



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 21, 2023 – 10:07 AM JST

PDB ID : 8I4S  
Title : the complex structure of SARS-CoV-2 Mpro with D8  
Authors : Lu, M.  
Deposited on : 2023-01-21  
Resolution : 2.20 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](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 : 2.36  
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) : 2.36

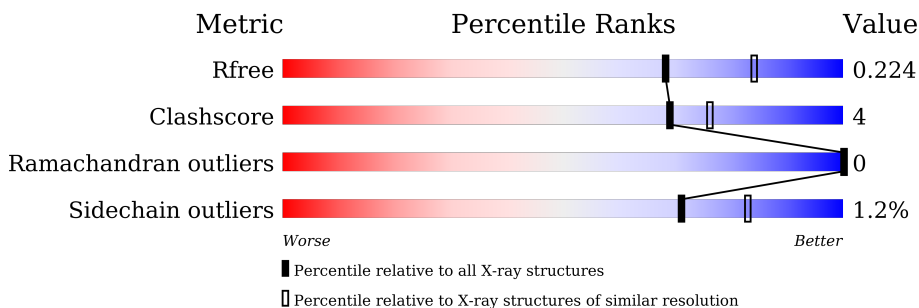
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	335	 81%      8% • 10%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2501 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 ORF1a polyprotein.

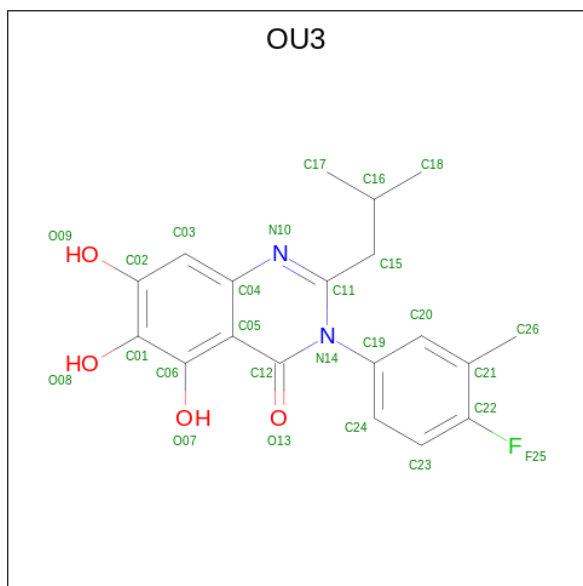
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	300	2323	1471	395	435	22	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	MET	-	initiating methionine	UNP A0A8B1JQ63
A	-17	GLY	-	expression tag	UNP A0A8B1JQ63
A	-16	SER	-	expression tag	UNP A0A8B1JQ63
A	-15	SER	-	expression tag	UNP A0A8B1JQ63
A	-14	HIS	-	expression tag	UNP A0A8B1JQ63
A	-13	HIS	-	expression tag	UNP A0A8B1JQ63
A	-12	HIS	-	expression tag	UNP A0A8B1JQ63
A	-11	HIS	-	expression tag	UNP A0A8B1JQ63
A	-10	HIS	-	expression tag	UNP A0A8B1JQ63
A	-9	HIS	-	expression tag	UNP A0A8B1JQ63
A	-8	SER	-	expression tag	UNP A0A8B1JQ63
A	-7	SER	-	expression tag	UNP A0A8B1JQ63
A	-6	GLY	-	expression tag	UNP A0A8B1JQ63
A	-5	MET	-	expression tag	UNP A0A8B1JQ63
A	307	GLY	-	expression tag	UNP A0A8B1JQ63
A	308	PRO	-	expression tag	UNP A0A8B1JQ63
A	309	LEU	-	expression tag	UNP A0A8B1JQ63
A	310	GLU	-	expression tag	UNP A0A8B1JQ63
A	311	HIS	-	expression tag	UNP A0A8B1JQ63
A	312	HIS	-	expression tag	UNP A0A8B1JQ63
A	313	HIS	-	expression tag	UNP A0A8B1JQ63
A	314	HIS	-	expression tag	UNP A0A8B1JQ63
A	315	HIS	-	expression tag	UNP A0A8B1JQ63
A	316	HIS	-	expression tag	UNP A0A8B1JQ63

- Molecule 2 is 3-(4-fluoranyl-3-methyl-phenyl)-2-(2-methylpropyl)-5,6,7-tris(oxidanyl)quinazolin-4-one (three-letter code: OU3) (formula: C<sub>19</sub>H<sub>19</sub>FN<sub>2</sub>O<sub>4</sub>) (labeled as "Ligand of

Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	F	N	O		
2	A	1	26	19	1	2	4	0	0


- Molecule 3 is water.

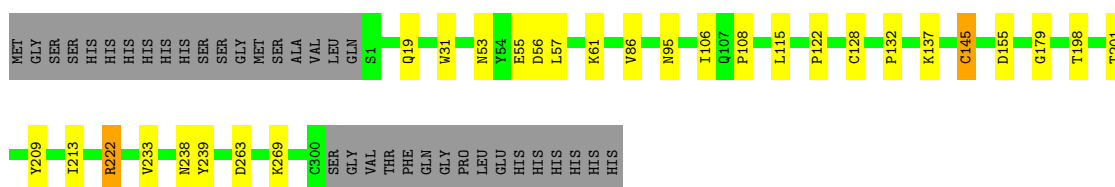
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	152	Total	O	0	0
			152	152		

### 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: ORF1a polypeptide

Chain A:  81% 8% 10%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	98.36Å 81.33Å 51.67Å 90.00° 114.74° 90.00°	Depositor
Resolution (Å)	42.41 – 2.20 44.67 – 1.69	Depositor EDS
% Data completeness (in resolution range)	96.8 (42.41-2.20) 66.9 (44.67-1.69)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.39 (at 1.69Å)	Xtrriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, $R_{free}$	0.185 , 0.224 0.186 , 0.224	Depositor DCC
$R_{free}$ test set	1915 reflections (5.85%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	17.0	Xtrriage
Anisotropy	0.415	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 30.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2501	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 7.41% 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: OU3

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	0.45	2/2375 (0.1%)	0.62	1/3228 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	128	CYS	CB-SG	-5.90	1.72	1.81
1	A	145	CYS	CB-SG	-5.20	1.73	1.81

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	106	ILE	C-N-CA	5.04	134.29	121.70

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	2323	0	2272	17	0
2	A	26	0	0	1	0
3	A	152	0	0	2	0
All	All	2501	0	2272	17	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 4.

All (17) 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:A:137:LYS:NZ	3:A:501:HOH:O	2.26	0.67
1:A:53:ASN:O	1:A:56:ASP:HB2	2.05	0.57
1:A:155:ASP:OD1	1:A:155:ASP:N	2.38	0.53
1:A:86:VAL:HG13	1:A:179:GLY:HA2	1.91	0.52
1:A:222:ARG:O	1:A:222:ARG:HG2	2.10	0.50
1:A:198:THR:HG22	1:A:238:ASN:OD1	2.12	0.49
1:A:115:LEU:HD11	1:A:122:PRO:HB3	1.94	0.49
1:A:145:CYS:SG	2:A:401:OU3:C01	3.03	0.46
1:A:31:TRP:CE2	1:A:95:ASN:HB2	2.51	0.46
1:A:233:VAL:HG11	1:A:269:LYS:HG3	1.98	0.46
1:A:19:GLN:HG2	3:A:514:HOH:O	2.17	0.45
1:A:31:TRP:CD2	1:A:95:ASN:HB2	2.53	0.44
1:A:57:LEU:O	1:A:61:LYS:HG2	2.19	0.43
1:A:108:PRO:HB3	1:A:132:PRO:HA	2.02	0.42
1:A:209:TYR:O	1:A:213:ILE:HG13	2.21	0.41
1:A:57:LEU:H	1:A:57:LEU:HG	1.70	0.41
1:A:201:THR:OG1	1:A:239:TYR:HB3	2.21	0.41

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	298/335 (89%)	293 (98%)	5 (2%)	0	<a href="#">100</a> <a href="#">100</a>

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	258/288 (90%)	255 (99%)	3 (1%)	71 83

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	55	GLU
1	A	222	ARG
1	A	263	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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](#)

1 ligand is 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	OU3	A	401	-	27,28,28	1.70	4 (14%)	33,42,42	2.15	8 (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
2	OU3	A	401	-	-	0/7/8/8	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	401	OU3	C05-C12	-4.55	1.36	1.47
2	A	401	OU3	C12-N14	-3.56	1.34	1.40
2	A	401	OU3	C04-N10	-3.35	1.33	1.39
2	A	401	OU3	C11-N10	2.94	1.35	1.29

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	401	OU3	C20-C21-C22	8.04	121.18	116.06
2	A	401	OU3	C05-C12-N14	4.57	119.69	114.87
2	A	401	OU3	C20-C19-N14	3.12	122.98	119.12
2	A	401	OU3	C24-C19-N14	-3.05	115.98	119.64
2	A	401	OU3	C19-C20-C21	-2.98	118.29	121.60
2	A	401	OU3	C23-C22-C21	-2.95	120.10	124.08
2	A	401	OU3	C06-C05-C12	2.84	122.02	118.60
2	A	401	OU3	O13-C12-C05	-2.03	120.45	124.25

There are no chirality outliers.

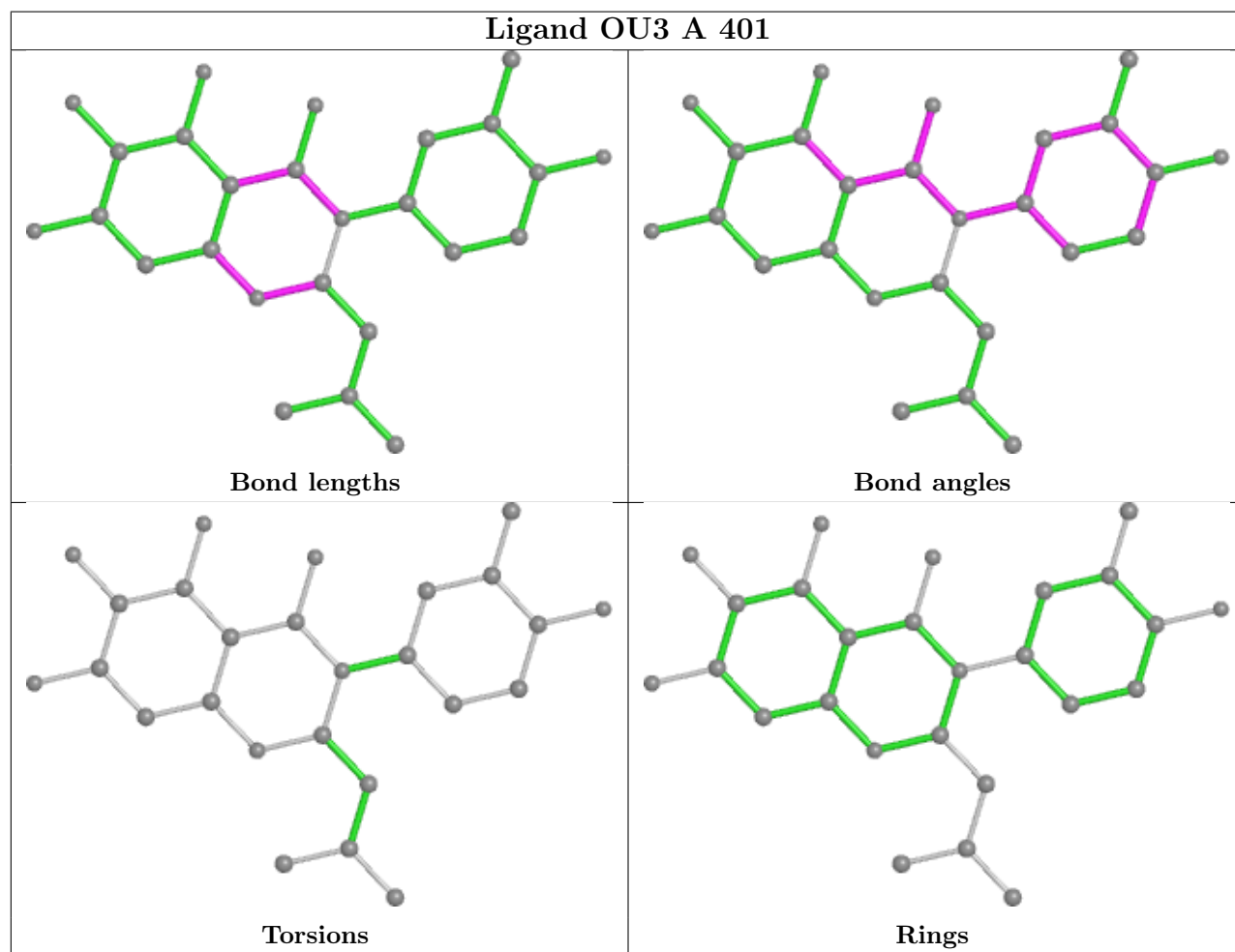
There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	401	OU3	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

There are no chain breaks in this entry.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

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

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

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

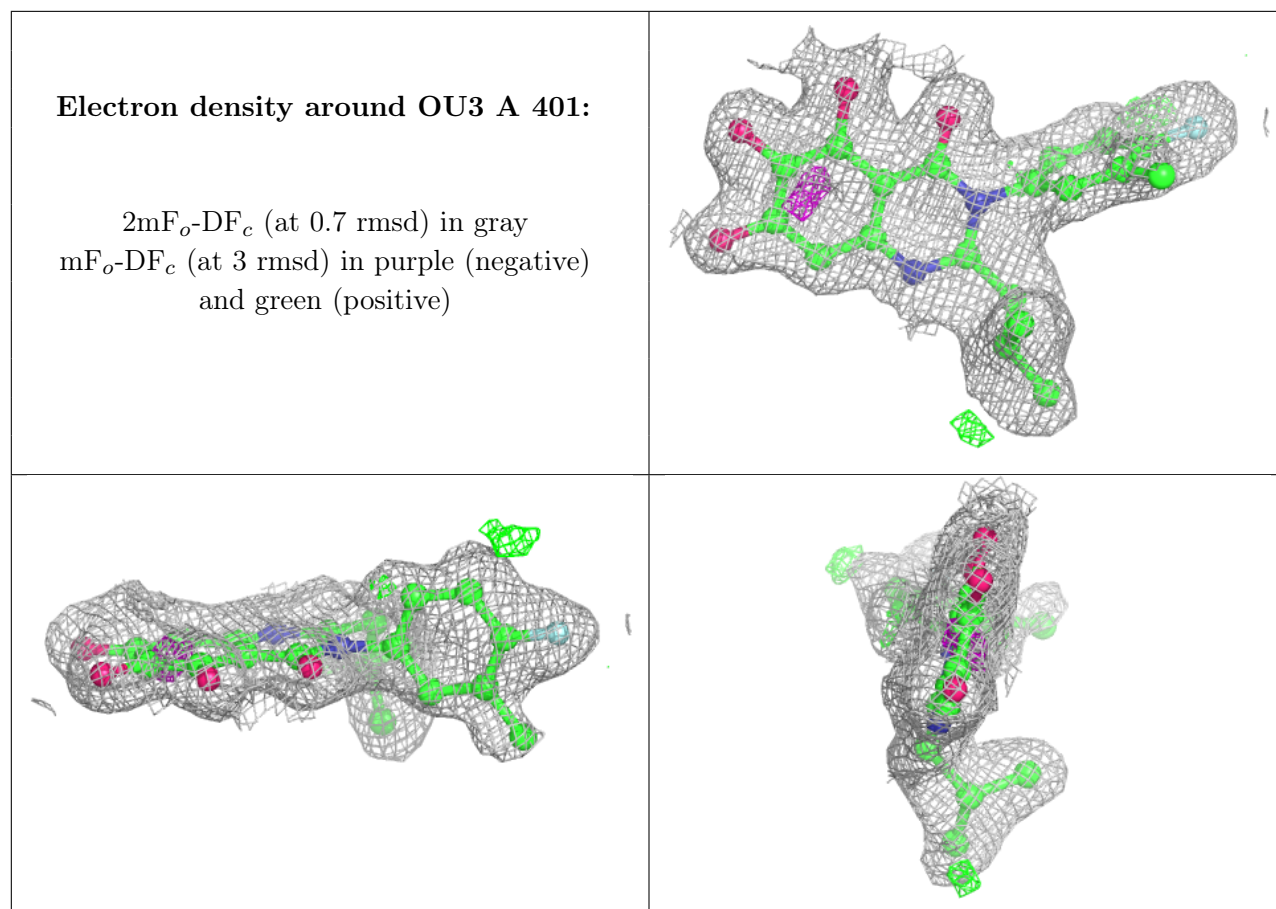
### 6.3 Carbohydrates

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

### 6.4 Ligands

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.



## 6.5 Other polymers [i](#)

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