



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 16, 2022 – 03:53 PM JST

PDB ID : 7WYP  
Title : Structure of the SARS-COV-2 main protease with EN102 inhibitor  
Authors : Qin, B.; Hou, P.; Gao, X.; Cui, S.  
Deposited on : 2022-02-16  
Resolution : 2.30 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 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 : 2.28.1  
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) : 2.28.1

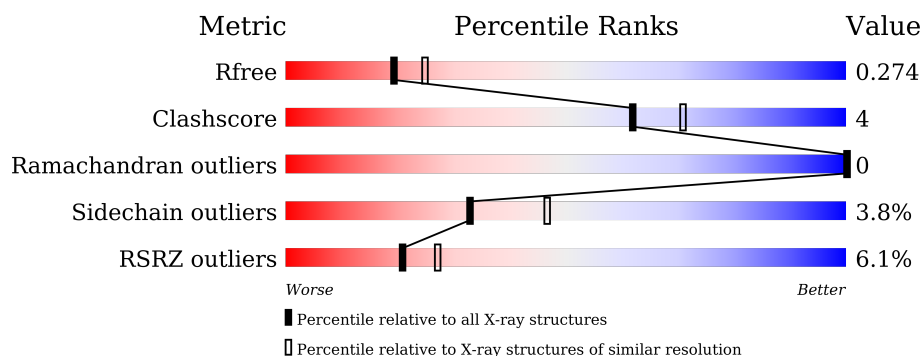
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	306	<div> <div>2%</div> <div>89%</div> <div>8%</div> <div>.</div> </div>
1	B	306	<div> <div>10%</div> <div>82%</div> <div>15%</div> <div>.</div> </div>

## 2 Entry composition [i](#)

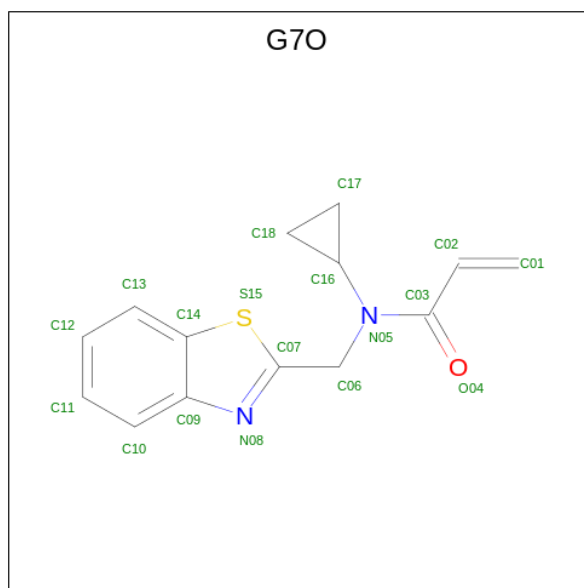
There are 3 unique types of molecules in this entry. The entry contains 4805 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 3C-like proteinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	300	Total	C	N	O	S	0	1	0
			2326	1473	395	435	23			
1	B	305	Total	C	N	O	S	0	1	0
			2361	1496	400	442	23			

- Molecule 2 is N-(1,3-benzothiazol-2-ylmethyl)-N-cyclopropyl-prop-2-enamide (three-letter code: G7O) (formula: C<sub>14</sub>H<sub>14</sub>N<sub>2</sub>OS) (labeled as "Ligand of Interest" by depositor).



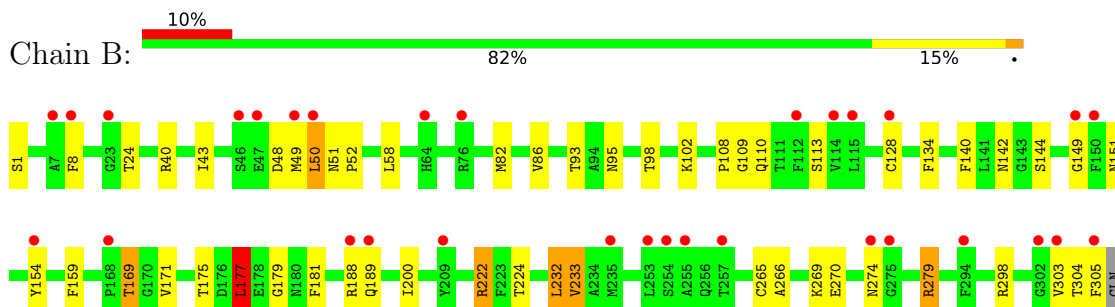
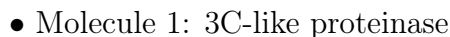
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			18	14	2	1	1		
2	B	1	Total	C	N	O	S	0	0
			18	14	2	1	1		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	57	Total 57	O 57	0	0
3	B	25	Total 25	O 25	0	0



- Molecule 1: 3C-like proteinase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.81Å 101.28Å 103.15Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.96 – 2.30 45.96 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.8 (45.96-2.30) 99.8 (45.96-2.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.34 (at 2.29Å)	Xtriage
Refinement program	PHENIX 1.20_4459	Depositor
R, $R_{free}$	0.224 , 0.273 0.225 , 0.274	Depositor DCC
$R_{free}$ test set	1525 reflections (4.72%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	45.5	Xtriage
Anisotropy	0.638	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 39.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.000 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	4805	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.01% 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: G7O

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	0.38	0/2381	0.61	1/3236 (0.0%)
1	B	0.37	0/2417	0.58	1/3285 (0.0%)
All	All	0.38	0/4798	0.59	2/6521 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	177	LEU	CA-CB-CG	-7.24	98.64	115.30
1	A	177	LEU	CA-CB-CG	-7.21	98.71	115.30

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	2326	0	2277	15	0
1	B	2361	0	2310	29	0
2	A	18	0	0	0	0
2	B	18	0	0	0	0
3	A	57	0	0	3	0
3	B	25	0	0	2	0
All	All	4805	0	4587	41	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 (41) 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:123:SER:HB2	1:B:304:THR:HG22	1.66	0.77
1:A:129:ALA:O	3:A:502:HOH:O	2.03	0.76
1:B:52:PRO:HD2	1:B:188:ARG:HG2	1.68	0.75
1:B:49:MET:HE2	1:B:189:GLN:HB2	1.68	0.74
1:B:169:THR:HG23	1:B:171:VAL:HG22	1.73	0.70
1:B:298:ARG:NH1	3:B:501:HOH:O	2.20	0.69
1:B:298:ARG:HG3	1:B:303:VAL:HB	1.78	0.66
1:A:119:ASN:OD1	3:A:504:HOH:O	2.17	0.60
1:A:166:GLU:OE1	1:B:1:SER:N	2.28	0.57
1:A:159:PHE:HB3	1:A:177:LEU:HD13	1.86	0.57
1:B:86:VAL:HG13	1:B:179:GLY:HA2	1.86	0.56
1:B:50:LEU:HD12	1:B:51:ASN:H	1.71	0.55
1:A:231:ASN:O	1:A:235:MET:HG3	2.06	0.55
1:B:108:PRO:HG3	1:B:134:PHE:CE1	2.43	0.54
1:A:109:GLY:HA2	1:A:200:ILE:HD13	1.90	0.54
1:A:118:TYR:HB2	1:B:304:THR:HG21	1.92	0.52
1:A:222:ARG:NH2	3:A:506:HOH:O	2.23	0.52
1:B:159:PHE:HB3	1:B:177:LEU:HD13	1.90	0.52
1:B:279:ARG:HA	1:B:279:ARG:HE	1.75	0.52
1:A:40:ARG:HD3	1:A:85:CYS:HA	1.92	0.52
1:B:8:PHE:HE1	1:B:305:PHE:CZ	2.28	0.51
1:B:233:VAL:HG11	1:B:269:LYS:HG3	1.91	0.51
1:B:266:ALA:O	1:B:270:GLU:HG2	2.11	0.51
1:A:113:SER:O	1:A:149:GLY:HA2	2.12	0.49
1:B:95:ASN:HB3	1:B:98:THR:OG1	2.13	0.49
1:B:8:PHE:HE2	1:B:151:ASN:HD22	1.61	0.49
1:A:95:ASN:HB3	1:A:98:THR:OG1	2.14	0.47
1:B:102:LYS:NZ	3:B:504:HOH:O	2.46	0.47
1:B:113:SER:O	1:B:149:GLY:HA2	2.14	0.47
1:A:10:SER:O	1:A:14:GLU:HG3	2.16	0.45
1:B:109:GLY:HA2	1:B:200:ILE:HD13	1.99	0.44
1:A:118:TYR:CE1	1:A:144:SER:HB3	2.52	0.44
1:B:175:THR:HG22	1:B:181:PHE:HA	2.00	0.44
1:B:265:CYS:O	1:B:269:LYS:N	2.43	0.44
1:B:140:PHE:HB3	1:B:144:SER:OG	2.18	0.43
1:A:48:ASP:O	1:A:52:PRO:HB3	2.18	0.42
1:B:40:ARG:O	1:B:43:ILE:HG12	2.20	0.41

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:222:ARG:H	1:B:222:ARG:HG3	1.46	0.41
1:B:232:LEU:HD12	1:B:232:LEU:HA	1.82	0.41
1:B:142:ASN:OD1	1:B:142:ASN:N	2.54	0.41
1:B:58:LEU:HD22	1:B:82:MET:HB2	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	299/306 (98%)	289 (97%)	10 (3%)	0	100	100
1	B	304/306 (99%)	296 (97%)	8 (3%)	0	100	100
All	All	603/612 (98%)	585 (97%)	18 (3%)	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	259/263 (98%)	253 (98%)	6 (2%)	50	67
1	B	263/263 (100%)	247 (94%)	16 (6%)	18	25
All	All	522/526 (99%)	500 (96%)	22 (4%)	33	42

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	24	THR
1	A	128[A]	CYS
1	A	128[B]	CYS
1	A	154	TYR
1	A	169	THR
1	A	177	LEU
1	B	24	THR
1	B	48	ASP
1	B	50	LEU
1	B	93	THR
1	B	110	GLN
1	B	128[A]	CYS
1	B	128[B]	CYS
1	B	154	TYR
1	B	169	THR
1	B	177	LEU
1	B	222	ARG
1	B	224	THR
1	B	232	LEU
1	B	233	VAL
1	B	274	ASN
1	B	279	ARG

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

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 5.6 Ligand geometry

2 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	G7O	A	401	1	18,20,20	2.45	6 (33%)	19,28,28	2.78	8 (42%)
2	G7O	B	401	1	18,20,20	2.57	5 (27%)	19,28,28	2.23	6 (31%)

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	G7O	A	401	1	-	3/13/16/16	0/3/3/3
2	G7O	B	401	1	-	3/13/16/16	0/3/3/3

All (11) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	401	G7O	C03-N05	7.62	1.47	1.34
2	A	401	G7O	C03-N05	6.60	1.45	1.34
2	B	401	G7O	C18-C16	-4.36	1.39	1.48
2	A	401	G7O	C17-C16	-4.28	1.39	1.48
2	A	401	G7O	C18-C16	-4.27	1.39	1.48
2	B	401	G7O	C17-C16	-4.03	1.40	1.48
2	A	401	G7O	C06-C07	2.48	1.54	1.51
2	A	401	G7O	C12-C13	2.43	1.42	1.36
2	B	401	G7O	C18-C17	-2.42	1.39	1.48
2	A	401	G7O	C18-C17	-2.25	1.40	1.48
2	B	401	G7O	C12-C13	2.18	1.41	1.36

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	401	G7O	C09-C14-S15	-6.00	103.90	111.85
2	B	401	G7O	C09-C14-S15	-5.97	103.93	111.85
2	A	401	G7O	C01-C02-C03	-5.76	109.54	121.33
2	B	401	G7O	C01-C02-C03	-5.01	111.08	121.33
2	A	401	G7O	C06-N05-C16	4.92	123.48	118.33
2	A	401	G7O	O04-C03-N05	-3.89	116.55	120.22
2	A	401	G7O	C14-C09-N08	3.21	115.52	108.04
2	B	401	G7O	C14-C09-N08	3.20	115.49	108.04
2	A	401	G7O	C07-C06-N05	-3.14	107.20	112.37
2	A	401	G7O	C02-C03-N05	3.04	121.38	117.86
2	B	401	G7O	C02-C03-N05	2.68	120.96	117.86
2	A	401	G7O	C10-C09-N08	-2.42	123.76	130.78
2	B	401	G7O	C10-C09-N08	-2.40	123.83	130.78
2	B	401	G7O	C06-N05-C16	2.35	120.79	118.33

There are no chirality outliers.

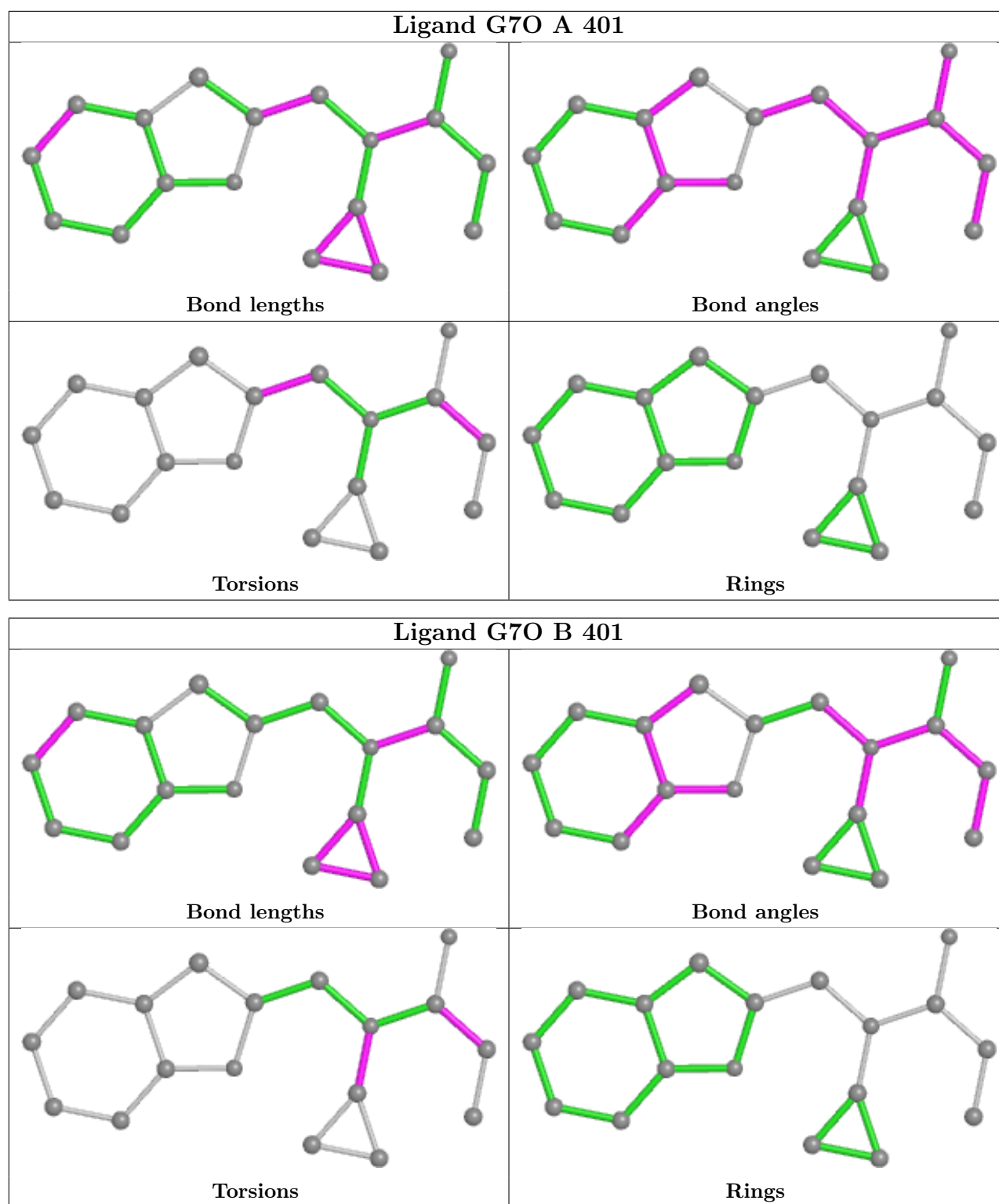
All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	401	G7O	C01-C02-C03-N05
2	B	401	G7O	C01-C02-C03-O04
2	A	401	G7O	C01-C02-C03-N05
2	A	401	G7O	C01-C02-C03-O04
2	B	401	G7O	C18-C16-N05-C03
2	A	401	G7O	N05-C06-C07-N08

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 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 ⓘ

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

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	300/306 (98%)	0.21	6 (2%) 65 71	41, 49, 59, 75	0
1	B	305/306 (99%)	0.66	31 (10%) 6 9	41, 59, 83, 95	0
All	All	605/612 (98%)	0.44	37 (6%) 21 27	41, 52, 79, 95	0

All (37) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	46	SER	4.4
1	B	294	PHE	4.2
1	B	154	TYR	3.9
1	B	49	MET	3.4
1	B	255	ALA	3.4
1	B	274	ASN	3.3
1	B	128[A]	CYS	3.2
1	B	50	LEU	3.2
1	B	114	VAL	3.0
1	B	112	PHE	2.9
1	A	154	TYR	2.9
1	A	294	PHE	2.9
1	B	23	GLY	2.8
1	B	302	GLY	2.8
1	B	254	SER	2.5
1	B	64	HIS	2.4
1	B	76	ARG	2.3
1	A	73	VAL	2.3
1	A	7	ALA	2.3
1	B	253	LEU	2.3
1	B	303	VAL	2.2
1	B	305	PHE	2.2
1	A	283	GLY	2.2
1	B	275	GLY	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	189	GLN	2.2
1	B	7	ALA	2.2
1	B	8	PHE	2.1
1	B	47	GLU	2.1
1	B	115	LEU	2.1
1	A	256	GLN	2.1
1	B	209	TYR	2.1
1	B	149	GLY	2.1
1	B	150	PHE	2.1
1	B	257	THR	2.1
1	B	168	PRO	2.0
1	B	235	MET	2.0
1	B	188	ARG	2.0

## 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](#)

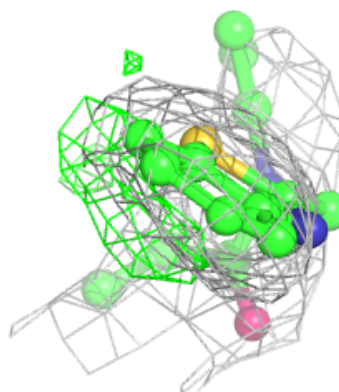
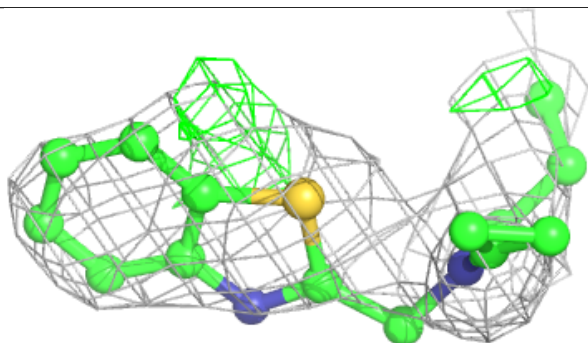
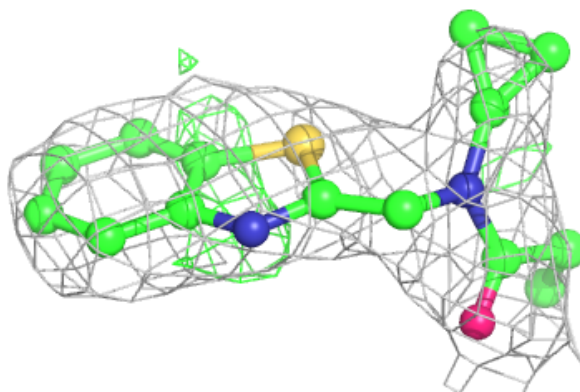
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 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
2	G7O	A	401	18/18	0.82	0.26	50,51,55,55	18
2	G7O	B	401	18/18	0.85	0.29	58,63,67,67	18

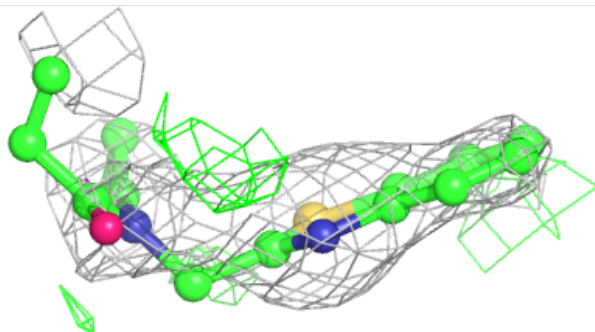
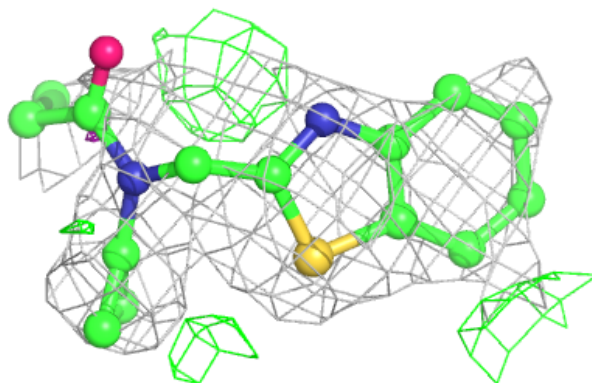
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

**Electron density around G7O A 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 G7O 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)



## 6.5 Other polymers [i](#)

There are no such residues in this entry.