



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 07:32 am BST

PDB ID : 4PGH  
Title : Caffeic acid O-methyltransferase from Sorghum bicolor  
Authors : Green, A.R.; Lewis, K.M.; Kang, C.  
Deposited on : 2014-05-02  
Resolution : 2.80 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 following versions of software and data (see [references ⓘ](#)) 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

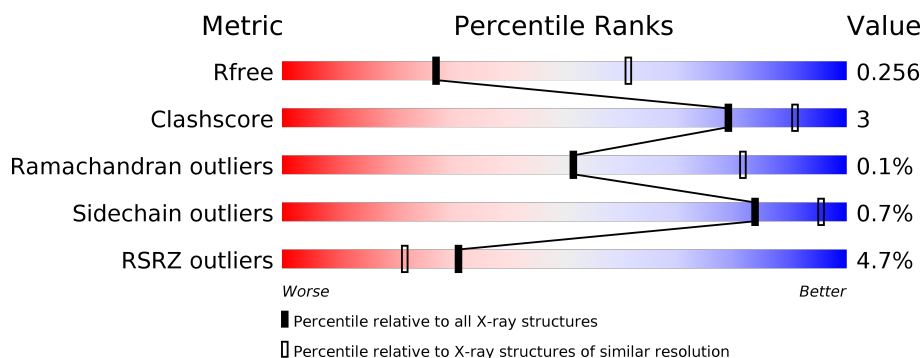
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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  $\geq 3$ , 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	358	<div> <div></div> <div>94%5%</div> </div>
1	B	358	<div> <div>%</div> <div>93%6%</div> </div>
1	C	358	<div> <div></div> <div>91%8%</div> </div>
1	D	358	<div> <div>17%</div> <div>88%12%</div> </div>

## 2 Entry composition [i](#)

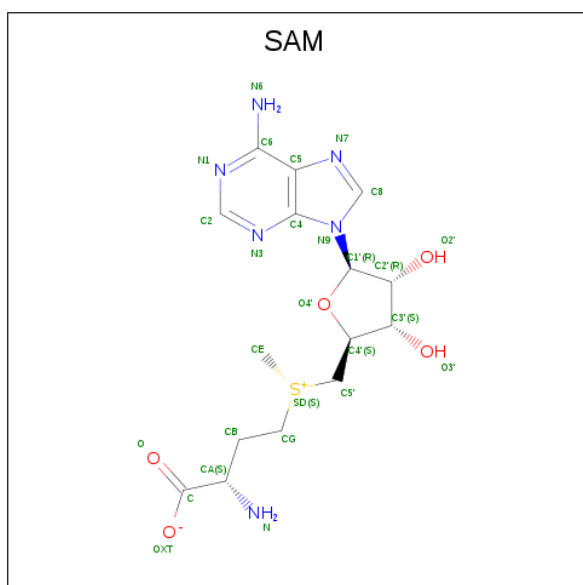
There are 3 unique types of molecules in this entry. The entry contains 10719 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Caffeic acid O-methyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	354	Total	C	N	O	S	0	0	0
			2616	1675	438	483	20			
1	B	357	Total	C	N	O	S	0	0	0
			2533	1608	436	469	20			
1	C	353	Total	C	N	O	S	0	0	0
			2612	1670	437	485	20			
1	D	358	Total	C	N	O	S	0	0	0
			2636	1688	443	485	20			

- Molecule 2 is S-ADENOSYLMETHIONINE (three-letter code: SAM) (formula: C<sub>15</sub>H<sub>22</sub>N<sub>6</sub>O<sub>5</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			27	15	6	5	1		
2	B	1	Total	C	N	O	S	0	0
			27	15	6	5	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	C	1	Total	C	N	O	S	0	0
			27	15	6	5	1		
2	D	1	Total	C	N	O	S	0	0
			27	15	6	5	1		

- Molecule 3 is water.

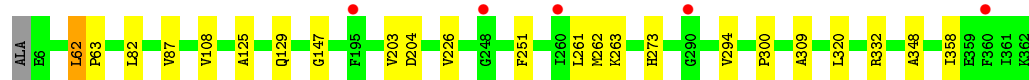
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	62	Total	O	0	0
			62	62		
3	B	43	Total	O	0	0
			43	43		
3	C	78	Total	O	0	0
			78	78		
3	D	31	Total	O	0	0
			31	31		



- Molecule 1: Caffeic acid O-methyltransferase

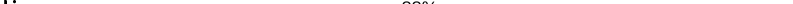


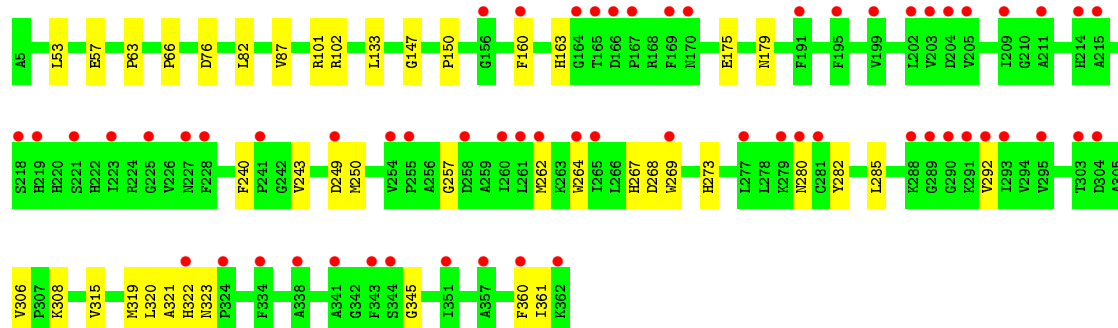
- Chain B:  93% 6%



- Chain C:  91% 8%



- Chain D:  17% 88% 12%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	76.78Å 63.90Å 152.47Å 90.00° 93.02° 90.00°	Depositor
Resolution (Å)	48.94 – 2.80 49.09 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.94-2.80) 97.3 (49.09-2.80)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.30 (at 2.81Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.204 , 0.252 0.208 , 0.256	Depositor DCC
$R_{free}$ test set	1999 reflections (5.45%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	58.7	Xtriage
Anisotropy	0.418	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 47.6	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	10719	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	69.0	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 53.81 % 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 3.9957e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: 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	0.20	0/2680	0.35	0/3655
1	B	0.20	0/2595	0.34	0/3551
1	C	0.20	0/2676	0.35	0/3652
1	D	0.20	0/2700	0.35	0/3684
All	All	0.20	0/10651	0.35	0/14542

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	2616	0	2436	11	0
1	B	2533	0	2220	14	0
1	C	2612	0	2426	16	0
1	D	2636	0	2496	23	0
2	A	27	0	22	0	0
2	B	27	0	22	1	0
2	C	27	0	22	0	0
2	D	27	0	22	0	0
3	A	62	0	0	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	43	0	0	0	0
3	C	78	0	0	4	0
3	D	31	0	0	0	0
All	All	10719	0	9666	57	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 3.

The worst 5 of 57 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:250:MET:O	1:D:280:ASN:ND2	2.23	0.71
1:C:229:ASP:O	3:C:553:HOH:O	2.11	0.69
1:C:248:GLY:N	3:C:553:HOH:O	2.24	0.67
1:D:306:VAL:HG22	1:D:308:LYS:H	1.63	0.64
1:D:76:ASP:OD2	1:D:101:ARG:NH2	2.30	0.61

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	352/358 (98%)	343 (97%)	9 (3%)	0	100	100
1	B	355/358 (99%)	347 (98%)	8 (2%)	0	100	100
1	C	351/358 (98%)	338 (96%)	13 (4%)	0	100	100
1	D	356/358 (99%)	342 (96%)	13 (4%)	1 (0%)	41	72
All	All	1414/1432 (99%)	1370 (97%)	43 (3%)	1 (0%)	51	81

All (1) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	D	63	PRO

### 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	243/287 (85%)	243 (100%)	0	100	100
1	B	213/287 (74%)	212 (100%)	1 (0%)	88	96
1	C	244/287 (85%)	240 (98%)	4 (2%)	62	88
1	D	255/287 (89%)	253 (99%)	2 (1%)	81	94
All	All	955/1148 (83%)	948 (99%)	7 (1%)	84	95

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	267	HIS
1	D	282	TYR
1	C	303	THR
1	C	81	LEU
1	D	53	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	267	HIS

### 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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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	SAM	D	401	-	21,29,29	1.80	3 (14%)	18,42,42	2.27	5 (27%)
2	SAM	C	401	-	21,29,29	1.79	3 (14%)	18,42,42	2.18	5 (27%)
2	SAM	B	401	-	21,29,29	1.93	3 (14%)	18,42,42	2.41	6 (33%)
2	SAM	A	401	-	21,29,29	1.79	3 (14%)	18,42,42	2.27	6 (33%)

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	SAM	D	401	-	-	3/8/33/33	0/3/3/3
2	SAM	C	401	-	-	5/8/33/33	0/3/3/3
2	SAM	B	401	-	-	3/8/33/33	0/3/3/3
2	SAM	A	401	-	-	3/8/33/33	0/3/3/3

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	401	SAM	C2'-C1'	-5.88	1.44	1.53
2	D	401	SAM	C2'-C1'	-5.76	1.45	1.53
2	C	401	SAM	C2'-C1'	-5.42	1.45	1.53
2	A	401	SAM	C2'-C1'	-5.34	1.45	1.53
2	B	401	SAM	O4'-C1'	4.37	1.47	1.41

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	401	SAM	C5'-SD-CG	6.95	121.13	103.40
2	A	401	SAM	C5'-SD-CG	6.79	120.73	103.40
2	D	401	SAM	C5'-SD-CG	6.61	120.27	103.40
2	C	401	SAM	C5'-SD-CG	6.51	120.01	103.40
2	D	401	SAM	N3-C2-N1	-4.29	121.97	128.68

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

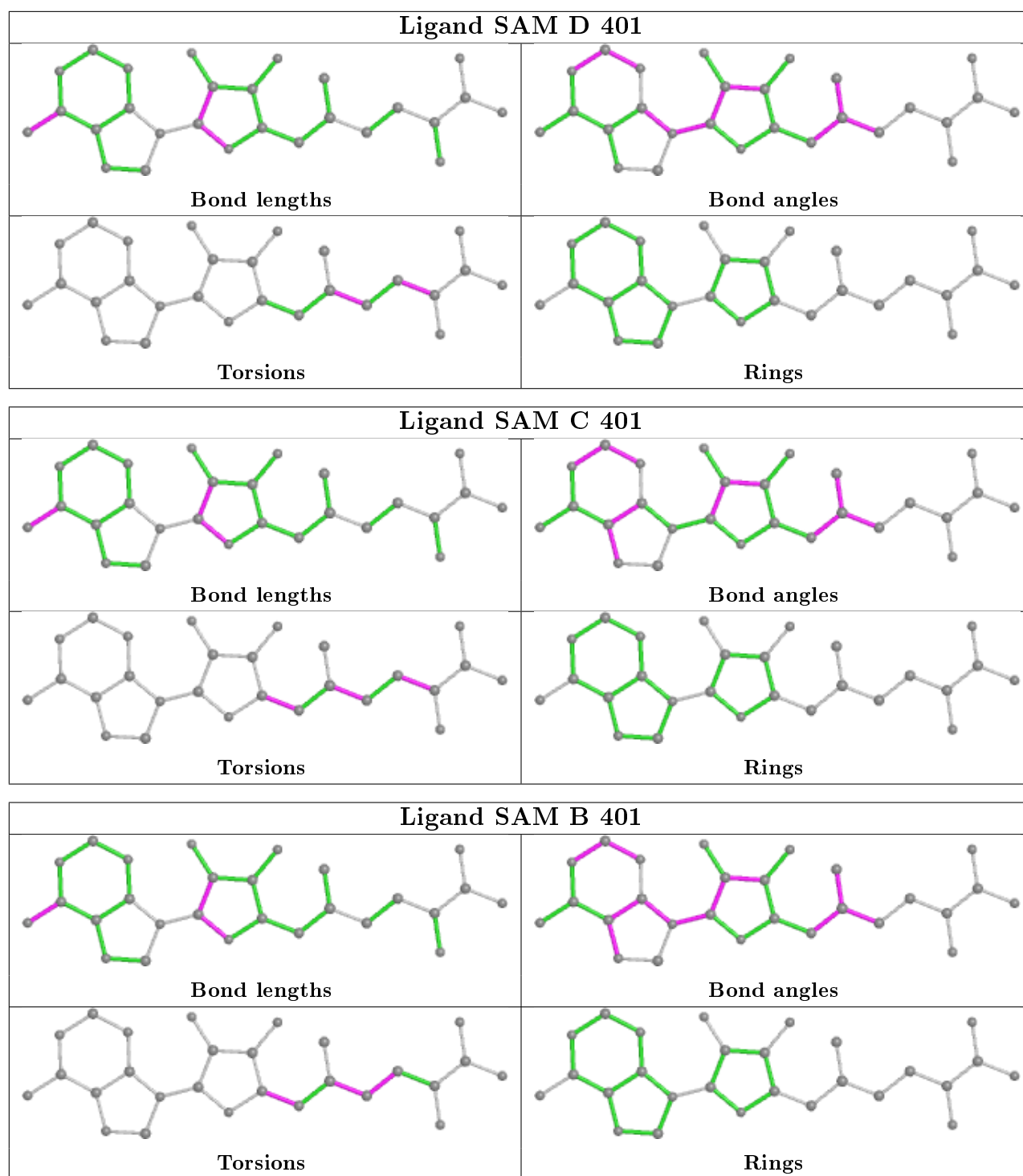
Mol	Chain	Res	Type	Atoms
2	D	401	SAM	N-CA-CB-CG
2	D	401	SAM	C-CA-CB-CG
2	D	401	SAM	CB-CG-SD-CE
2	C	401	SAM	N-CA-CB-CG
2	C	401	SAM	C-CA-CB-CG

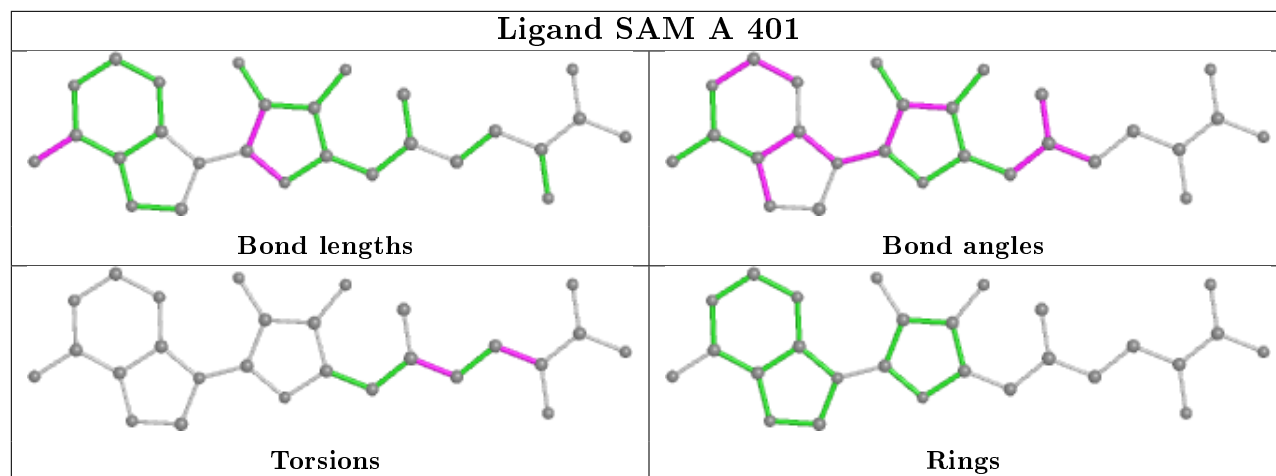
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	401	SAM	1	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	354/358 (98%)	-0.37	0 <span>100</span> <span>100</span>	32, 53, 90, 108	0
1	B	357/358 (99%)	-0.11	5 (1%) <span>75</span> <span>70</span>	29, 69, 103, 130	0
1	C	353/358 (98%)	-0.34	1 (0%) <span>94</span> <span>93</span>	37, 56, 93, 154	0
1	D	358/358 (100%)	0.69	61 (17%) <span>1</span> <span>1</span>	33, 95, 149, 166	0
All	All	1422/1432 (99%)	-0.03	67 (4%) <span>31</span> <span>22</span>	29, 63, 127, 166	0

The worst 5 of 67 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	290	GLY	10.7
1	D	164	GLY	6.6
1	D	204	ASP	6.6
1	D	281	CYS	5.4
1	D	205	VAL	5.1

### 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](#)

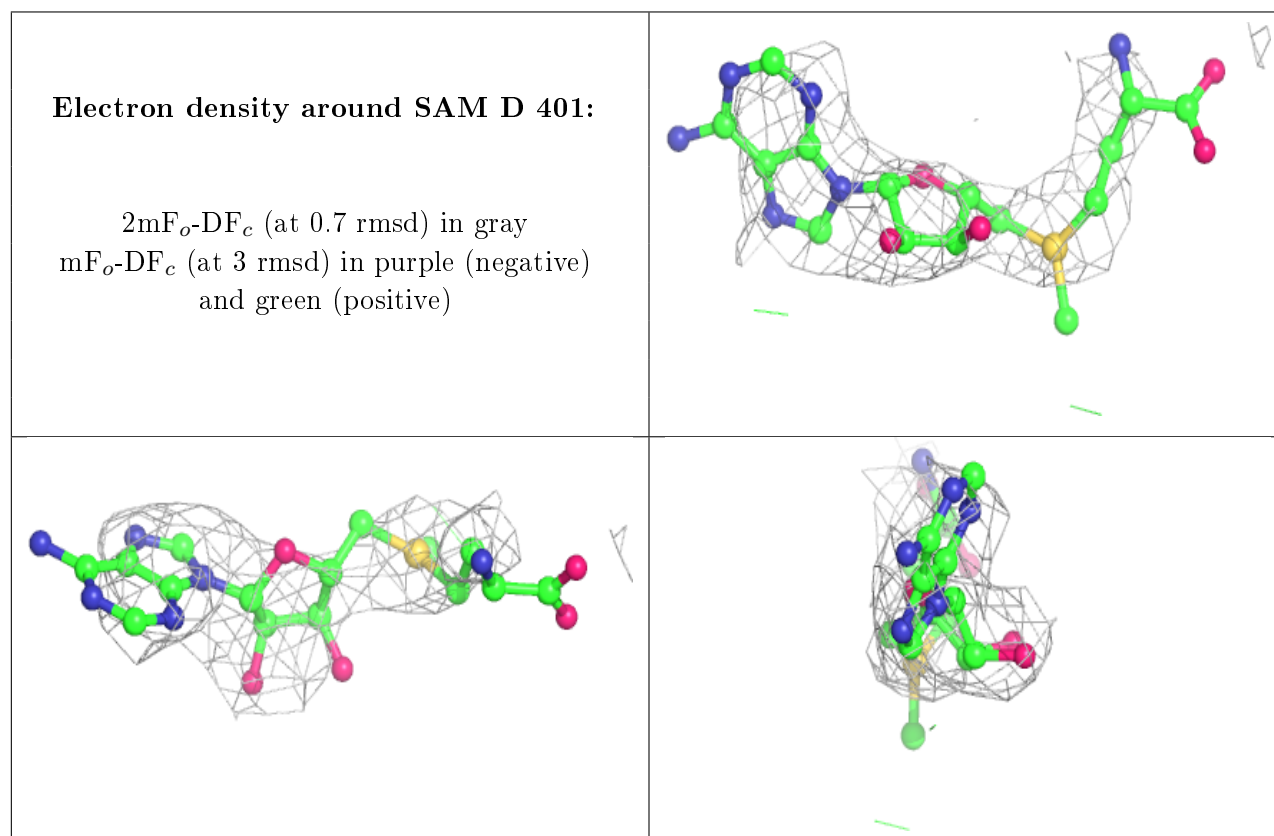
There are no carbohydrates in this entry.

### 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, 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.

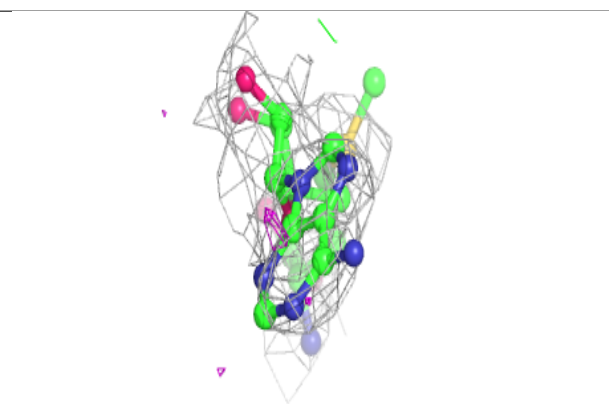
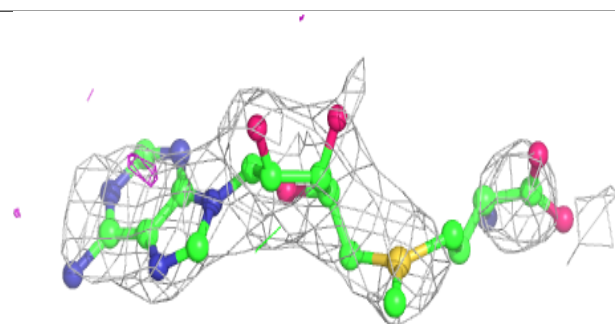
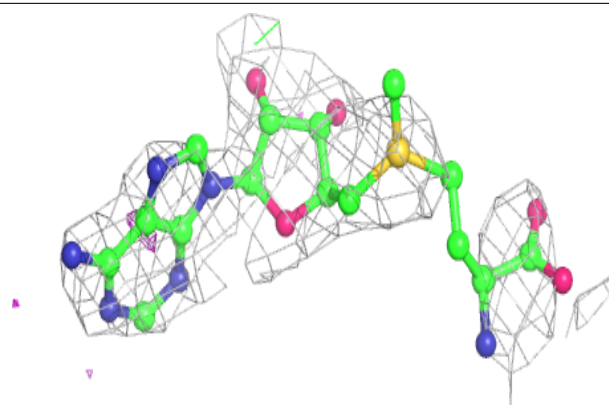
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	SAM	D	401	27/27	0.75	0.39	124,145,169,179	0
2	SAM	B	401	27/27	0.81	0.36	93,104,108,110	0
2	SAM	C	401	27/27	0.95	0.17	48,58,80,90	0
2	SAM	A	401	27/27	0.95	0.18	44,55,77,81	0

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.

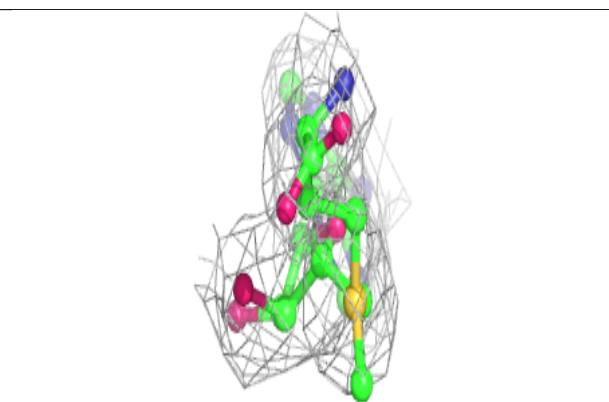
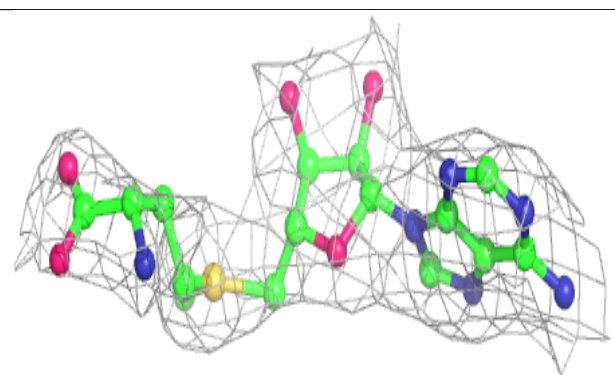
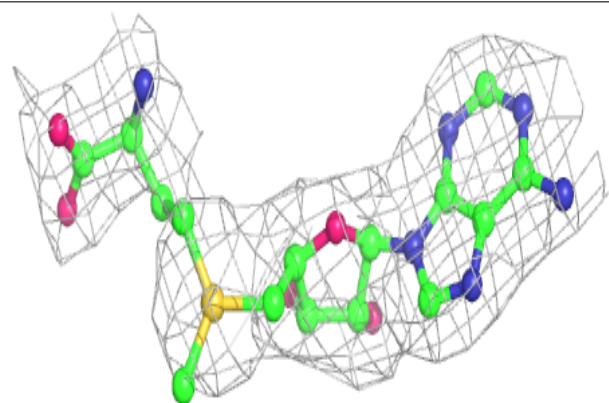


**Electron density around SAM B 401:**

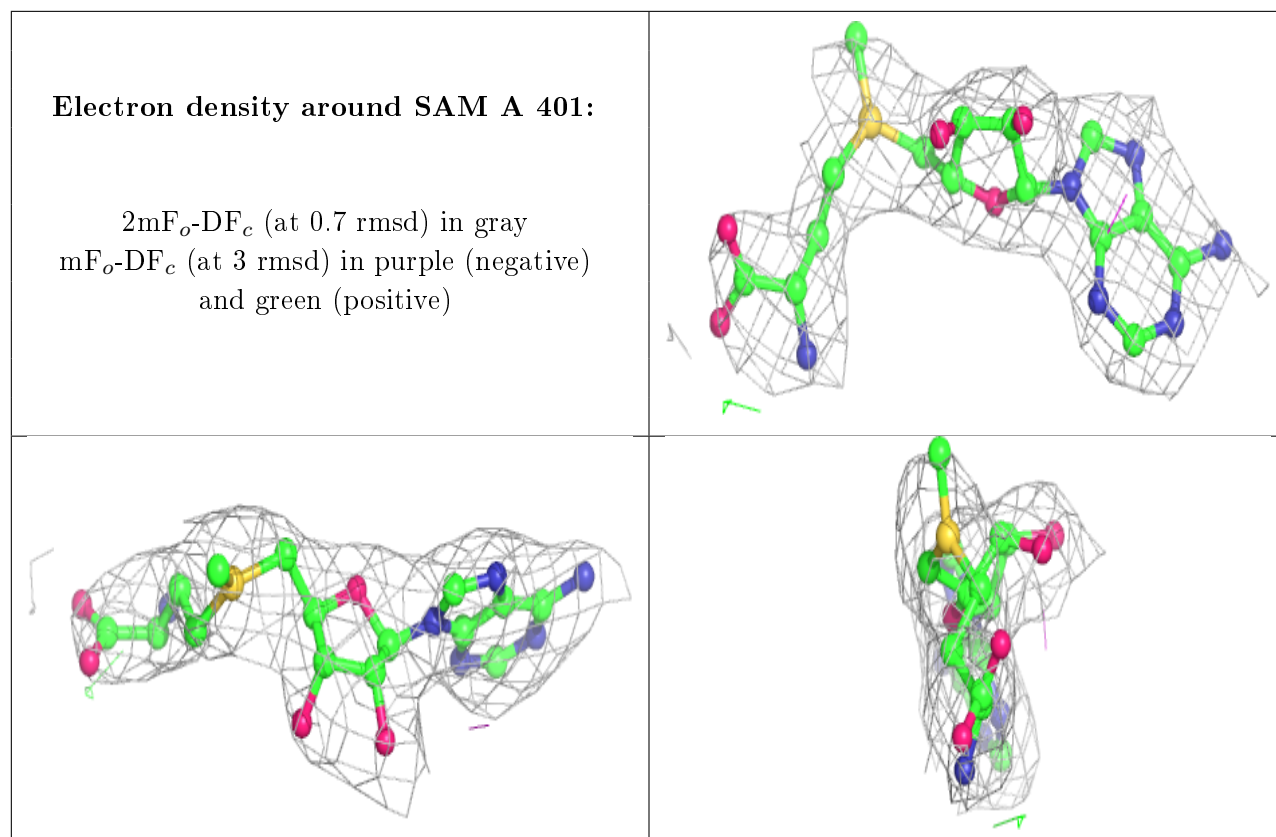
$2mF_o-DF_c$  (at 0.7 rmsd) in gray  
 $mF_o-DF_c$  (at 3 rmsd) in purple (negative)  
and green (positive)

**Electron density around SAM C 401:**

$2mF_o-DF_c$  (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.