease NS2B regulatory subunit; Flavivirin protease NS3 catalytic subunit; Non-structural protein 4A (NS4A); Non-structural protein 4B (NS4B); RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)]

Chain A:



• Molecule 1: Genome polyprotein [Contains: Capsid protein C (Core protein); Envelope protein M (Matrix protein); Major envelope protein E; Non-structural protein 1 (NS1); Non-structural protein 2A (NS2A); Flavivirin protease NS2B regulatory subunit; Flavivirin protease NS3 catalytic subunit; Non-structural protein 4A (NS4A); Non-structural protein 4B (NS4B); RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)]

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	46.40Å 68.30Å 82.00Å 90.00° 98.40° 90.00°	Depositor
Resolution (Å)	27.40 – 2.00 27.40 – 1.90	Depositor EDS
% Data completeness (in resolution range)	81.7 (27.40-2.00) 84.0 (27.40-1.90)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.19 (at 1.91Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.200 , 0.273 0.221 , 0.288	Depositor DCC
R_{free} test set	1661 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	19.7	Xtriage
Anisotropy	1.020	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 61.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	4409	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 43.70 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.6968e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: CL, 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	0.61	0/2067	0.67	0/2780
1	B	0.59	0/2034	0.69	0/2737
All	All	0.60	0/4101	0.68	0/5517

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	2025	0	2032	30	0
1	B	1992	0	1994	30	0
2	A	3	0	0	0	0
2	B	4	0	0	1	0
3	A	26	0	19	0	0
3	B	26	0	19	0	0
4	A	187	0	0	6	0
4	B	146	0	0	4	0
All	All	4409	0	4064	60	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 7.

All (60) 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:B:248:ILE:O	1:B:248:ILE:HD12	1.76	0.85
1:B:216:ASN:OD1	1:B:218:GLU:HG2	1.81	0.80
1:B:59:THR:CG2	1:B:258:VAL:HG11	2.12	0.80
1:A:74:ILE:HG13	1:A:96:ASN:HB3	1.71	0.72
1:A:228:ASN:HD22	1:A:231:HIS:H	1.38	0.72
1:B:59:THR:HG21	1:B:258:VAL:HG11	1.73	0.71
1:A:72:GLN:HG2	4:A:566:HOH:O	1.91	0.70
1:B:59:THR:HG22	1:B:258:VAL:HG11	1.78	0.66
1:B:166:GLU:HG2	4:B:629:HOH:O	1.96	0.65
1:A:50:VAL:HG13	4:A:674:HOH:O	1.96	0.64
1:B:65:LEU:CD2	1:B:207:VAL:HG21	2.31	0.61
1:A:33:LEU:O	1:A:211:LEU:HA	2.02	0.59
1:A:35:VAL:HG22	4:A:805:HOH:O	2.00	0.59
1:A:181:ILE:HD12	1:A:181:ILE:N	2.19	0.56
1:B:28:ARG:HD3	4:B:842:HOH:O	2.07	0.54
1:B:37:ARG:O	1:B:38:THR:C	2.43	0.54
1:B:158:GLU:HG3	1:B:191:VAL:HG23	1.90	0.54
1:A:55:VAL:HG22	1:A:86:GLY:HA2	1.90	0.54
1:B:65:LEU:HD21	1:B:207:VAL:HG21	1.90	0.54
1:A:28:ARG:HD2	4:A:837:HOH:O	2.08	0.53
1:A:84:ARG:HD2	1:A:114:MET:SD	2.49	0.52
1:A:216:ASN:OD1	1:A:218:GLU:HG2	2.09	0.52
1:A:138:GLU:HG2	1:A:171:TRP:CZ3	2.45	0.51
1:B:74:ILE:HG13	1:B:96:ASN:HB3	1.91	0.51
1:B:217:HIS:O	1:B:217:HIS:ND1	2.42	0.51
1:B:30:GLU:HA	1:B:30:GLU:OE1	2.11	0.51
1:B:153:SER:HB3	1:B:156:ILE:HD12	1.94	0.49
1:B:114:MET:HE3	4:B:749:HOH:O	2.13	0.49
1:B:8:LEU:HD13	1:B:238:GLN:HG3	1.96	0.48
1:A:179:PHE:HB2	1:A:181:ILE:HD11	1.96	0.48
1:A:77:VAL:HG22	1:A:142:THR:HB	1.95	0.48
1:A:143:LEU:HD23	1:A:172:LEU:HD23	1.95	0.48
1:B:35:VAL:HG11	1:B:257:ASP:CG	2.34	0.48
1:A:50:VAL:HG21	1:A:117:GLN:H	1.79	0.47
1:B:10:GLU:O	1:B:14:GLU:HG3	2.16	0.46
1:A:218:GLU:OE1	4:A:606:HOH:O	2.20	0.46
1:A:34:GLU:O	1:A:254:TYR:HA	2.15	0.46
1:A:9:GLY:HA3	1:A:187:TYR:CG	2.51	0.45
1:A:143:LEU:HD23	1:A:172:LEU:CD2	2.45	0.45
1:A:59:THR:HG22	1:A:258:VAL:HG21	1.97	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:35:VAL:HG11	1:B:257:ASP:OD1	2.17	0.44
1:A:198:LEU:HB3	1:A:221:TRP:CH2	2.52	0.44
1:A:57:ARG:HD3	4:A:805:HOH:O	2.17	0.44
1:A:195:LEU:HB3	1:A:229:ILE:HD13	1.99	0.44
1:A:74:ILE:HG13	1:A:96:ASN:CB	2.44	0.44
1:B:54:PRO:HG3	1:B:258:VAL:HG12	1.99	0.44
1:B:165:LEU:HD22	1:B:198:LEU:CD1	2.48	0.44
1:A:50:VAL:CG2	1:A:117:GLN:H	2.31	0.43
1:A:175:GLY:N	1:A:176:PRO:CD	2.81	0.43
1:A:117:GLN:HA	1:A:121:TRP:CD1	2.54	0.42
1:B:35:VAL:CG1	1:B:36:ASP:N	2.82	0.42
1:B:37:ARG:O	1:B:40:ALA:N	2.53	0.42
1:B:175:GLY:N	1:B:176:PRO:CD	2.83	0.42
1:B:66:VAL:HG21	1:B:90:TYR:HE2	1.84	0.42
1:B:54:PRO:HG3	1:B:258:VAL:CG1	2.50	0.41
1:A:37:ARG:HG3	1:A:57:ARG:CZ	2.50	0.41
1:B:6:ARG:O	4:B:877:HOH:O	2.22	0.41
1:B:208:ARG:N	2:B:529:CL:CL	2.74	0.40
1:A:138:GLU:HG2	1:A:171:TRP:CH2	2.56	0.40
1:B:101:ARG:HD3	1:B:125:THR:HB	2.03	0.40

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	251/269 (93%)	249 (99%)	2 (1%)	0	100	100
1	B	247/269 (92%)	242 (98%)	5 (2%)	0	100	100
All	All	498/538 (93%)	491 (99%)	7 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	218/225 (97%)	212 (97%)	6 (3%)	43	44
1	B	215/225 (96%)	208 (97%)	7 (3%)	38	37
All	All	433/450 (96%)	420 (97%)	13 (3%)	41	41

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	29	LYS
1	A	35	VAL
1	A	105	LYS
1	A	166	GLU
1	A	247	LYS
1	A	259	ASN
1	B	21	LYS
1	B	53	HIS
1	B	57	ARG
1	B	95	LYS
1	B	101	ARG
1	B	174	ARG
1	B	247	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	98	GLN
1	A	228	ASN
1	A	259	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 7 are monoatomic - leaving 2 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	SAH	B	500	-	24,28,28	1.18	2 (8%)	25,40,40	1.53	4 (16%)
3	SAH	A	500	-	24,28,28	1.15	3 (12%)	25,40,40	1.73	6 (24%)

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	SAH	B	500	-	-	1/11/31/31	0/3/3/3
3	SAH	A	500	-	-	0/11/31/31	0/3/3/3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	500	SAH	C2-N3	3.79	1.38	1.32
3	B	500	SAH	C2-N3	3.68	1.38	1.32
3	B	500	SAH	C2-N1	2.98	1.39	1.33
3	A	500	SAH	C2-N1	2.47	1.38	1.33
3	A	500	SAH	OXT-C	-2.03	1.23	1.30

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	500	SAH	N3-C2-N1	-5.54	120.02	128.68
3	B	500	SAH	N3-C2-N1	-5.03	120.81	128.68
3	A	500	SAH	C5'-SD-CG	-2.91	93.52	102.27
3	B	500	SAH	OXT-C-CA	2.84	123.07	113.38
3	B	500	SAH	OXT-C-O	-2.81	117.71	124.09
3	A	500	SAH	OXT-C-O	-2.73	117.88	124.09
3	A	500	SAH	CB-CG-SD	-2.28	108.20	113.31
3	B	500	SAH	C1'-N9-C4	-2.15	122.87	126.64
3	A	500	SAH	C1'-N9-C4	-2.14	122.88	126.64
3	A	500	SAH	C4-C5-N7	-2.03	107.28	109.40

There are no chirality outliers.

All (1) torsion outliers are listed below:

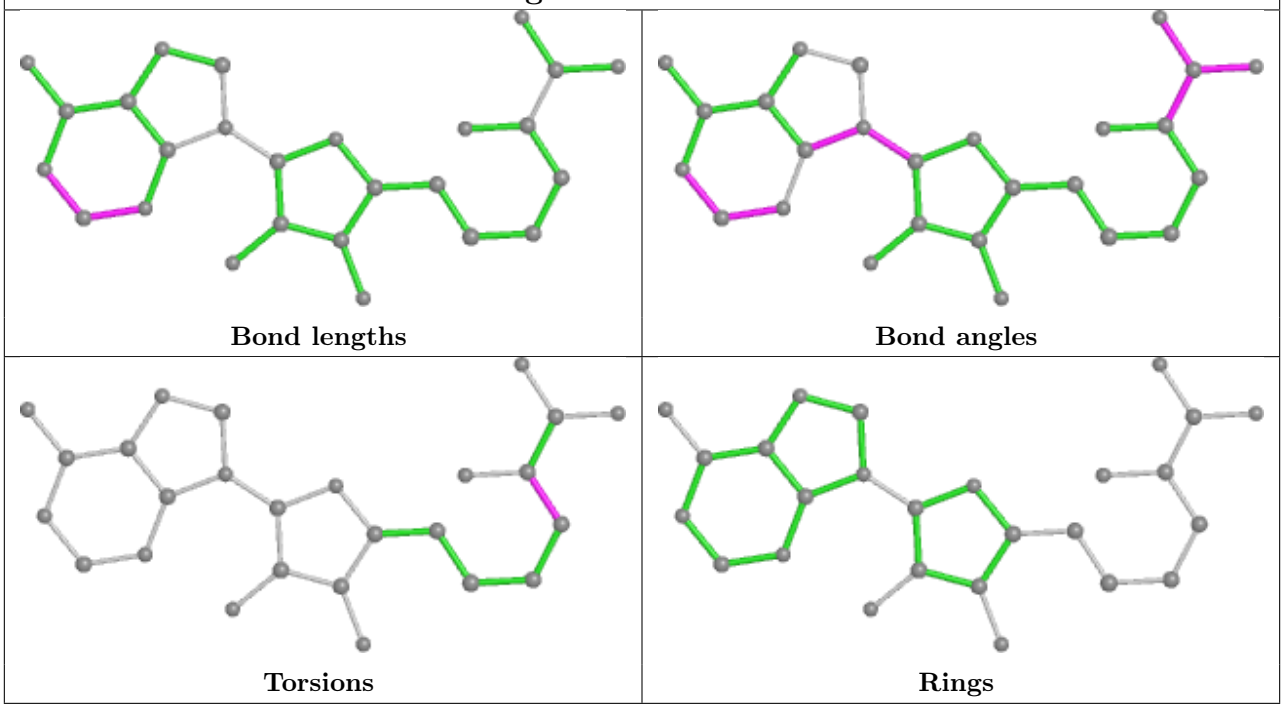
Mol	Chain	Res	Type	Atoms
3	B	500	SAH	C-CA-CB-CG

There are no ring outliers.

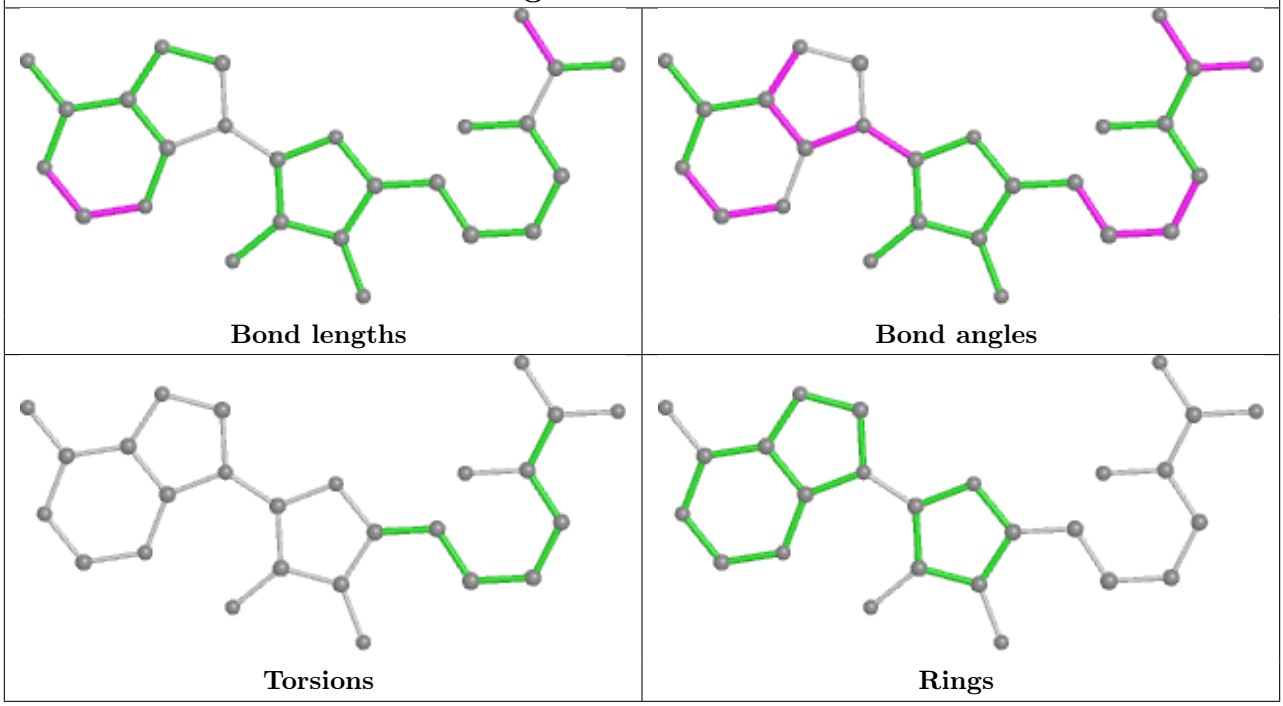
No monomer is involved in short contacts.

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 500



Ligand SAH A 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.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

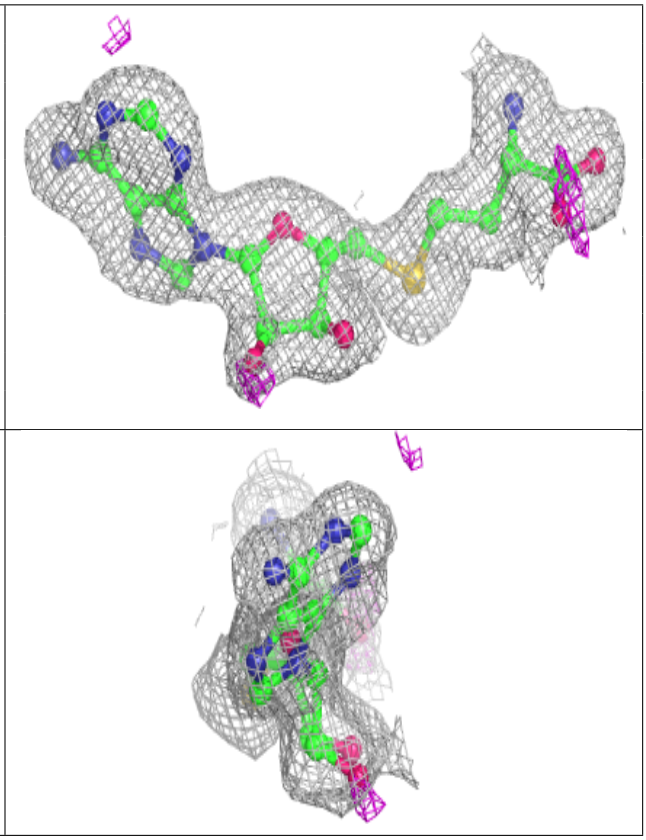
6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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.

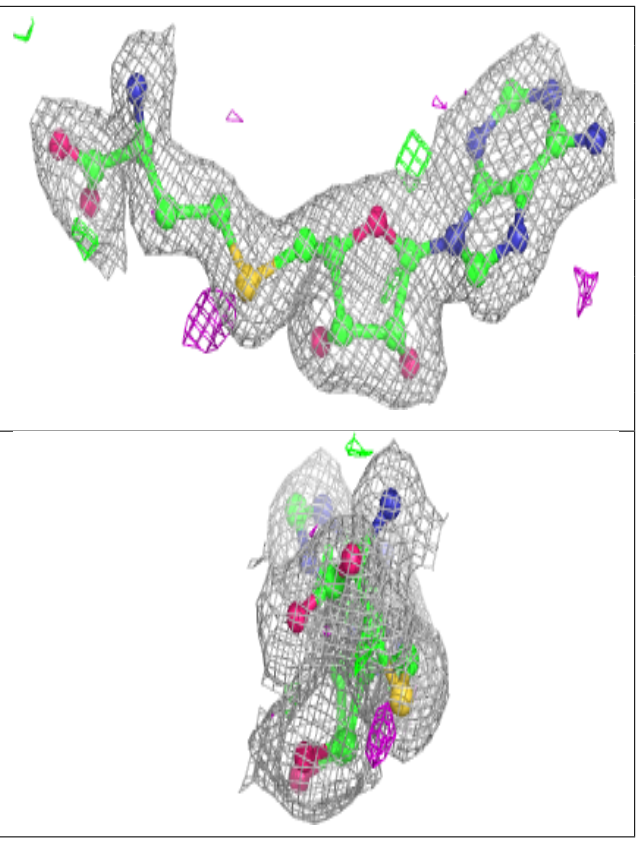
Electron density around SAH A 500:

$2mF_o$-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 500:

$2mF_o$-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)

Unable to reproduce the depositors R factor - this section is therefore empty.