



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 26, 2020 – 04:12 PM EST

PDB ID : 7KHS  
Title : OgOGA IN COMPLEX WITH LIGAND 55  
Authors : Shaffer, P.L.  
Deposited on : 2020-10-21  
Resolution : 1.78 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 : 2.14.6  
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) : 2.14.6

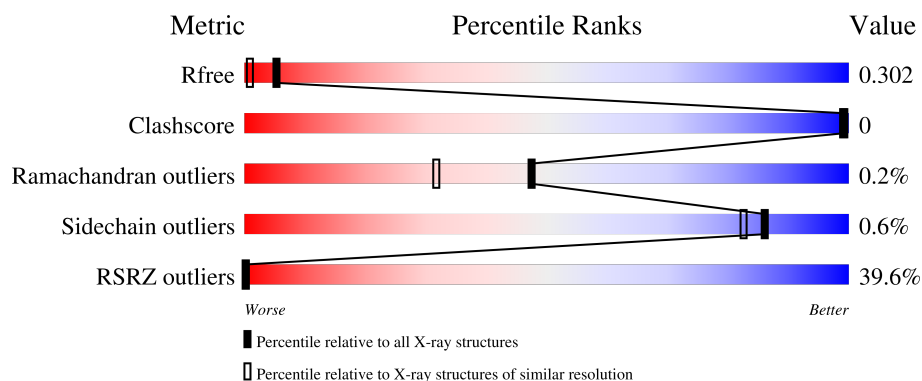
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.78 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	9185 (1.80-1.76)
Clashscore	141614	10184 (1.80-1.76)
Ramachandran outliers	138981	10051 (1.80-1.76)
Sidechain outliers	138945	10050 (1.80-1.76)
RSRZ outliers	127900	9032 (1.80-1.76)

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	449	
1	B	449	
1	C	449	
1	D	449	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 28033 atoms, of which 13715 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.

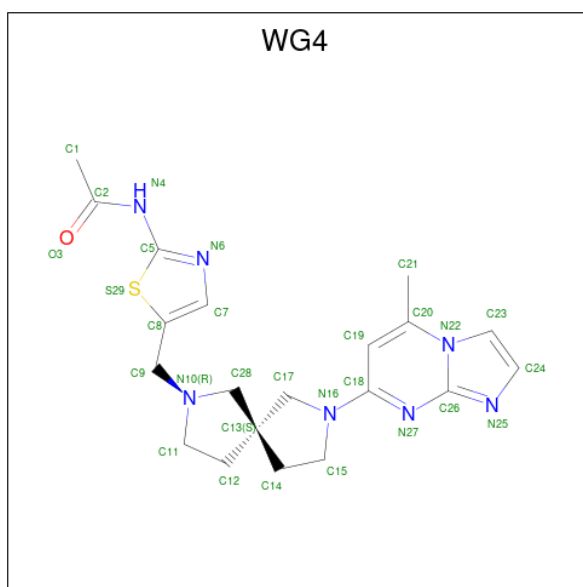
- Molecule 1 is a protein called Protein O-GlcNAcase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	449	Total	C	H	N	O	S	198	2	0
			7042	2263	3474	652	641	12			
1	B	430	Total	C	H	N	O	S	218	0	0
			6715	2159	3317	619	608	12			
1	C	427	Total	C	H	N	O	S	386	2	0
			6710	2156	3316	622	604	12			
1	D	448	Total	C	H	N	O	S	133	4	0
			7059	2268	3484	653	642	12			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP Q2CEE3
A	0	SER	-	expression tag	UNP Q2CEE3
B	-1	GLY	-	expression tag	UNP Q2CEE3
B	0	SER	-	expression tag	UNP Q2CEE3
C	-1	GLY	-	expression tag	UNP Q2CEE3
C	0	SER	-	expression tag	UNP Q2CEE3
D	-1	GLY	-	expression tag	UNP Q2CEE3
D	0	SER	-	expression tag	UNP Q2CEE3

- Molecule 2 is N-(5-([(5S)-7-(5-methylimidazo[1,2-a]pyrimidin-7-yl)-2,7-diazaspiro[4.4]nonan-2-yl)methyl]-1,3-thiazol-2-yl)acetamide (three-letter code: WG4) (formula: C<sub>20</sub>H<sub>25</sub>N<sub>7</sub>OS) (labeled as "Ligand of Interest" by author).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	S	0	0
			54	20	25	7	1	1		
2	B	1	Total	C	H	N	O	S	0	0
			54	20	25	7	1	1		
2	C	1	Total	C	H	N	O	S	0	0
			54	20	25	7	1	1		
2	D	1	Total	C	H	N	O	S	0	0
			54	20	25	7	1	1		

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Ca	0	0
			1	1		
3	D	1	Total	Ca	0	0
			1	1		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	C	1	Total	C	H	O	1	0
			10	2	6	2		

- Molecule 5 is water.

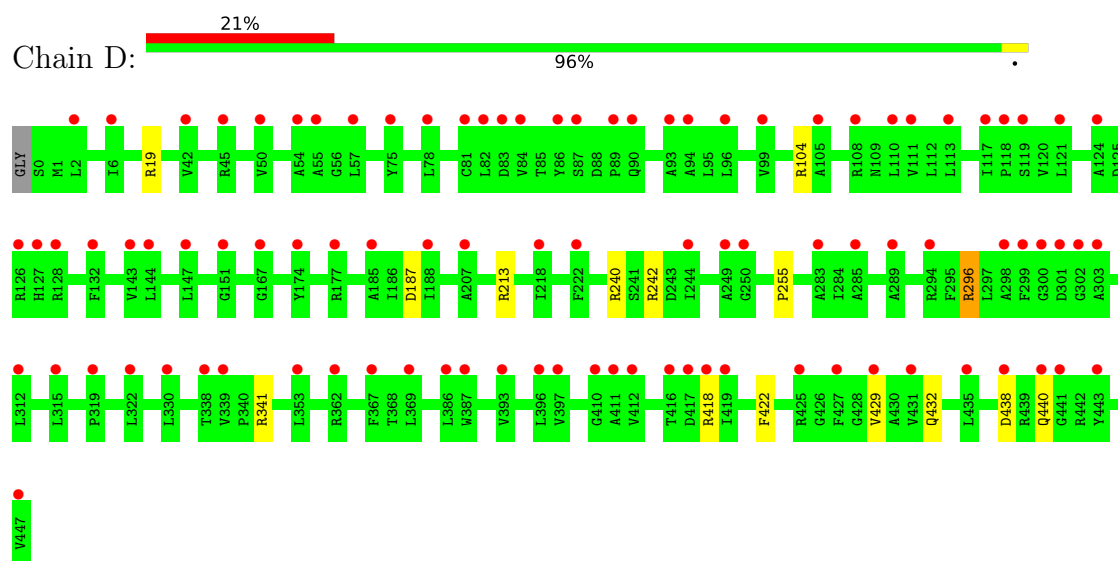
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	116	Total	H	O	0	0
			118	2	116		
5	B	38	Total	H	O	0	0
			46	8	38		
5	C	18	Total	O		0	0
			18	18			
5	D	89	Total	H	O	0	0
			97	8	89		



● Molecule 1: Protein O-GlcNAcase



● Molecule 1: Protein O-GlcNAcase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.23Å 83.08Å 122.96Å 90.00° 90.25° 90.00°	Depositor
Resolution (Å)	44.79 – 1.78 44.67 – 1.79	Depositor EDS
% Data completeness (in resolution range)	87.4 (44.79-1.78) 87.4 (44.67-1.79)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.37 (at 1.79Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
R, $R_{free}$	0.273 , 0.296 0.277 , 0.302	Depositor DCC
$R_{free}$ test set	2508 reflections (1.42%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.9	Xtriage
Anisotropy	0.283	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 17.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.399 for h,-k,-l	Xtriage
Reported twinning fraction	0.625 for H, K, L 0.375 for -h,-k,l	Depositor
Outliers	0 of 176418 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	28033	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.41% 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: WG4, CA, EDO

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	0.84	0/3666	0.99	20/4999 (0.4%)
1	B	0.76	0/3488	0.95	17/4754 (0.4%)
1	C	0.74	1/3484 (0.0%)	0.93	11/4747 (0.2%)
1	D	0.80	0/3673	0.97	11/5009 (0.2%)
All	All	0.79	1/14311 (0.0%)	0.96	59/19509 (0.3%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	350	ARG	CD-NE	10.34	1.64	1.46

The worst 5 of 59 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	215	ARG	NE-CZ-NH2	-7.41	116.59	120.30
1	A	240	ARG	NE-CZ-NH1	7.27	123.93	120.30
1	D	213	ARG	NE-CZ-NH1	7.19	123.90	120.30
1	C	296	ARG	NE-CZ-NH1	6.75	123.68	120.30
1	B	296	ARG	NE-CZ-NH1	6.67	123.64	120.30

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	3568	3474	3459	2	0
1	B	3398	3317	3305	2	0
1	C	3394	3316	3302	3	0
1	D	3575	3484	3469	2	0
2	A	29	25	0	0	0
2	B	29	25	0	1	0
2	C	29	25	0	0	0
2	D	29	25	0	1	0
3	B	1	0	0	0	0
3	D	1	0	0	0	0
4	C	4	6	6	0	0
5	A	116	2	0	0	0
5	B	38	8	0	0	0
5	C	18	0	0	0	0
5	D	89	8	0	0	0
All	All	14318	13715	13541	9	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 0.

The worst 5 of 9 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:C:294[B]:ARG:HB2	1:C:294[B]:ARG:HH11	1.31	0.91
1:C:294[B]:ARG:NH1	1:C:294[B]:ARG:HB2	2.09	0.65
1:D:429:VAL:HA	1:D:432:GLN:HG2	1.80	0.64
1:A:227:TYR:CE1	1:B:422:PHE:HB2	2.50	0.46
1:B:242:ARG:HD2	1:B:272:TYR:O	2.17	0.45

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	449/449 (100%)	435 (97%)	13 (3%)	1 (0%)	47	32
1	B	424/449 (94%)	411 (97%)	12 (3%)	1 (0%)	47	32
1	C	423/449 (94%)	409 (97%)	13 (3%)	1 (0%)	47	32
1	D	450/449 (100%)	436 (97%)	13 (3%)	1 (0%)	47	32
All	All	1746/1796 (97%)	1691 (97%)	51 (3%)	4 (0%)	47	32

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	255	PRO
1	A	255	PRO
1	C	255	PRO
1	D	255	PRO

### 5.3.2 Protein sidechains ⓘ

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	358/356 (101%)	355 (99%)	3 (1%)	81	76
1	B	339/356 (95%)	338 (100%)	1 (0%)	92	90
1	C	338/356 (95%)	336 (99%)	2 (1%)	86	82
1	D	360/356 (101%)	357 (99%)	3 (1%)	81	76
All	All	1395/1424 (98%)	1386 (99%)	9 (1%)	86	82

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	218	ILE
1	D	440	GLN
1	D	296	ARG
1	A	438	ASP
1	C	438	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such

sidechains are listed below:

Mol	Chain	Res	Type
1	B	390	GLN
1	D	390	GLN
1	C	390	GLN
1	A	390	GLN
1	D	109	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 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	$\# Z  > 2$
2	WG4	C	501	-	27,33,33	1.79	5 (18%)	24,49,49	2.59	12 (50%)
2	WG4	A	501	-	27,33,33	1.60	4 (14%)	24,49,49	2.92	10 (41%)
2	WG4	D	501	-	27,33,33	1.53	7 (25%)	24,49,49	2.84	11 (45%)
4	EDO	C	502	-	3,3,3	0.47	0	2,2,2	0.52	0
2	WG4	B	501	-	27,33,33	1.55	6 (22%)	24,49,49	2.38	7 (29%)

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	WG4	C	501	-	-	2/9/34/34	0/5/5/5
2	WG4	A	501	-	-	4/9/34/34	0/5/5/5
2	WG4	D	501	-	-	2/9/34/34	0/5/5/5
4	EDO	C	502	-	-	1/1/1/1	-
2	WG4	B	501	-	-	2/9/34/34	0/5/5/5

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	501	WG4	C28-N10	6.45	1.52	1.46
2	A	501	WG4	C28-N10	5.25	1.51	1.46
2	D	501	WG4	C28-N10	4.09	1.50	1.46
2	B	501	WG4	C7-N6	-3.47	1.31	1.36
2	A	501	WG4	C9-C8	-3.38	1.47	1.51

The worst 5 of 40 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	501	WG4	C9-N10-C28	7.49	122.44	112.93
2	D	501	WG4	C18-N27-C26	7.01	122.92	117.65
2	C	501	WG4	C9-N10-C28	6.52	121.21	112.93
2	B	501	WG4	C9-N10-C28	5.61	120.06	112.93
2	D	501	WG4	C9-N10-C28	5.47	119.87	112.93

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

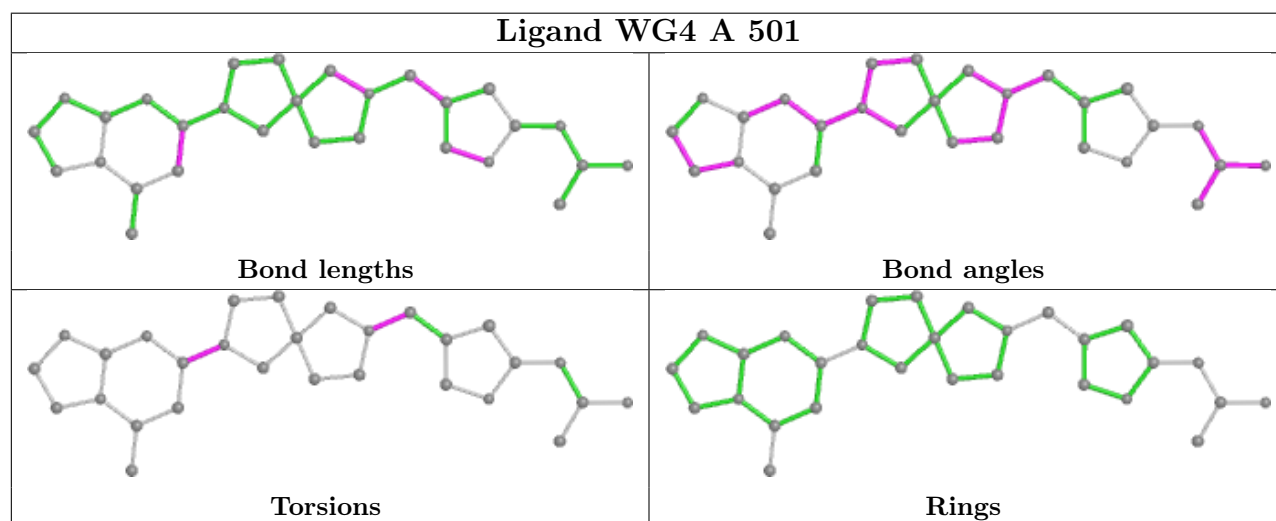
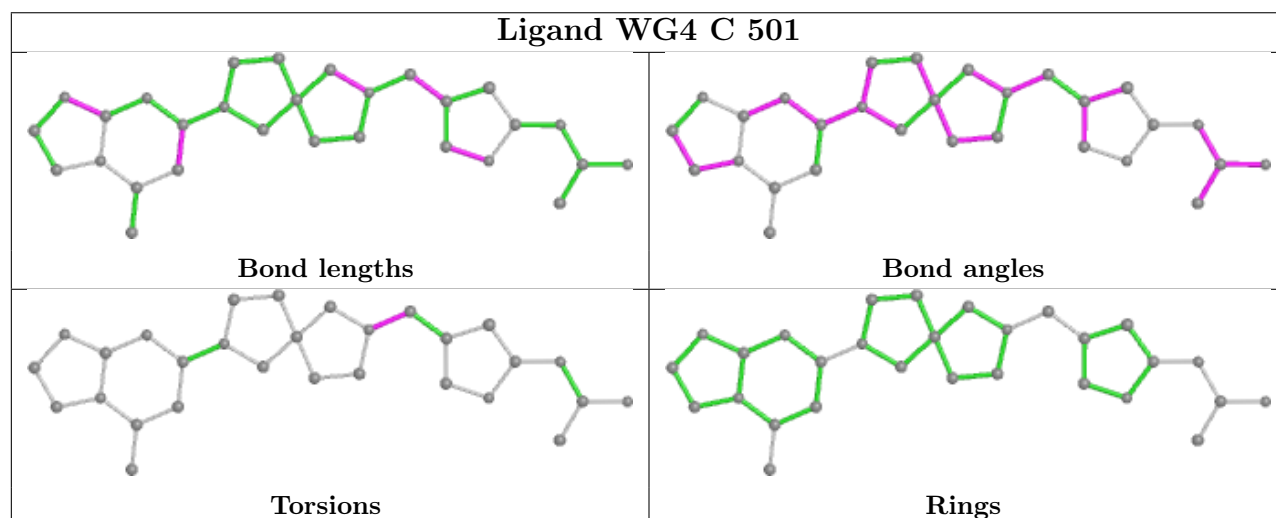
Mol	Chain	Res	Type	Atoms
2	C	501	WG4	C8-C9-N10-C11
2	C	501	WG4	C8-C9-N10-C28
2	A	501	WG4	C8-C9-N10-C11
2	A	501	WG4	C8-C9-N10-C28
2	D	501	WG4	C8-C9-N10-C28

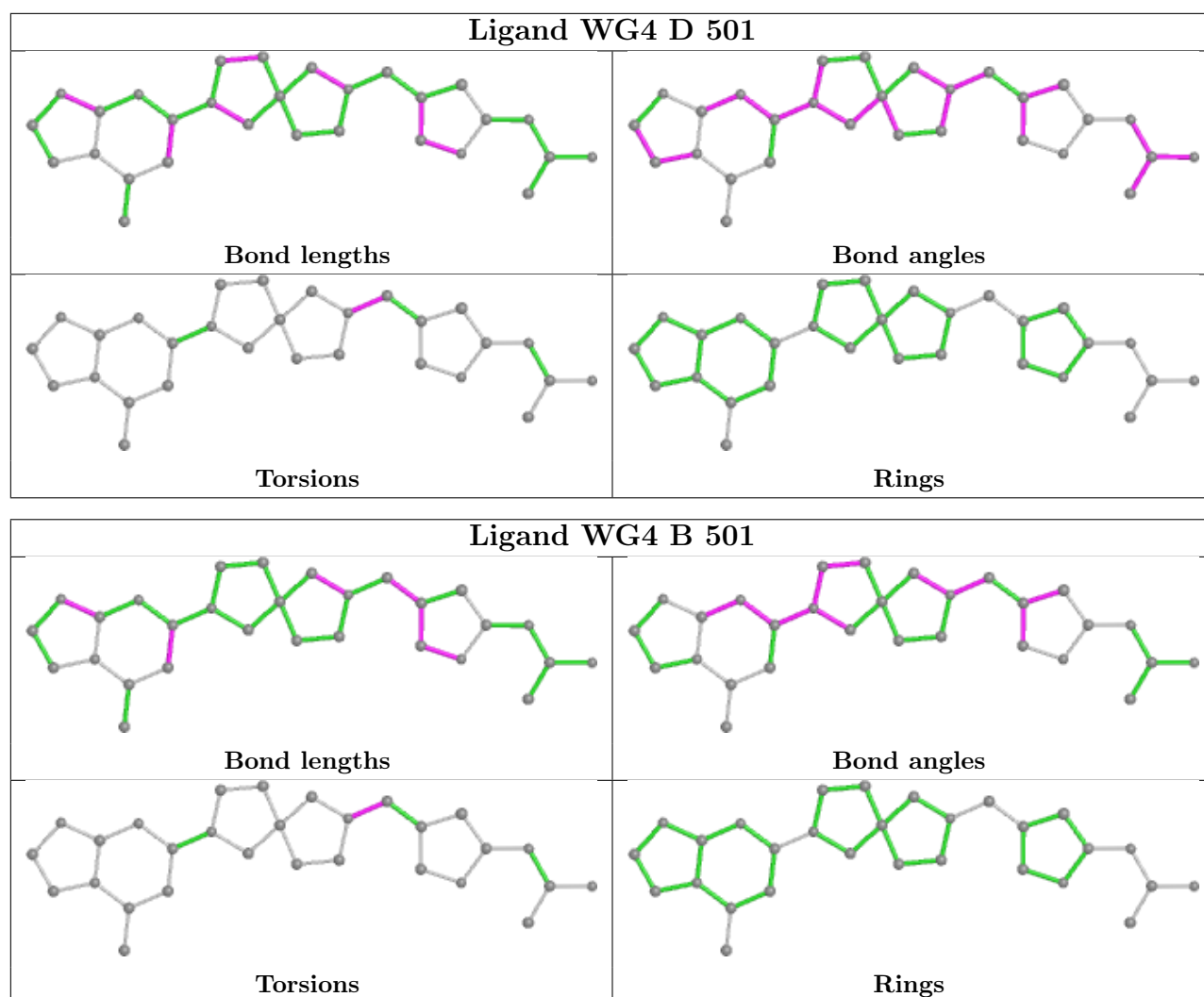
There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	501	WG4	1	0
2	B	501	WG4	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	449/449 (100%)	1.39	86 (19%) <b>1</b> <b>1</b>	20, 29, 52, 73	18 (4%)
1	B	430/449 (95%)	2.56	231 (53%) <b>0</b> <b>0</b>	23, 39, 65, 94	20 (4%)
1	C	427/449 (95%)	3.09	284 (66%) <b>0</b> <b>0</b>	25, 49, 72, 89	39 (9%)
1	D	448/449 (99%)	1.52	94 (20%) <b>1</b> <b>0</b>	20, 30, 48, 65	10 (2%)
All	All	1754/1796 (97%)	2.13	695 (39%) <b>0</b> <b>0</b>	20, 35, 65, 94	87 (4%)

The worst 5 of 695 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	82	LEU	16.6
1	B	86	TYR	12.5
1	B	112	LEU	12.2
1	D	299	PHE	12.2
1	C	132	PHE	11.5

### 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.

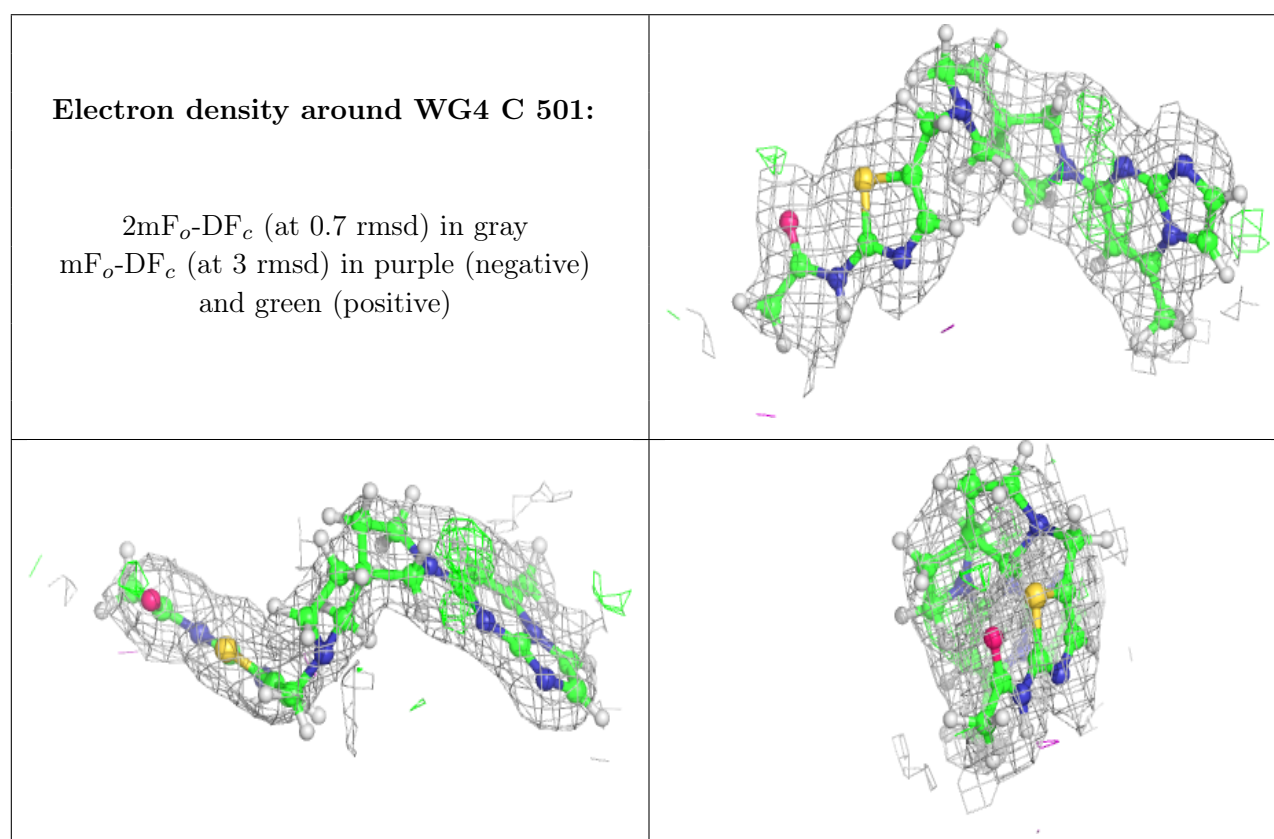
### 6.4 Ligands [i](#)

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.



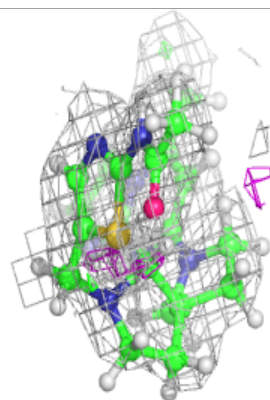
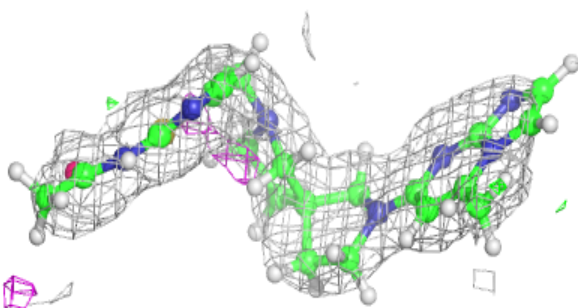
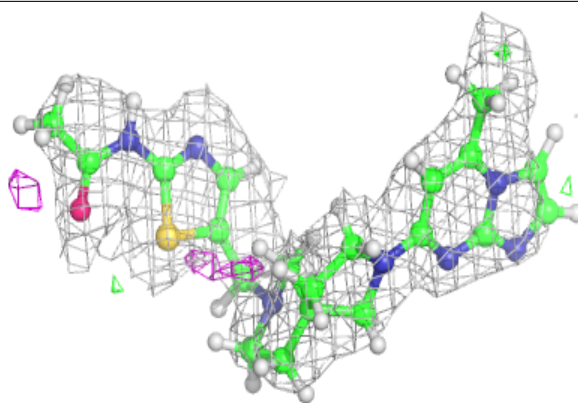
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	CA	B	502	1/1	0.80	0.07	37,37,37,37	0
2	WG4	C	501	29/29	0.81	0.17	25,31,40,44	0
2	WG4	B	501	29/29	0.83	0.18	20,30,43,50	0
2	WG4	A	501	29/29	0.84	0.15	24,31,38,41	0
2	WG4	D	501	29/29	0.91	0.13	19,25,40,44	0
4	EDO	C	502	4/4	0.92	0.17	18,19,25,26	1
3	CA	D	502	1/1	0.98	0.07	29,29,29,29	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.

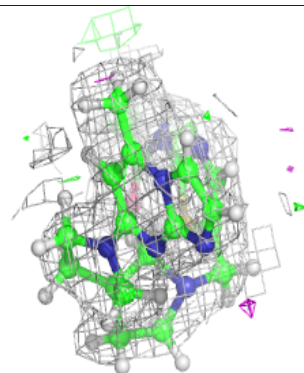
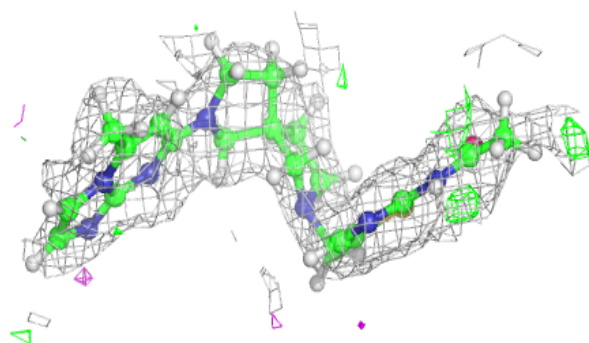
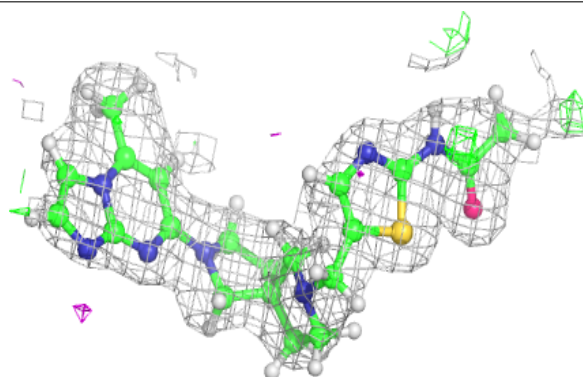


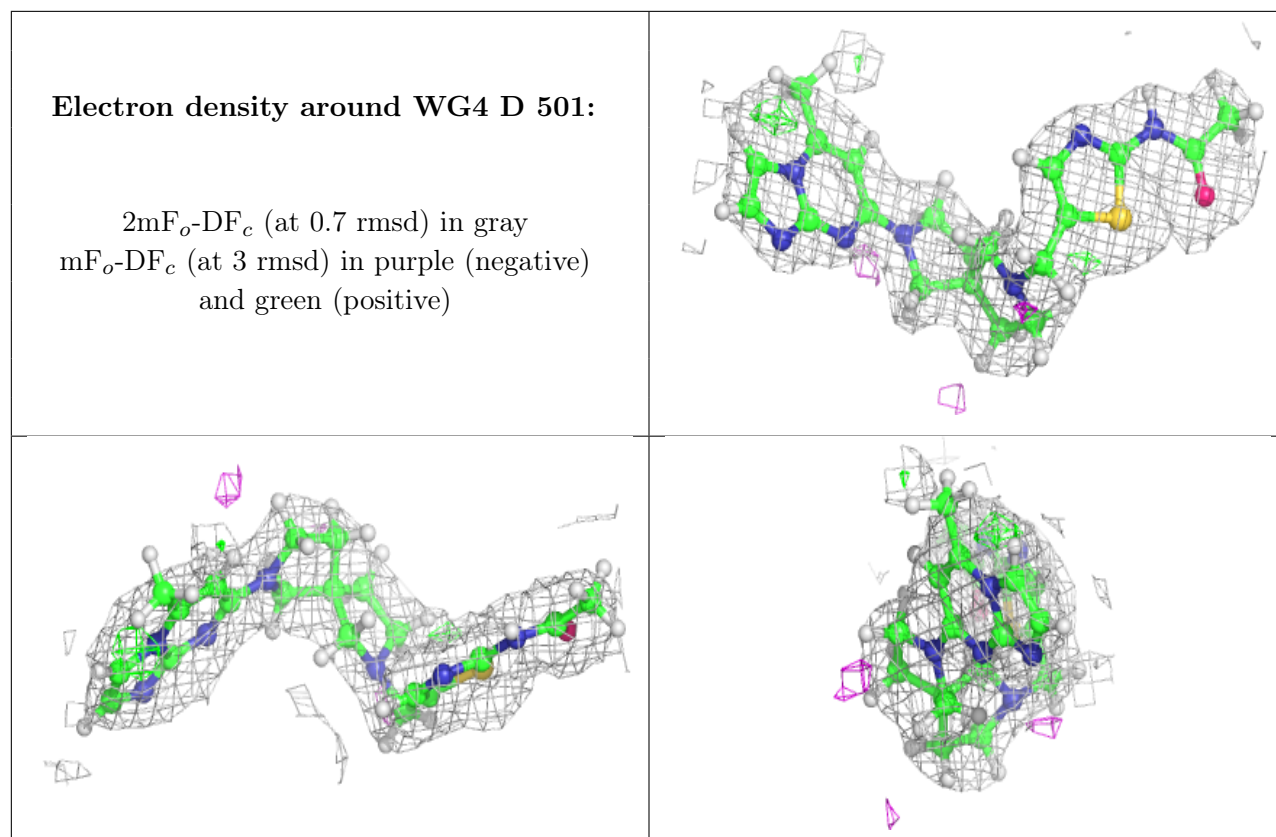
**Electron density around WG4 B 501:**

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

**Electron density around WG4 A 501:**

$2mF_o-DF_c$  (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.