



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 4, 2023 – 04:30 PM JST

PDB ID : 8K9Y  
Title : Crystal structure of Arabidopsis thaliana sulfotransferase SOT16 involved in glucosinolate biosynthesis  
Authors : Iwamoto, Y.; Saito, S.; Teramoto, T.; Kakuta, Y.  
Deposited on : 2023-08-02  
Resolution : 1.82 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35  
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

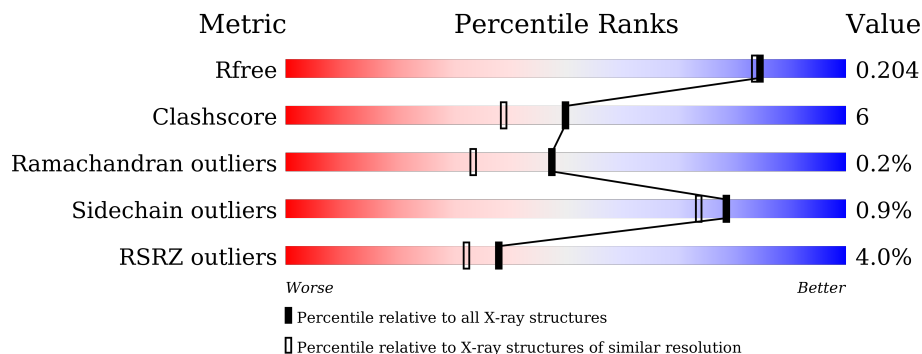
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.82 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	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RSRZ outliers	127900	7371 (1.84-1.80)

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 of 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	323	 87% 12% .
1	B	323	 7% 90% 9% ..

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
3	GOL	B	403	-	X	-	-

## 2 Entry composition [i](#)

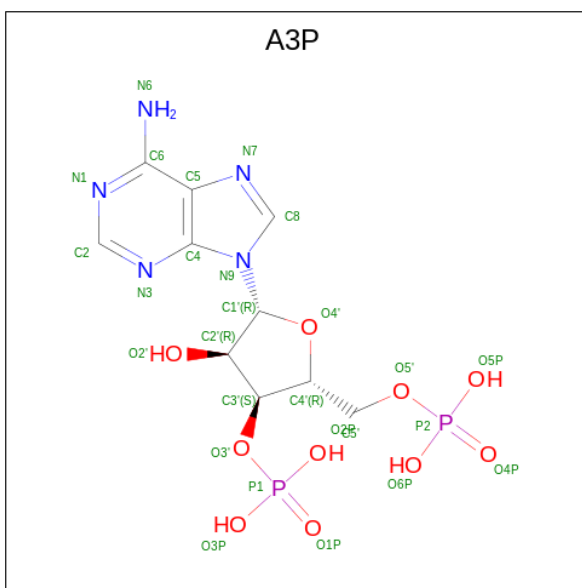
There are 6 unique types of molecules in this entry. The entry contains 6180 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Cytosolic sulfotransferase 16.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	322	Total	C	N	O	S	0	5	0
			2688	1732	448	497	11			
1	B	321	Total	C	N	O	S	0	0	0
			2642	1705	438	488	11			

- Molecule 2 is ADENOSINE-3'-5'-DIPHOSPHATE (three-letter code: A3P) (formula:  $C_{10}H_{15}N_5O_{10}P_2$ ).



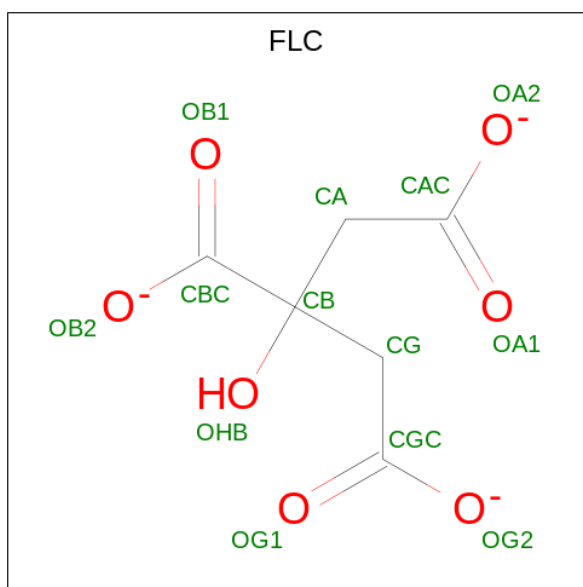
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
2	A	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
2	B	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	B	1	Total	C	O	0	0
			6	3	3		
3	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 4 is CITRATE ANION (three-letter code: FLC) (formula:  $C_6H_5O_7$ ).



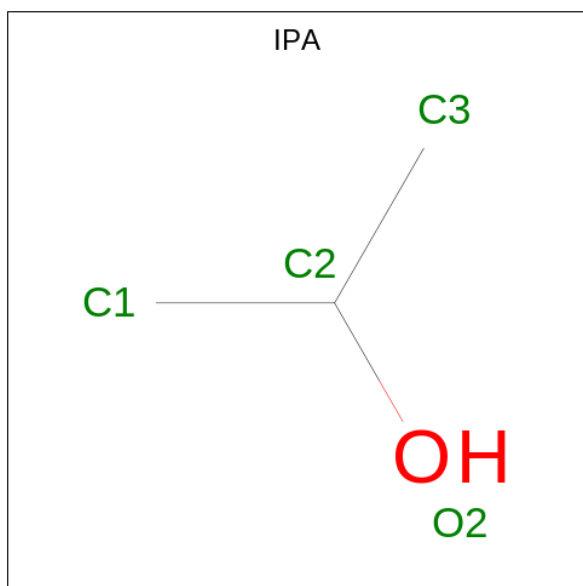
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			13	6	7		

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	O	0	0
			13	6	7		

- Molecule 5 is ISOPROPYL ALCOHOL (three-letter code: IPA) (formula: C<sub>3</sub>H<sub>8</sub>O).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	O	0	0
			4	3	1		

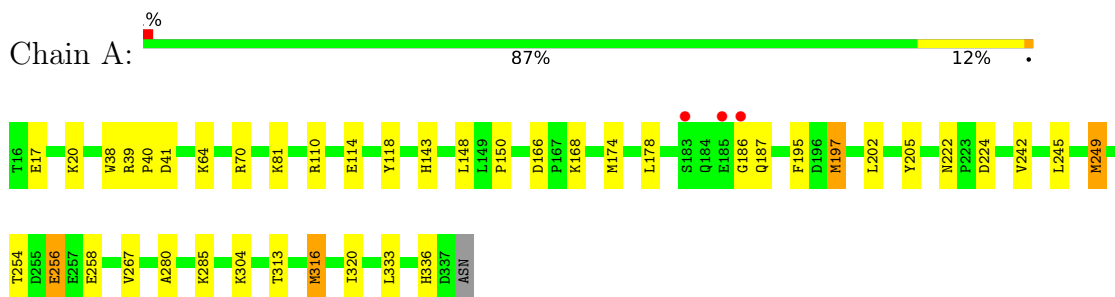
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	432	Total	O	0	0
			432	432		
6	B	316	Total	O	0	0
			316	316		

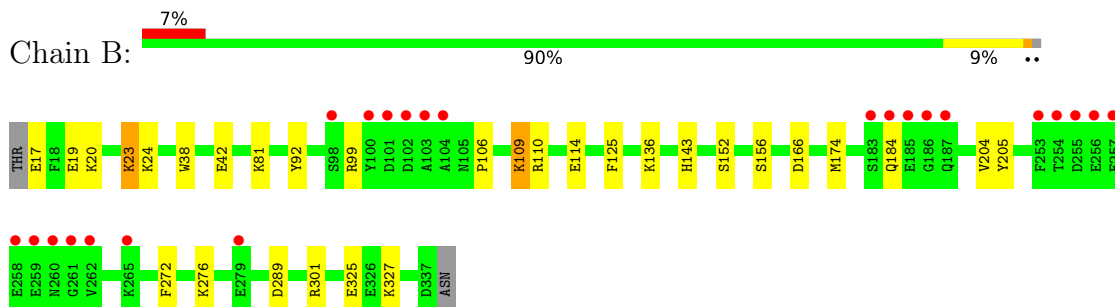
### 3 Residue-property plots [i](#)

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: Cytosolic sulfotransferase 16



- Molecule 1: Cytosolic sulfotransferase 16



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.43Å 138.24Å 63.13Å 90.00° 104.88° 90.00°	Depositor
Resolution (Å)	47.58 – 1.82 47.58 – 1.82	Depositor EDS
% Data completeness (in resolution range)	94.1 (47.58-1.82) 94.1 (47.58-1.82)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.20 (at 1.82Å)	Xtrriage
Refinement program	PHENIX (1.19.2_4158: ???)	Depositor
R, $R_{free}$	0.163 , 0.204 0.163 , 0.204	Depositor DCC
$R_{free}$ test set	3651 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	26.1	Xtrriage
Anisotropy	0.030	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 50.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	6180	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.78% of the height of the origin peak. No significant pseudotranslation is detected.*

<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: IPA, GOL, A3P, FLC

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	0.90	0/2766	0.85	3/3747 (0.1%)
1	B	0.80	0/2720	0.78	1/3685 (0.0%)
All	All	0.85	0/5486	0.82	4/7432 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	197	MET	CA-CB-CG	8.55	127.83	113.30
1	A	249	MET	N-CA-CB	-7.07	97.87	110.60
1	B	166	ASP	CB-CG-OD1	6.70	124.33	118.30
1	A	166	ASP	CB-CG-OD1	6.05	123.75	118.30

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	2688	0	2593	36	0
1	B	2642	0	2549	24	0
2	A	27	0	11	0	0
2	B	27	0	11	1	0
3	A	6	0	8	3	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	12	0	15	4	0
4	A	13	0	5	3	0
4	B	13	0	5	1	0
5	B	4	0	8	0	1
6	A	432	0	0	11	4
6	B	316	0	0	7	3
All	All	6180	0	5205	62	4

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 6.

The worst 5 of 62 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:23:LYS:NZ	6:B:502:HOH:O	2.10	0.77
1:B:23:LYS:HD2	1:B:24:LYS:N	1.99	0.77
1:B:174:MET:HG2	6:B:737:HOH:O	1.87	0.74
1:A:285:LYS:NZ	6:A:501:HOH:O	2.11	0.74
1:A:197:MET:HB3	1:A:202:LEU:HB3	1.72	0.72

All (4) 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)
5:B:402:IPA:O2	6:A:742:HOH:O[2_455]	2.00	0.20
6:A:837:HOH:O	6:B:789:HOH:O[2_446]	2.07	0.13
6:A:543:HOH:O	6:B:726:HOH:O[2_446]	2.09	0.11
6:A:837:HOH:O	6:B:679:HOH:O[2_446]	2.13	0.07

## 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	325/323 (101%)	318 (98%)	7 (2%)	0	100	100
1	B	319/323 (99%)	308 (97%)	10 (3%)	1 (0%)	41	27
All	All	644/646 (100%)	626 (97%)	17 (3%)	1 (0%)	47	33

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	184	GLN

### 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	287/284 (101%)	285 (99%)	2 (1%)	84	80
1	B	282/284 (99%)	279 (99%)	3 (1%)	73	67
All	All	569/568 (100%)	564 (99%)	5 (1%)	78	74

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	256	GLU
1	A	316	MET
1	B	23	LYS
1	B	109	LYS
1	B	125	PHE

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

Mol	Chain	Res	Type
1	A	187	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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

8 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
5	IPA	B	402	-	3,3,3	0.66	0	3,3,3	0.69	0
4	FLC	A	403	-	12,12,12	1.86	3 (25%)	17,17,17	2.13	5 (29%)
3	GOL	B	404	-	5,5,5	1.18	0	5,5,5	1.46	0
4	FLC	B	405	-	12,12,12	1.59	3 (25%)	17,17,17	2.89	7 (41%)
2	A3P	A	401	-	26,29,29	1.43	4 (15%)	31,45,45	1.19	3 (9%)
3	GOL	A	402	-	5,5,5	1.43	1 (20%)	5,5,5	1.83	1 (20%)
2	A3P	B	401	-	26,29,29	1.08	1 (3%)	31,45,45	0.93	1 (3%)
3	GOL	B	403	-	5,5,5	1.39	1 (20%)	5,5,5	1.44	1 (20%)

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	FLC	A	403	-	-	2/16/16/16	-

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	B	404	-	-	2/4/4/4	-
4	FLC	B	405	-	-	3/16/16/16	-
2	A3P	A	401	-	-	0/11/31/31	0/3/3/3
3	GOL	A	402	-	-	2/4/4/4	-
2	A3P	B	401	-	-	2/11/31/31	0/3/3/3
3	GOL	B	403	-	-	4/4/4/4	-

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	401	A3P	O4'-C1'	3.59	1.46	1.41
4	A	403	FLC	OA1-CAC	3.36	1.33	1.22
4	A	403	FLC	OG1-CGC	3.27	1.33	1.22
4	A	403	FLC	OG2-CGC	-3.24	1.19	1.30
2	B	401	A3P	P1-O3'	3.09	1.65	1.59

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	405	FLC	OB1-CBC-CB	-7.69	111.37	122.25
4	B	405	FLC	OB2-CBC-CB	5.72	122.98	113.05
4	A	403	FLC	CG-CB-CA	-4.68	96.94	109.16
4	A	403	FLC	OHB-CB-CA	4.50	119.92	109.40
4	A	403	FLC	OA1-CAC-CA	-3.78	111.89	122.94

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	402	GOL	O1-C1-C2-C3
3	B	404	GOL	O1-C1-C2-C3
4	B	405	FLC	CA-CB-CG-CGC
4	B	405	FLC	CBC-CB-CG-CGC
4	B	405	FLC	OHB-CB-CG-CGC

There are no ring outliers.

7 monomers are involved in 13 short contacts:

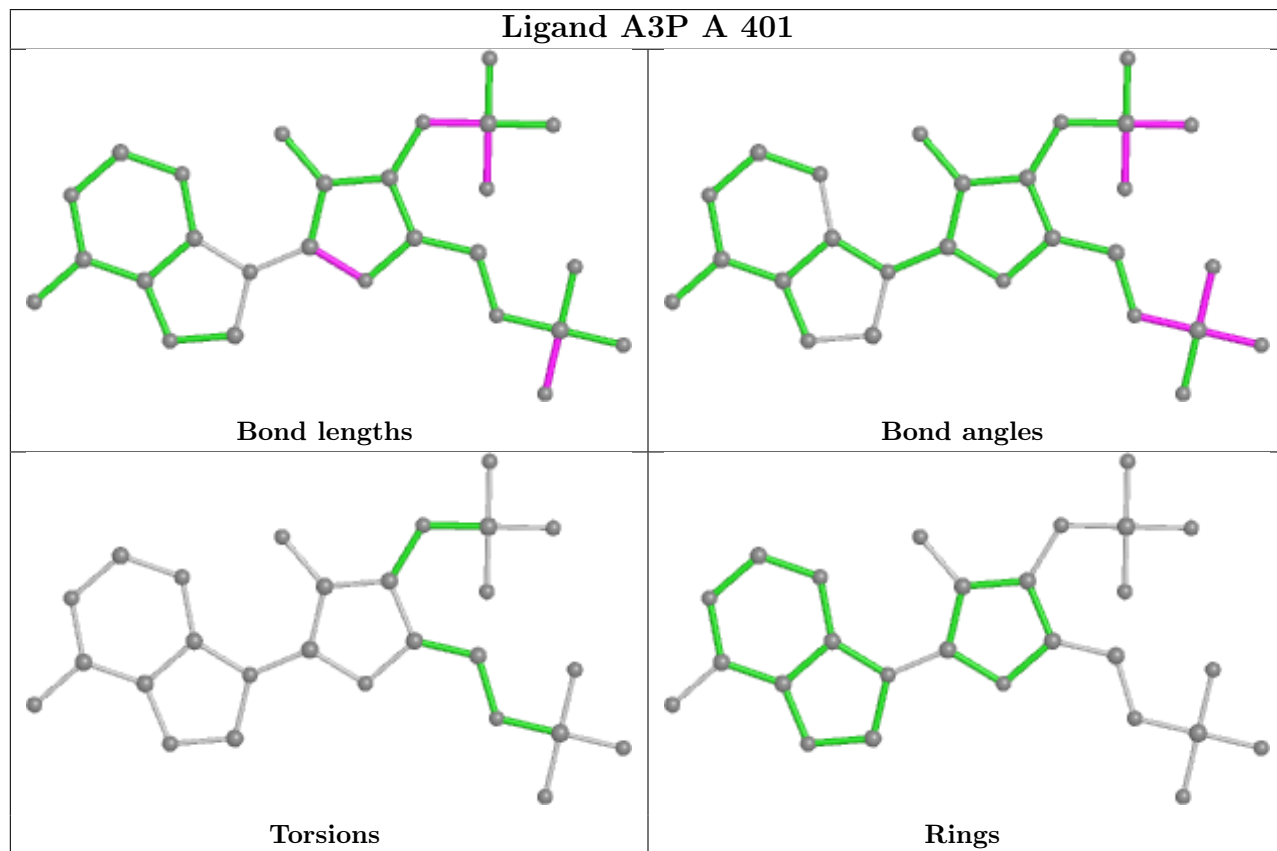
Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	B	402	IPA	0	1

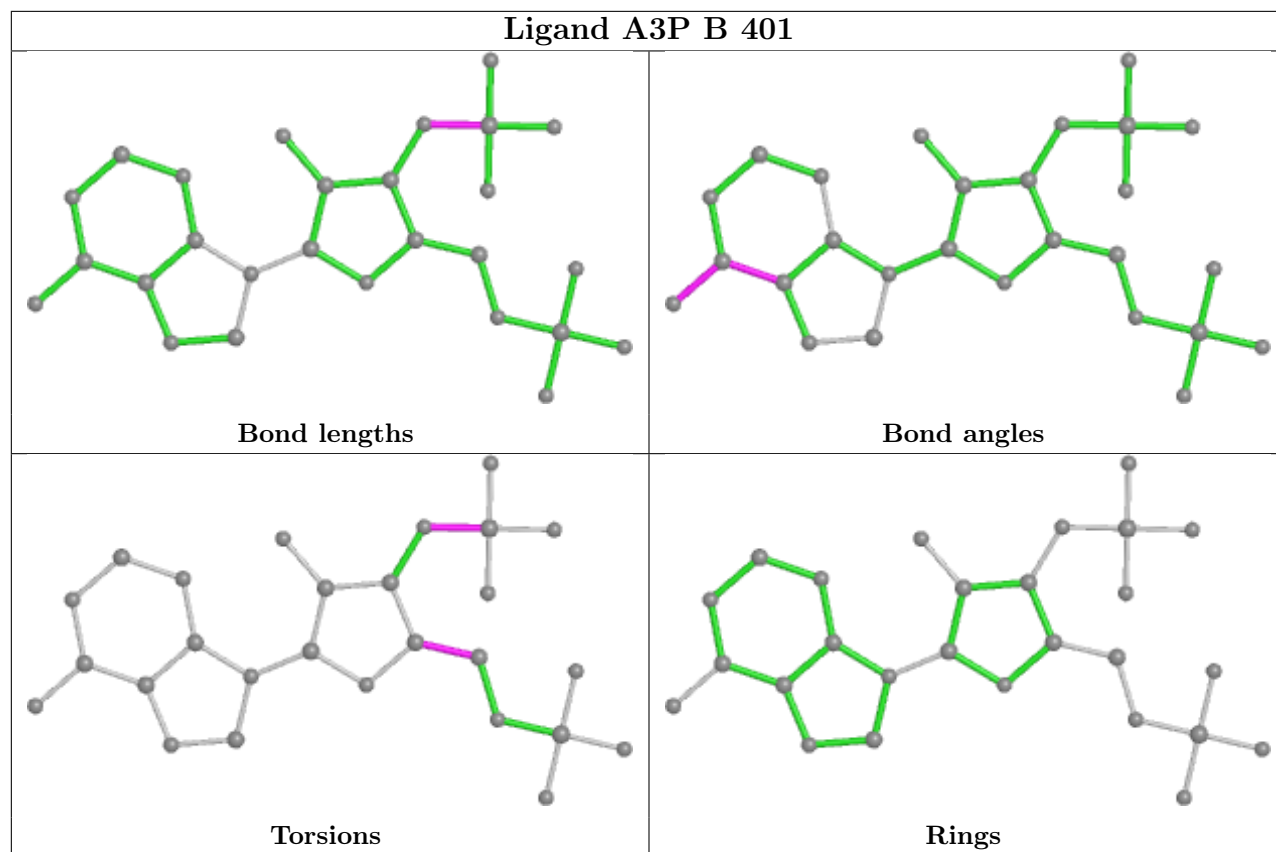
*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	403	FLC	3	0
3	B	404	GOL	1	0
4	B	405	FLC	1	0
3	A	402	GOL	3	0
2	B	401	A3P	1	0
3	B	403	GOL	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 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	322/323 (99%)	-0.19	3 (0%) 84 82	15, 27, 46, 65	0
1	B	321/323 (99%)	-0.04	23 (7%) 15 12	15, 31, 58, 90	0
All	All	643/646 (99%)	-0.11	26 (4%) 38 32	15, 29, 53, 90	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	183	SER	5.1
1	B	185	GLU	4.7
1	B	254	THR	4.6
1	B	262	VAL	4.6
1	B	255	ASP	4.3

### 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 monosaccharides in this entry.

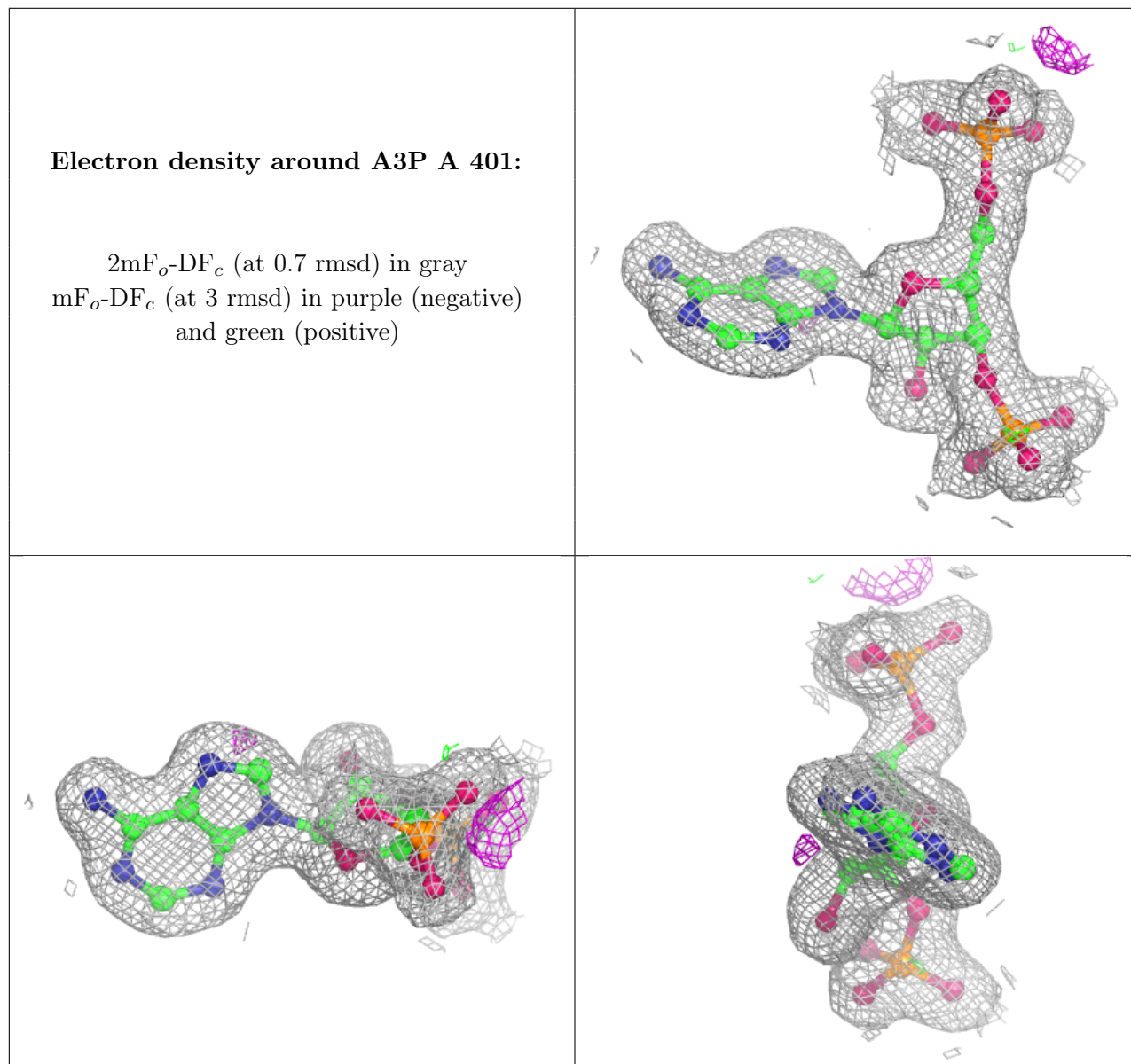
### 6.4 Ligands [i](#)

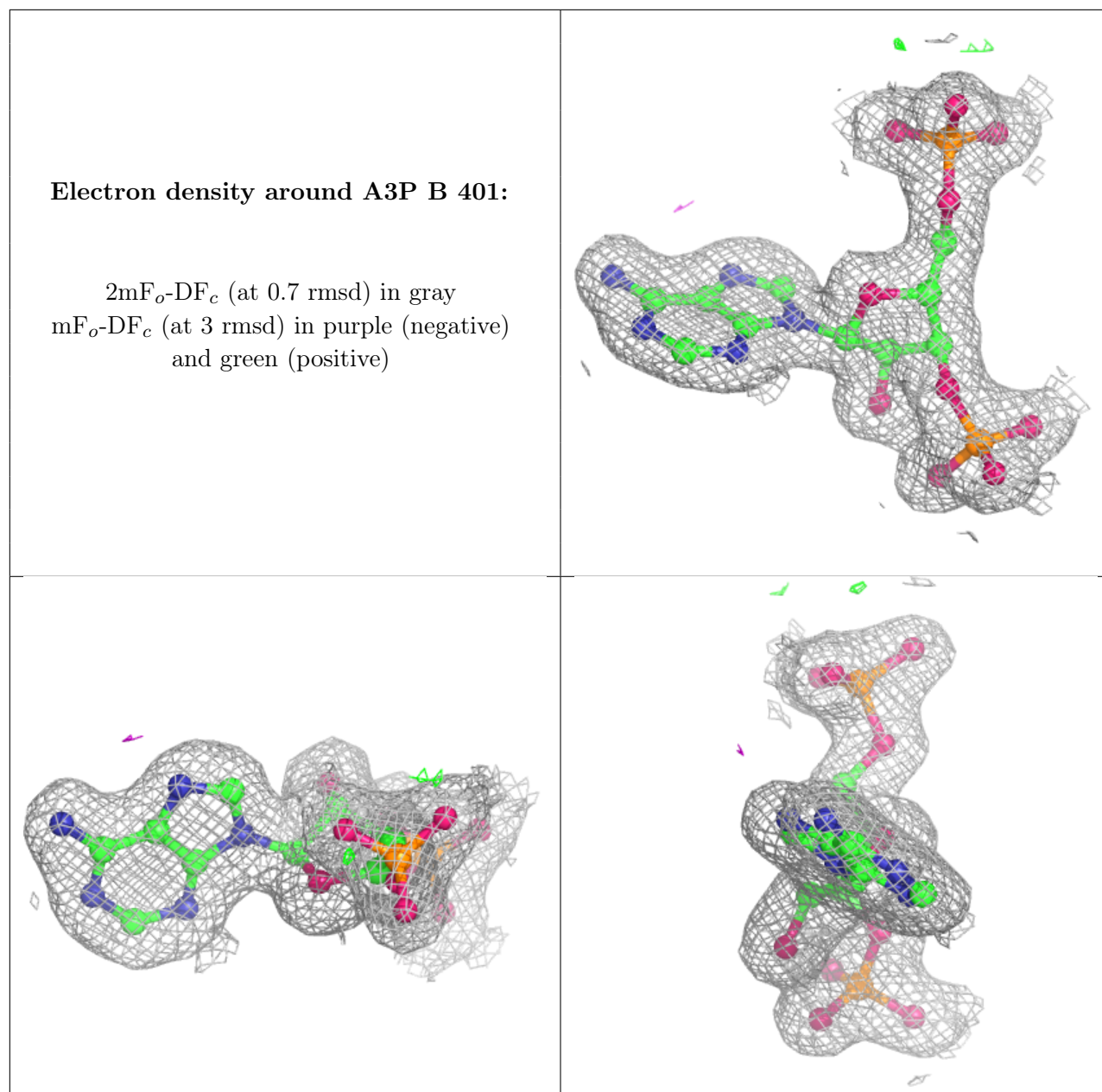
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q < 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
4	FLC	A	403	13/13	0.87	0.17	30,39,57,70	0
4	FLC	B	405	13/13	0.89	0.24	30,47,74,91	0
3	GOL	A	402	6/6	0.91	0.10	32,37,40,40	0
3	GOL	B	403	6/6	0.93	0.10	28,33,38,38	0
5	IPA	B	402	4/4	0.94	0.14	31,38,41,44	0
3	GOL	B	404	6/6	0.95	0.12	31,40,47,48	0
2	A3P	A	401	27/27	0.98	0.09	16,19,22,24	0
2	A3P	B	401	27/27	0.99	0.08	20,23,29,29	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.





## 6.5 Other polymers [i](#)

There are no such residues in this entry.