



# Full wwPDB X-ray Structure Validation Report ⓘ

Jan 20, 2024 – 05:35 pm GMT

PDB ID : 7APH

Title : Structure of SARS-CoV-2 Main Protease bound to Tofogliflozin.

Authors : Guenther, S.; Reinke, P.; Oberthuer, D.; Yefanov, O.; Gelisio, L.; Ginn, H.; Lieske, J.; Domaracky, M.; Brehm, W.; Rahmani Mashour, A.; White, T.A.; Knoska, J.; Pena Esperanza, G.; Koua, F.; Tolstikova, A.; Groessler, M.; Fischer, P.; Hennicke, V.; Fleckenstein, H.; Trost, F.; Galchenkova, M.; Gevorkov, Y.; Li, C.; Awel, S.; Paulraj, L.X.; Ullah, N.; Falke, S.; Alves Franca, B.; Schwinzer, M.; Brognaro, H.; Werner, N.; Perbandt, M.; Tidow, H.; Seychell, B.; Beck, T.; Meier, S.; Doyle, J.J.; Giseler, H.; Melo, D.; Dunkel, I.; Lane, T.J.; Peck, A.; Saouane, S.; Hakanpaeae, J.; Meyer, J.; Noei, H.; Gribbon, P.; Ellinger, B.; Kuzikov, M.; Wolf, M.; Zhang, L.; Ehrt, C.; Pletzer-Zelgert, J.; Wollenhaupt, J.; Feiler, C.; Weiss, M.; Schulz, E.C.; Mehrabi, P.; Norton-Baker, B.; Schmidt, C.; Lorenzen, K.; Schubert, R.; Han, H.; Chari, A.; Fernandez Garcia, Y.; Turk, D.; Hilgenfeld, R.; Rarey, M.; Zaliani, A.; Chapman, H.N.; Pearson, A.; Betzel, C.; Meents, A.

Deposited on : 2020-10-16

Resolution : 1.65 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>.

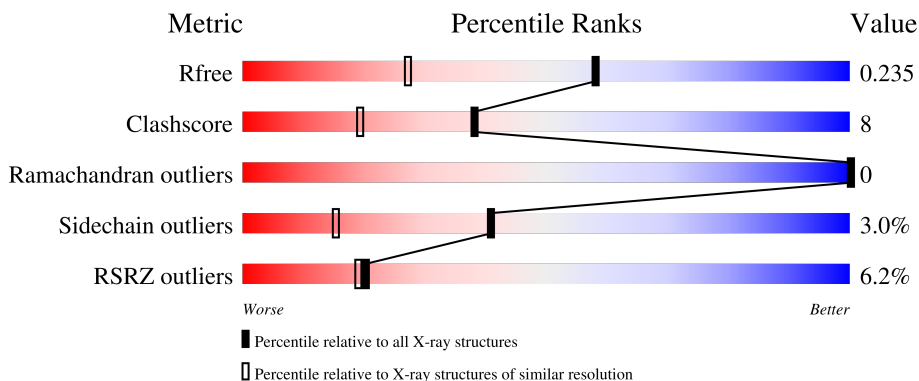
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.65 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	1827 (1.66-1.66)
Clashscore	141614	1931 (1.66-1.66)
Ramachandran outliers	138981	1891 (1.66-1.66)
Sidechain outliers	138945	1891 (1.66-1.66)
RSRZ outliers	127900	1791 (1.66-1.66)

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.

MolProbity	: 4.02b-467
Mogul	: 1.8.4, CSD as541be (2020)
Xtrriage (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

Mol	Chain	Length	Quality of chain
1	AA	306	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into three segments: a red segment on the left labeled '6%', a large green segment in the middle labeled '83%', and a yellow segment on the right labeled '16%'. A small black dot is located at the far right end of the bar.</p>

## 2 Entry composition [i](#)

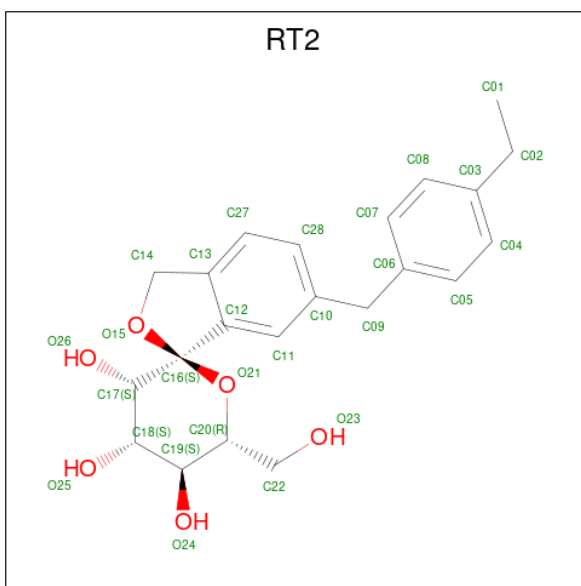
There are 5 unique types of molecules in this entry. The entry contains 6217 atoms, of which 3287 are hydrogens and 0 are deuteriums.

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 3C-like proteinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	AA	306	4935	1572	2447	425	464	27	2464	17	0

- Molecule 2 is Tofogliflozin (three-letter code: RT2) (formula:  $C_{22}H_{26}O_6$ ) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
2	AA	1	54	22	26	6	26	0

- Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula:  $C_2H_6OS$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	H	O	S		
3	AA	1	10	2	6	1	1	6	0
3	AA	1	10	2	6	1	1	6	0
3	AA	1	10	2	6	1	1	6	0
3	AA	1	10	2	6	1	1	6	0
3	AA	1	10	2	6	1	1	6	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	AA	2	Total	Cl	0	0
			2	2		

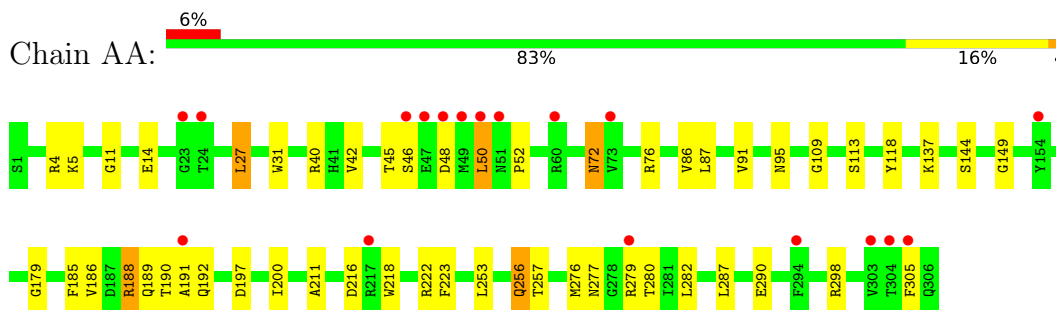
- Molecule 5 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	AA	392	Total	H	O	784	0
			1176	784	392		

### 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: 3C-like proteinase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	113.98Å 53.32Å 44.63Å 90.00° 103.07° 90.00°	Depositor
Resolution (Å)	48.06 – 1.65 48.06 – 1.31	Depositor EDS
% Data completeness (in resolution range)	100.0 (48.06-1.65) 79.6 (48.06-1.31)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.96 (at 1.31Å)	Xtrriage
Refinement program	MAIN	Depositor
R, $R_{free}$	0.234 , 0.266 0.239 , 0.235	Depositor DCC
$R_{free}$ test set	1715 reflections (3.41%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	23.0	Xtrriage
Anisotropy	0.267	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 38.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	6217	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.02% 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: RT2, CL, DMS

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	AA	0.83	0/2566	0.98	0/3478

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	AA	2488	2447	2424	38	2
2	AA	28	26	0	1	0
3	AA	20	30	30	0	0
4	AA	2	0	0	1	0
5	AA	392	784	0	1	0
All	All	2930	3287	2454	38	2

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 8.

All (38) 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:AA:5:LYS:HE3	5:AA:743:HOH:O	1.76	0.85
1:AA:50:LEU:O	1:AA:188:ARG:NH2	2.18	0.75
1:AA:190:THR:HG23	1:AA:192:GLN:HG3	1.77	0.66
1:AA:86:VAL:HG13	1:AA:179:GLY:HA2	1.85	0.58
1:AA:276[A]:MET:HE1	1:AA:280:THR:HA	1.86	0.57
1:AA:5:LYS:HZ2	1:AA:290:GLU:HB2	1.71	0.56
1:AA:256[A]:GLN:HG3	2:AA:401:RT2:O21	2.07	0.55
1:AA:165[A]:MET:CE	1:AA:185:PHE:HB3	2.37	0.55
1:AA:5:LYS:HB3	4:AA:408:CL:CL	2.45	0.53
1:AA:109:GLY:HA2	1:AA:200:ILE:HD13	1.91	0.53
1:AA:190:THR:OG1	1:AA:191:ALA:N	2.41	0.53
1:AA:45:THR:H	1:AA:48:ASP:HB2	1.72	0.52
1:AA:218:TRP:CD1	1:AA:279[B]:ARG:NH2	2.79	0.50
1:AA:190:THR:HG23	1:AA:192:GLN:CG	2.42	0.50
1:AA:165[A]:MET:HE2	1:AA:185:PHE:HB3	1.93	0.49
1:AA:276[B]:MET:HE2	1:AA:287:LEU:HD23	1.94	0.49
1:AA:276[A]:MET:HE1	1:AA:279[A]:ARG:O	2.15	0.47
1:AA:27:LEU:HD21	1:AA:42:VAL:HB	1.97	0.47
1:AA:5:LYS:NZ	1:AA:290:GLU:HB2	2.30	0.47
1:AA:253:LEU:O	1:AA:257:THR:HG23	2.15	0.46
1:AA:31:TRP:CD2	1:AA:95:ASN:HB2	2.51	0.46
1:AA:45:THR:N	1:AA:48:ASP:OD2	2.49	0.46
1:AA:223:PHE:CD1	1:AA:223:PHE:N	2.79	0.46
1:AA:276[B]:MET:CE	1:AA:287:LEU:HD23	2.46	0.46
1:AA:76:ARG:O	1:AA:91:VAL:HA	2.16	0.45
1:AA:40:ARG:HA	1:AA:87:LEU:HG	1.98	0.44
1:AA:48:ASP:O	1:AA:52:PRO:HD3	2.18	0.44
1:AA:211:ALA:HB1	1:AA:216:ASP:HB2	1.99	0.43
1:AA:45:THR:HG23	1:AA:48:ASP:OD2	2.19	0.43
1:AA:137:LYS:CE	1:AA:197:ASP:OD2	2.66	0.43
1:AA:276[B]:MET:HE2	1:AA:287:LEU:CD2	2.49	0.42
1:AA:118:TYR:CE1	1:AA:144:SER:HB3	2.55	0.42
1:AA:282:LEU:HD23	1:AA:282:LEU:HA	1.73	0.41
1:AA:298:ARG:HD3	1:AA:305:PHE:HZ	1.86	0.41
1:AA:218:TRP:HD1	1:AA:279[B]:ARG:NH2	2.19	0.41
1:AA:113:SER:O	1:AA:149:GLY:HA2	2.21	0.41
1:AA:276[B]:MET:O	1:AA:277:ASN:HB2	2.21	0.40
1:AA:165[A]:MET:HE1	1:AA:186:VAL:O	2.22	0.40

All (2) 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)
1:AA:72:ASN:O	1:AA:222:ARG:HH22[1_546]	1.43	0.17
1:AA:11:GLY:H	1:AA:14:GLU:OE2[2_555]	1.58	0.02

### 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	AA	321/306 (105%)	315 (98%)	6 (2%)	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	AA	280/263 (106%)	270 (96%)	10 (4%)	35 11

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AA	4[A]	ARG
1	AA	4[B]	ARG
1	AA	27	LEU
1	AA	46	SER
1	AA	50	LEU
1	AA	72	ASN
1	AA	188	ARG

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	AA	189	GLN
1	AA	256[A]	GLN
1	AA	256[B]	GLN

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

Mol	Chain	Res	Type
1	AA	41	HIS
1	AA	238	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 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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	DMS	AA	406	-	3,3,3	0.56	0	3,3,3	0.72	0
3	DMS	AA	404	-	3,3,3	0.14	0	3,3,3	0.58	0
2	RT2	AA	401	-	29,31,31	1.28	4 (13%)	34,46,46	1.30	2 (5%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	DMS	AA	402	-	3,3,3	1.34	0	3,3,3	0.42	0
3	DMS	AA	405	-	3,3,3	0.70	0	3,3,3	0.12	0
3	DMS	AA	407	-	3,3,3	0.26	0	3,3,3	0.41	0

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	RT2	AA	401	-	-	0/8/43/43	0/4/4/4

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	AA	401	RT2	C13-C12	-3.88	1.34	1.39
2	AA	401	RT2	C16-C12	2.77	1.54	1.51
2	AA	401	RT2	C14-C13	2.54	1.53	1.50
2	AA	401	RT2	C11-C12	2.30	1.42	1.39

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	AA	401	RT2	C14-C13-C12	-4.59	106.54	108.80
2	AA	401	RT2	O21-C20-C22	3.45	111.27	106.56

There are no chirality outliers.

There are no torsion outliers.

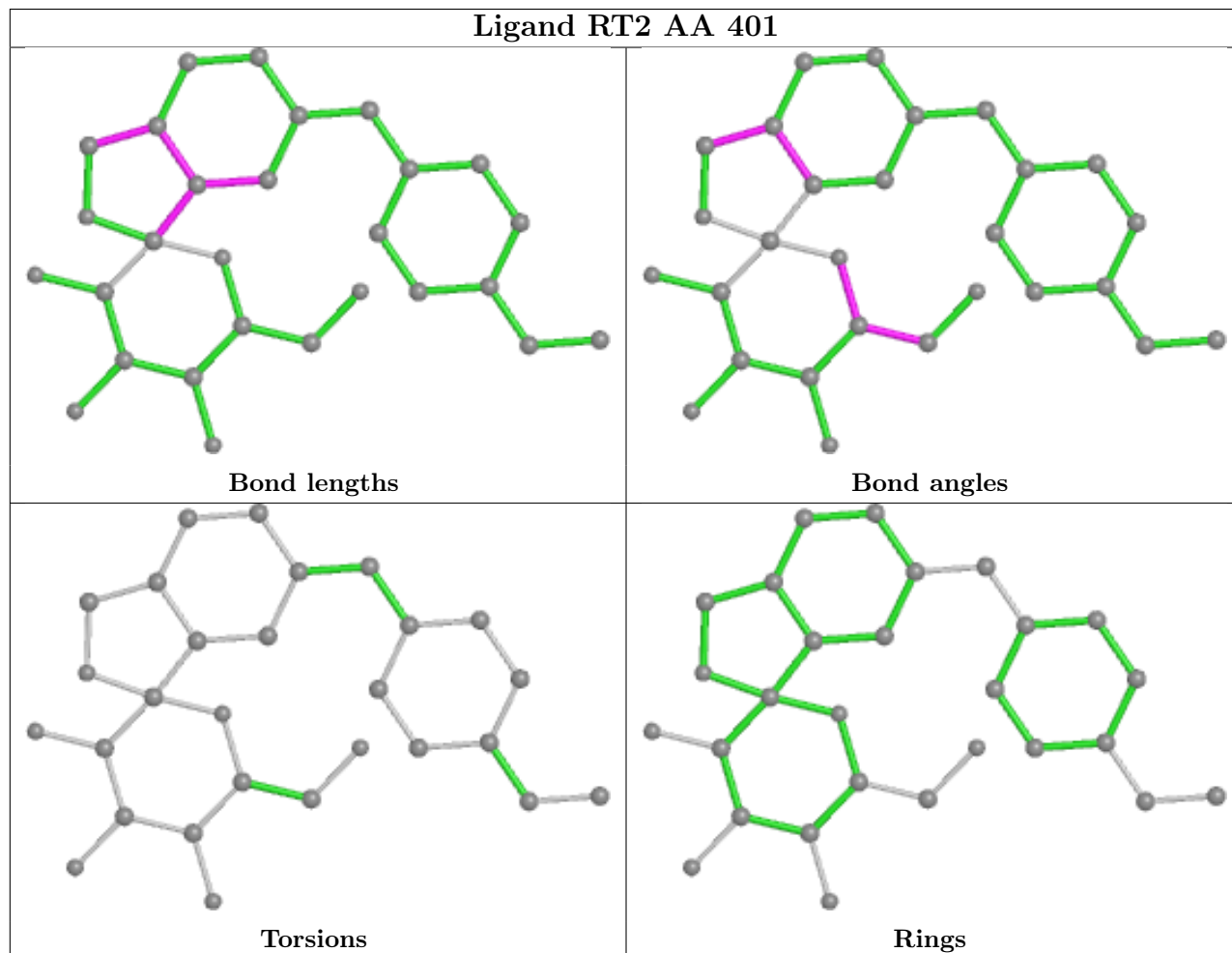
There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	AA	401	RT2	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	AA	305/306 (99%)	0.24	19 (6%) <b>20</b> <b>19</b>	16, 26, 56, 160	2 (0%)

All (19) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AA	60	ARG	5.8
1	AA	46	SER	5.1
1	AA	294[A]	PHE	4.8
1	AA	50	LEU	4.3
1	AA	49[A]	MET	3.8
1	AA	47	GLU	3.5
1	AA	305	PHE	3.5
1	AA	168	PRO	3.4
1	AA	154	TYR	3.3
1	AA	304	THR	3.1
1	AA	279[A]	ARG	3.1
1	AA	48	ASP	3.1
1	AA	303	VAL	3.0
1	AA	24	THR	2.7
1	AA	73[A]	VAL	2.6
1	AA	191	ALA	2.5
1	AA	23	GLY	2.5
1	AA	51	ASN	2.3
1	AA	217[A]	ARG	2.2

### 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](#)

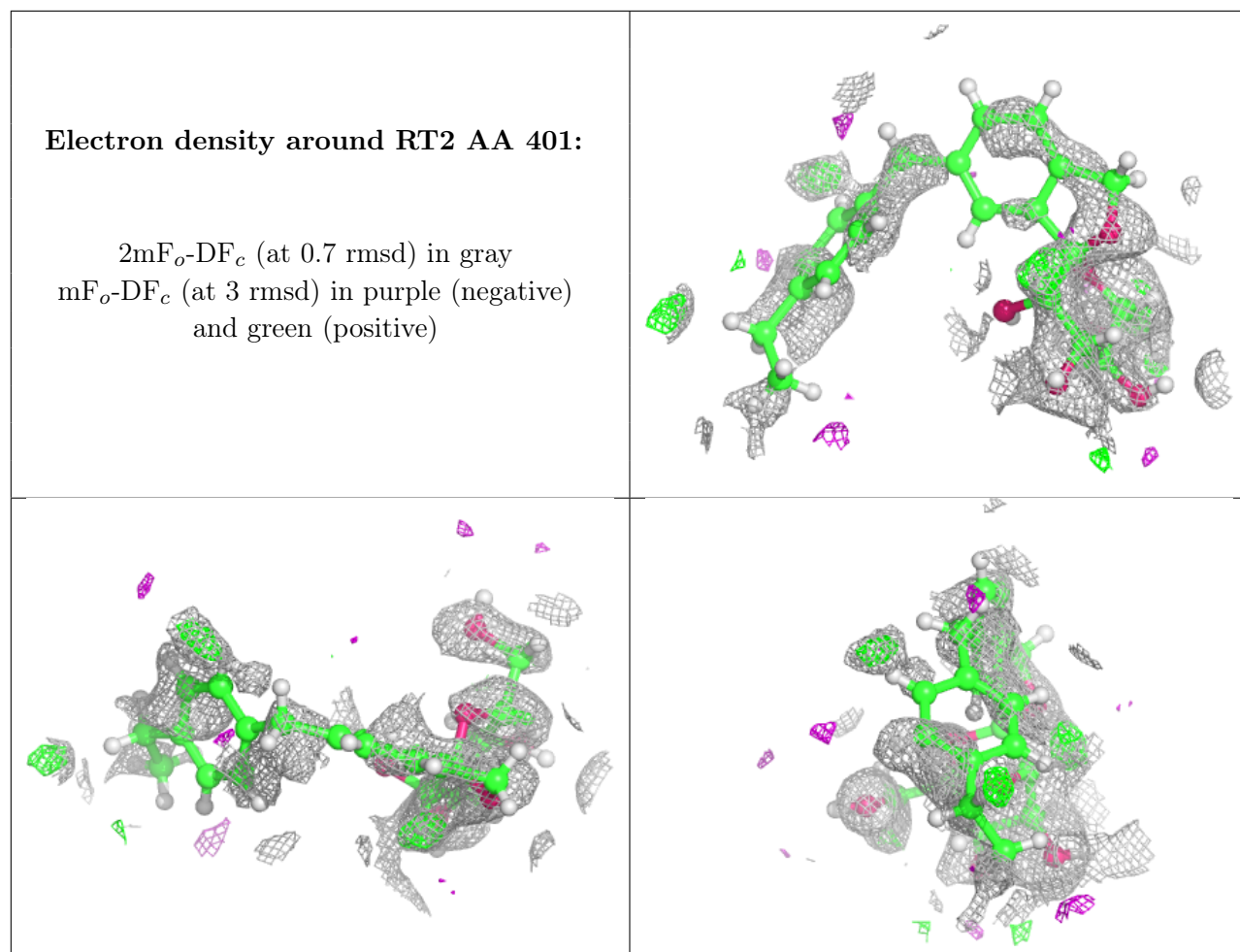
There are no monosaccharides in this entry.

### 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(Å <sup>2</sup> )	Q<0.9
2	RT2	AA	401	28/28	0.52	0.38	1,23,40,44	54
3	DMS	AA	406	4/4	0.60	0.20	1,1,66,72	6
4	CL	AA	408	1/1	0.65	0.19	86,86,86,86	1
3	DMS	AA	404	4/4	0.82	0.19	1,1,59,60	6
3	DMS	AA	407	4/4	0.83	0.14	1,1,80,86	6
3	DMS	AA	405	4/4	0.84	0.17	1,1,48,56	6
3	DMS	AA	402	4/4	0.94	0.09	1,1,27,28	6
4	CL	AA	403	1/1	0.98	0.06	34,34,34,34	1

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.