



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 23, 2024 – 06:25 PM EDT

PDB ID : 5M7T  
Title : Structure of human O-GlcNAc hydrolase with PugNAc type inhibitor  
Authors : Roth, C.; Chan, S.; Offen, W.A.; Hemsworth, G.R.; Willems, L.I.; King, D.;  
Varghese, V.; Britton, R.; Vocadlo, D.J.; Davies, G.J.  
Deposited on : 2016-10-28  
Resolution : 2.60 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)  
Xtrriage (Phenix) : 1.13  
EDS : 2.37.1  
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) : 2.37.1

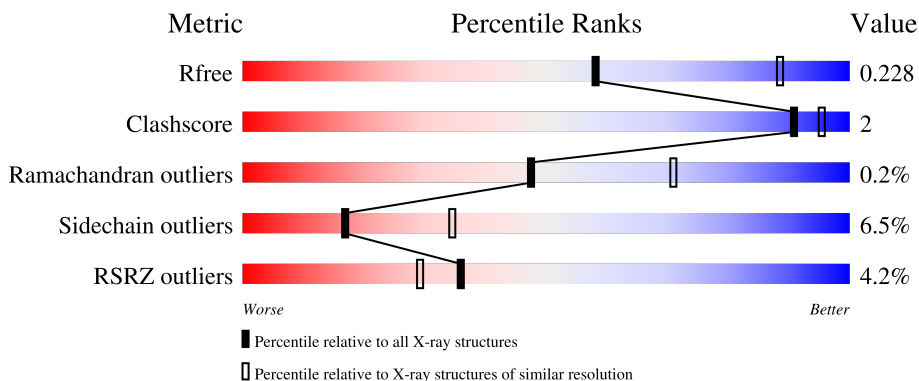
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	916	 3% 44% 5% 51%
1	B	916	 3% 44% 5% 51%

## 2 Entry composition [i](#)

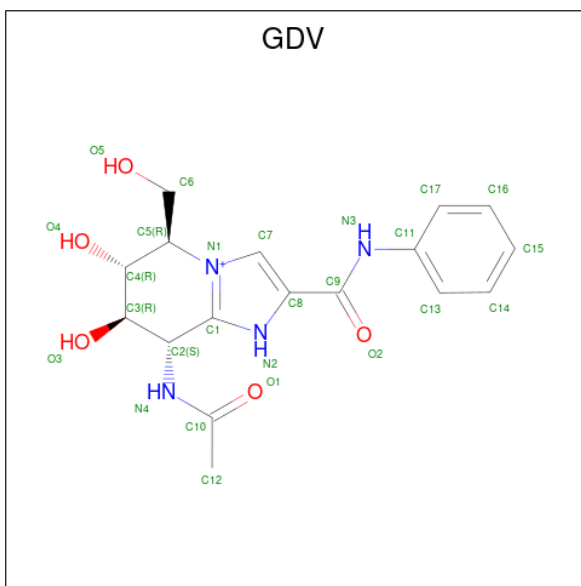
There are 2 unique types of molecules in this entry. The entry contains 7411 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Protein O-GlcNAcase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	451	Total 3695	C 2390	N 610	O 671	S 24	0	0	0
1	B	447	Total 3664	C 2372	N 605	O 663	S 24	0	0	0

- Molecule 2 is (5R,6R,7R,8S)-8-(ACETYLAMINO)-6,7-DIHYDROXY-5-(HYDROXYMETHYL)-N-PHENYL-1,5,6,7,8,8A-HEXAHYDROIMIDAZO[1,2-A]PYRIDINE-2-CARBOXAMIDE (three-letter code: GDV) (formula: C<sub>17</sub>H<sub>21</sub>N<sub>4</sub>O<sub>5</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	Total 26	C 17	N 4	O 5	0	0
2	B	1	Total 26	C 17	N 4	O 5	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.15Å 101.15Å 283.29Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	95.44 – 2.60 82.31 – 2.60	Depositor EDS
% Data completeness (in resolution range)	100.0 (95.44-2.60) 100.0 (82.31-2.60)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.74 (at 2.62Å)	Xtrriage
Refinement program	REFMAC 5.8.0155	Depositor
R, $R_{free}$	0.203 , 0.235 0.204 , 0.228	Depositor DCC
$R_{free}$ test set	1411 reflections (3.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	80.6	Xtrriage
Anisotropy	0.382	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 65.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7411	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	105.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.17% 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: GDV

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	0.67	0/3795	0.81	8/5139 (0.2%)
1	B	0.59	0/3763	0.76	2/5096 (0.0%)
All	All	0.63	0/7558	0.78	10/10235 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	634	ARG	NE-CZ-NH2	-7.77	116.42	120.30
1	A	74	VAL	CG1-CB-CG2	6.05	120.58	110.90
1	A	634	ARG	NE-CZ-NH1	5.87	123.24	120.30
1	A	104	ARG	NE-CZ-NH2	-5.36	117.62	120.30
1	B	608	ARG	NE-CZ-NH1	5.33	122.96	120.30
1	A	580	ARG	NE-CZ-NH1	5.10	122.85	120.30
1	A	156	ARG	NE-CZ-NH1	5.09	122.84	120.30
1	B	691	ARG	NE-CZ-NH1	5.08	122.84	120.30
1	A	382	LEU	CA-CB-CG	5.05	126.91	115.30
1	A	608	ARG	NE-CZ-NH1	5.01	122.80	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	664	ARG	Peptide
1	B	371	ASP	Peptide

## 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	3695	0	3619	13	0
1	B	3664	0	3592	17	0
2	A	26	0	21	0	0
2	B	26	0	21	0	0
All	All	7411	0	7253	25	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 2.

All (25) 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:75:MET:HE1	1:A:122:LEU:HD22	1.45	0.98
1:A:75:MET:CE	1:A:122:LEU:HD22	2.21	0.70
1:A:686:ALA:HB1	1:B:650:ILE:HD12	1.78	0.65
1:A:686:ALA:HB1	1:B:650:ILE:CD1	2.28	0.63
1:A:75:MET:HE1	1:A:122:LEU:CD2	2.26	0.61
1:A:684:GLY:CA	1:B:290:ARG:HD3	2.31	0.61
1:B:247:VAL:CG1	1:B:271:ILE:HD12	2.32	0.60
1:B:247:VAL:HG11	1:B:271:ILE:HD12	1.84	0.58
1:B:110:MET:CE	1:B:157:LYS:HA	2.34	0.57
1:B:604:ILE:HG23	1:B:608:ARG:HE	1.70	0.56
1:B:626:THR:O	1:B:630:ASN:ND2	2.38	0.53
1:A:381:LYS:O	1:A:385:THR:HG23	2.11	0.51
1:B:110:MET:HE1	1:B:157:LYS:HA	1.94	0.49
1:A:650:ILE:HD12	1:B:686:ALA:HB1	1.96	0.48
1:A:622:MET:O	1:A:626:THR:HG23	2.16	0.45
1:A:153:THR:HG23	1:A:156:ARG:NH2	2.32	0.44

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:161:VAL:HG23	1:B:169:PHE:HE1	1.83	0.44
1:A:96:ALA:N	1:A:97:PRO:HD3	2.33	0.44
1:A:650:ILE:CD1	1:B:686:ALA:HB1	2.48	0.43
1:B:96:ALA:N	1:B:97:PRO:HD3	2.33	0.43
1:B:653:MET:HG3	1:B:679:TRP:CZ2	2.55	0.42
1:B:247:VAL:HG13	1:B:271:ILE:HD12	2.01	0.42
1:A:625:PHE:HD1	1:A:644:VAL:HG12	1.85	0.41
1:B:158:LEU:HA	1:B:161:VAL:HG22	2.03	0.41
1:B:581:GLU:OE2	1:B:610:ARG:NH1	2.54	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	441/916 (48%)	420 (95%)	20 (4%)	1 (0%)	47	71
1	B	437/916 (48%)	413 (94%)	23 (5%)	1 (0%)	47	71
All	All	878/1832 (48%)	833 (95%)	43 (5%)	2 (0%)	47	71

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	314	PRO
1	B	314	PRO

#### 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	402/797 (50%)	373 (93%)	29 (7%)	14	29
1	B	397/797 (50%)	374 (94%)	23 (6%)	20	40
All	All	799/1594 (50%)	747 (94%)	52 (6%)	17	34

All (52) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	74	VAL
1	A	75	MET
1	A	101	TYR
1	A	149	LYS
1	A	150	GLU
1	A	156	ARG
1	A	163	GLN
1	A	216	PRO
1	A	230	GLN
1	A	253	LYS
1	A	273	ARG
1	A	303	LEU
1	A	308	LYS
1	A	336	MET
1	A	337	ASN
1	A	372	VAL
1	A	537	GLN
1	A	568	PRO
1	A	593	SER
1	A	594	VAL
1	A	599	LYS
1	A	602	GLU
1	A	605	GLU
1	A	626	THR
1	A	629	SER
1	A	634	ARG
1	A	642	SER
1	A	648	LYS
1	A	649	SER
1	B	83	ARG
1	B	101	TYR
1	B	117	GLU
1	B	119	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	127	ARG
1	B	149	LYS
1	B	156	ARG
1	B	253	LYS
1	B	273	ARG
1	B	308	LYS
1	B	336	MET
1	B	337	ASN
1	B	372	VAL
1	B	537	GLN
1	B	539	VAL
1	B	549	THR
1	B	605	GLU
1	B	608	ARG
1	B	629	SER
1	B	642	SER
1	B	648	LYS
1	B	649	SER
1	B	664	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	103	HIS
1	A	193	HIS
1	A	230	GLN
1	A	288	GLN
1	A	630	ASN
1	B	103	HIS
1	B	193	HIS
1	B	377	GLN
1	B	537	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](#)

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	GDV	B	1001	-	24,28,28	0.97	1 (4%)	27,40,40	1.39	4 (14%)
2	GDV	A	1001	-	24,28,28	1.14	3 (12%)	27,40,40	1.19	2 (7%)

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	GDV	B	1001	-	-	1/10/34/34	0/2/3/3
2	GDV	A	1001	-	-	1/10/34/34	0/2/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1001	GDV	C7-N1	-2.71	1.34	1.38
2	A	1001	GDV	C11-N3	-2.33	1.36	1.41
2	A	1001	GDV	C3-C2	2.30	1.56	1.52
2	A	1001	GDV	C7-N1	-2.30	1.34	1.38

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1001	GDV	O4-C4-C5	3.24	115.52	109.77
2	B	1001	GDV	O2-C9-C8	-2.81	114.97	121.08
2	A	1001	GDV	O2-C9-C8	-2.62	115.37	121.08

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1001	GDV	C8-C9-N3	2.52	120.15	114.04
2	B	1001	GDV	C7-N1-C1	-2.49	107.78	109.31
2	A	1001	GDV	C6-C5-C4	2.23	116.47	112.45

There are no chirality outliers.

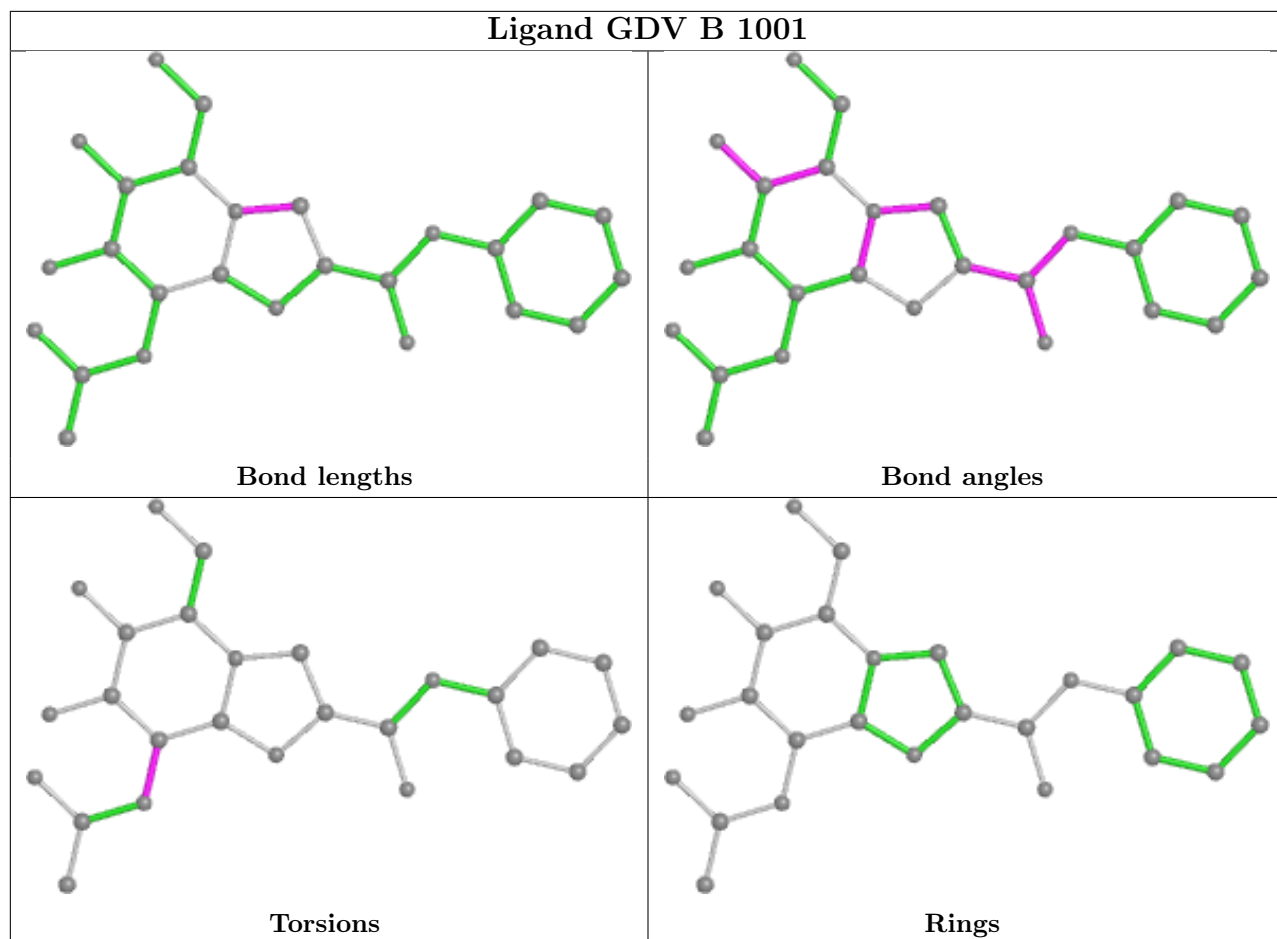
All (2) torsion outliers are listed below:

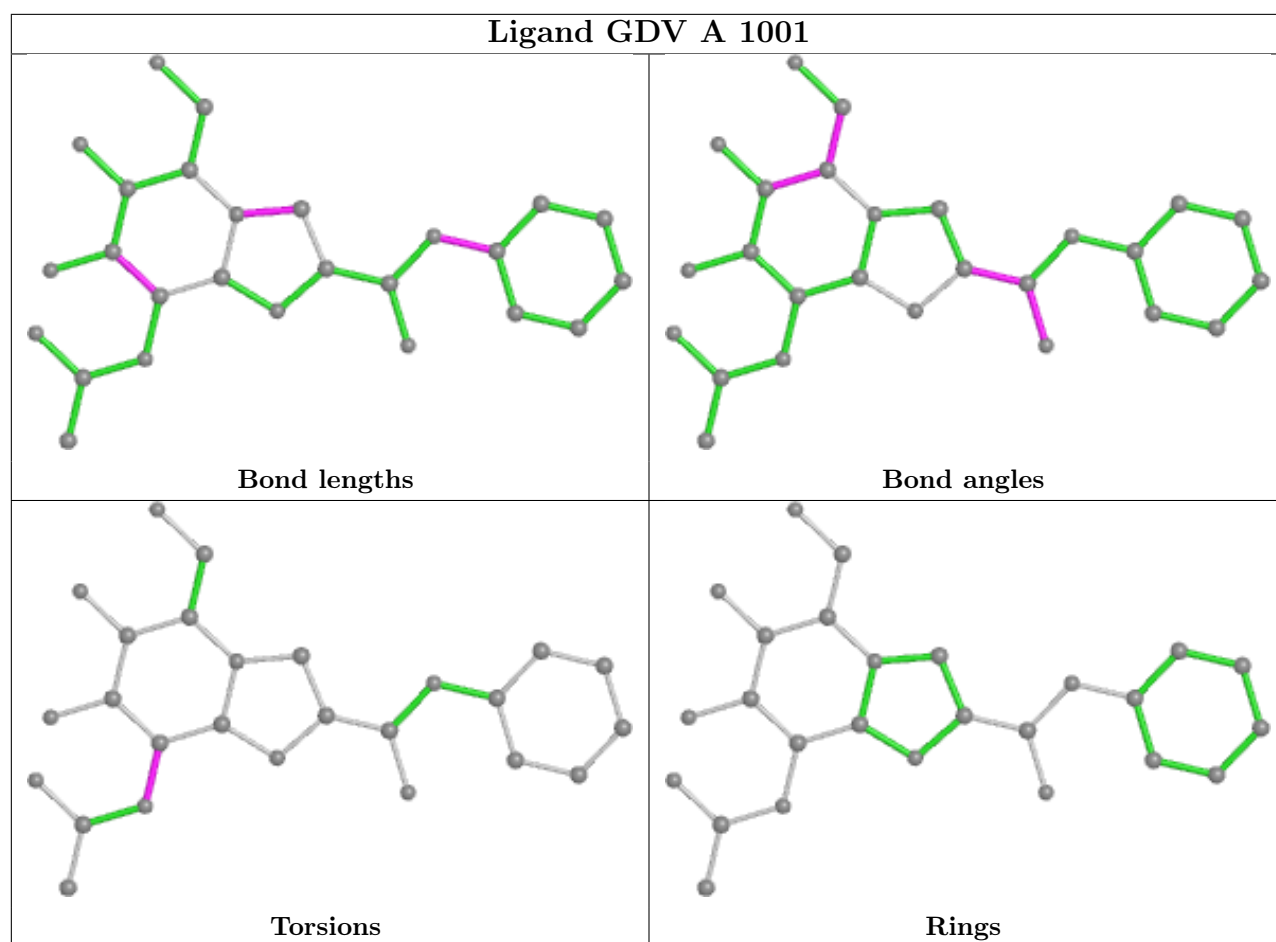
Mol	Chain	Res	Type	Atoms
2	A	1001	GDV	C1-C2-N4-C10
2	B	1001	GDV	C1-C2-N4-C10

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 [\(i\)](#)

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

### 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	451/916 (49%)	0.21	9 (1%) 65 60	55, 86, 159, 214	0
1	B	447/916 (48%)	0.50	29 (6%) 18 14	60, 113, 179, 212	0
All	All	898/1832 (49%)	0.35	38 (4%) 36 29	55, 102, 171, 214	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	371	ASP	12.8
1	B	338	GLY	9.4
1	B	339	VAL	7.8
1	B	340	ARG	7.4
1	B	537	GLN	7.1
1	B	548	TYR	4.4
1	A	594	VAL	4.1
1	B	538	PHE	3.8
1	B	664	ARG	3.5
1	B	63	GLY	3.5
1	A	679	TRP	3.4
1	B	373	LEU	3.4
1	A	548	TYR	3.3
1	A	371	ASP	3.0
1	B	191	PHE	3.0
1	A	592	VAL	3.0
1	B	695	ILE	2.8
1	B	547	LEU	2.7
1	B	225	TYR	2.7
1	A	665	SER	2.7
1	A	538	PHE	2.7
1	B	333	LYS	2.6
1	B	549	THR	2.5
1	B	539	VAL	2.4

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	545	LYS	2.4
1	B	306	ARG	2.3
1	B	212	PHE	2.3
1	B	653	MET	2.2
1	A	693	LEU	2.2
1	B	679	TRP	2.2
1	B	259	ILE	2.1
1	B	303	LEU	2.1
1	B	621	VAL	2.1
1	B	156	ARG	2.1
1	A	685	LEU	2.0
1	B	127	ARG	2.0
1	B	171	LEU	2.0
1	B	540	PRO	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](#)

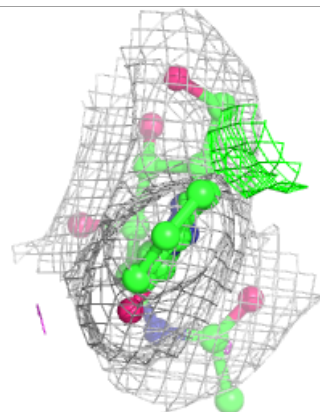
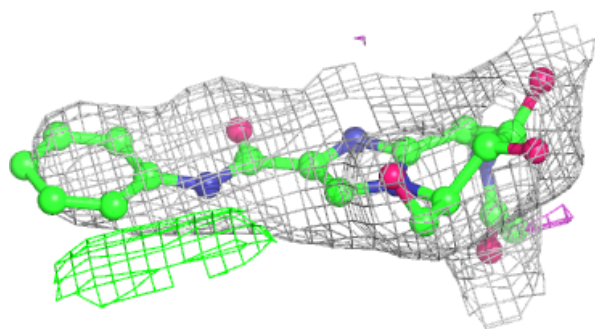
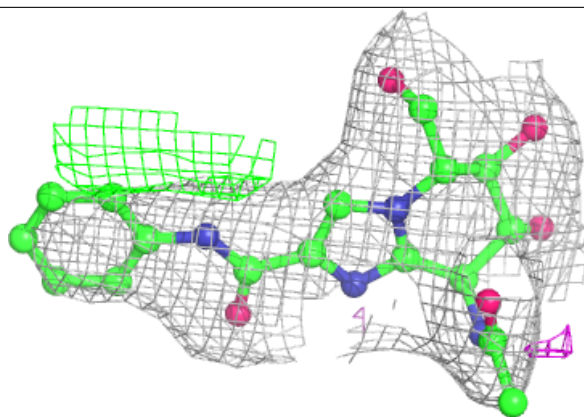
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q<0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	GDV	B	1001	26/26	0.96	0.15	73,84,133,135	0
2	GDV	A	1001	26/26	0.97	0.17	62,77,117,120	0

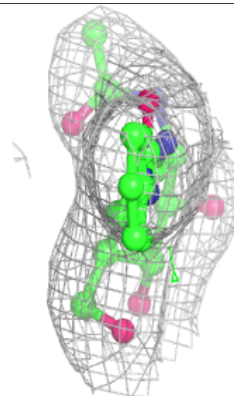
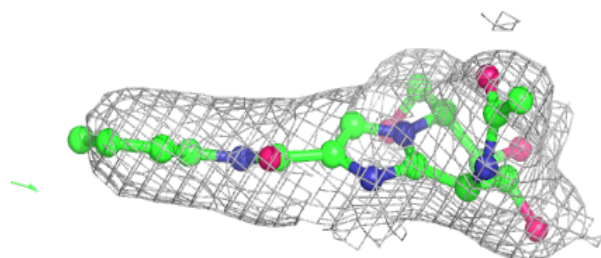
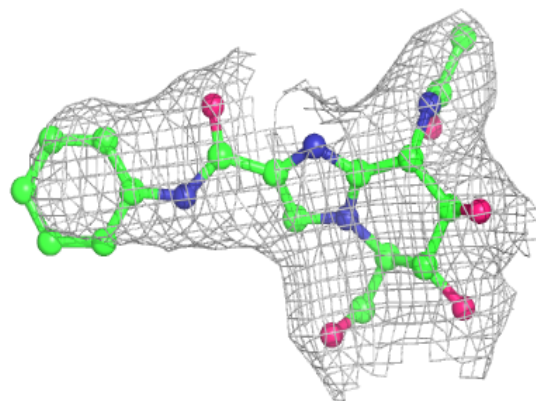
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

**Electron density around GDV B 1001:**

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

**Electron density around GDV A 1001:**

$2mF_o-DF_c$  (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.