Full wwPDB X-ray Structure Validation Report ⓘ

May 15, 2020 – 11:59 pm BST

PDB ID	:	5MPQ
Title	:	Bulgecin A: The key to a broad-spectrum inhibitor that targets lytic transglycosylases
Authors	:	Williams, A.H.
Deposited on	:	2016-12-17
Resolution	:	1.78 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*
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:

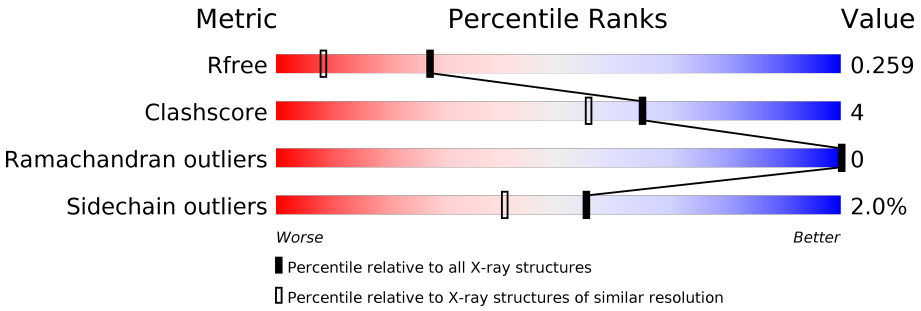
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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)

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 >=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 <=5%

Mol	Chain	Length	Quality of chain
1	A	576	92% 7% •

2 Entry composition

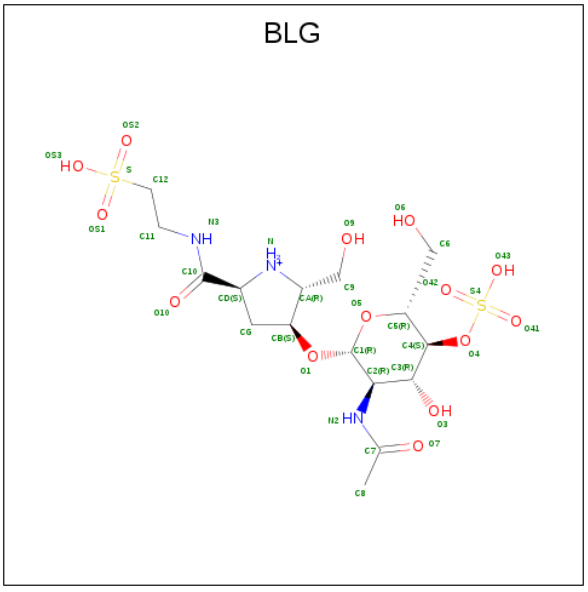
There are 5 unique types of molecules in this entry. The entry contains 4948 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Transglycosylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	576	4517	2808	838	856	15	0	5	0

- Molecule 2 is 4-O-(4-O-SULFONYL-N-ACETYLGLUCOSAMININYL)-5-METHYLHYDROXY-L-PROLINE-TAURINE (three-letter code: BLG) (formula: $C_{16}H_{30}N_3O_{14}S_2$).

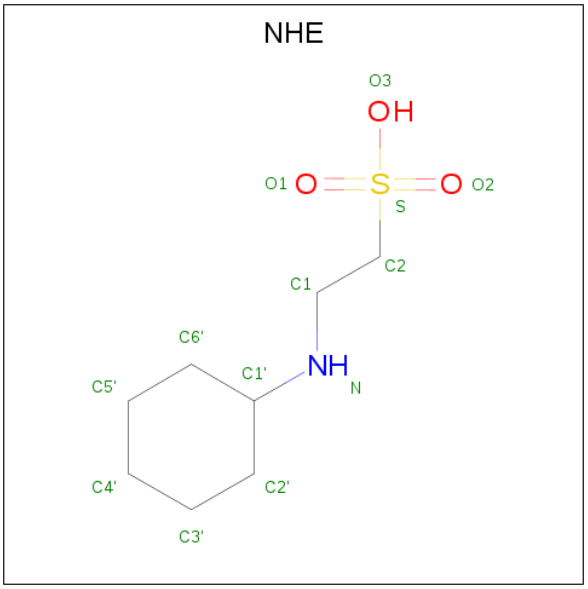


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	35	16	3	14	2	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
3	A	2	2	2	0	0

- Molecule 4 is 2-[N-CYCLOHEXYLAMINO]ETHANE SULFONIC ACID (three-letter code: NHE) (formula: $C_8H_{17}NO_3S$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	S	0	0
			12	7	1	3	1		

- Molecule 5 is water.

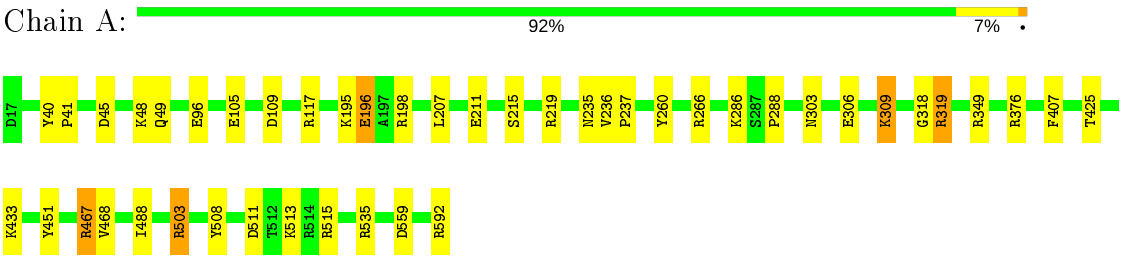
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	382	Total	O	0	0
			382	382		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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: Transglycosylase

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	65.83Å 71.56Å 121.12Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.84 – 1.78 48.45 – 1.78	Depositor EDS
% Data completeness (in resolution range)	84.0 (45.84-1.78) 97.6 (48.45-1.78)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.00 (at 1.78Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	(Not available) , (Not available) 0.206 , 0.259	Depositor DCC
R_{free} test set	2719 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	21.7	Xtriage
Anisotropy	0.481	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 42.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4948	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.73% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NHE, BLG, MG

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	0.39	0/4607	0.52	0/6234

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

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	4517	0	4428	36	0
2	A	35	0	30	3	0
3	A	2	0	0	0	0
4	A	12	0	11	6	0
5	A	382	0	0	8	1
All	All	4948	0	4469	37	1

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 (37) 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:211:GLU:HG3	1:A:219:ARG:HD2	1.58	0.86
4:A:604:NHE:C4'	4:A:604:NHE:C2'	2.54	0.84
1:A:319:ARG:H	4:A:604:NHE:HC22	1.46	0.80
1:A:96:GLU:OE2	5:A:701:HOH:O	2.10	0.69
1:A:349:ARG:NH1	5:A:706:HOH:O	2.25	0.68
1:A:105:GLU:OE1	1:A:117:ARG:NH2	2.33	0.61
1:A:109:ASP:OD2	1:A:117:ARG:NH1	2.30	0.58
1:A:45:ASP:HA	1:A:48:LYS:HD3	1.86	0.57
1:A:513:LYS:NZ	5:A:702:HOH:O	2.12	0.55
1:A:515:ARG:O	1:A:535[A]:ARG:NH2	2.29	0.54
1:A:488:ILE:HD12	1:A:503:ARG:NH1	2.22	0.54
1:A:235:ASN:OD1	5:A:703:HOH:O	2.18	0.53
1:A:433:LYS:HA	1:A:451:TYR:OH	2.08	0.53
1:A:319:ARG:N	4:A:604:NHE:HC22	2.22	0.53
1:A:309:LYS:NZ	5:A:726:HOH:O	2.43	0.52
1:A:508:TYR:CD2	2:A:601:BLG:H83	2.45	0.52
1:A:196[A]:GLU:H	1:A:196[A]:GLU:CD	2.16	0.49
1:A:559[A]:ASP:OD2	5:A:704:HOH:O	2.20	0.47
1:A:318:GLY:HA3	4:A:604:NHE:HC21	1.96	0.47
1:A:288:PRO:HG2	4:A:604:NHE:H5'2	1.96	0.46
1:A:467:ARG:HE	2:A:601:BLG:H112	1.80	0.46
1:A:195:LYS:HD2	1:A:198:ARG:CZ	2.45	0.46
1:A:535[B]:ARG:O	1:A:592:ARG:HD3	2.15	0.45
1:A:511:ASP:O	1:A:515:ARG:HG3	2.17	0.45
1:A:376:ARG:HD3	5:A:962:HOH:O	2.16	0.45
1:A:40:TYR:HB3	1:A:41:PRO:HD3	1.99	0.45
1:A:303:ASN:HB3	1:A:306:GLU:CD	2.37	0.44
1:A:207:LEU:O	1:A:211:GLU:HB2	2.18	0.44
1:A:109:ASP:OD2	1:A:117:ARG:HD2	2.18	0.44
1:A:467:ARG:HG2	1:A:468:VAL:N	2.34	0.43
1:A:236:VAL:HB	1:A:237:PRO:HD3	2.00	0.43
1:A:286:LYS:NZ	5:A:745:HOH:O	2.52	0.42
1:A:535[A]:ARG:O	1:A:592:ARG:HD3	2.19	0.42
1:A:215:SER:O	1:A:219:ARG:HG3	2.20	0.42
1:A:467:ARG:HD3	2:A:601:BLG:OS2	2.20	0.41
1:A:45:ASP:O	1:A:49:GLN:HG2	2.20	0.41
1:A:319:ARG:HH21	4:A:604:NHE:HO3	1.69	0.41

All (1) 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)
5:A:741:HOH:O	5:A:1033:HOH:O[3_545]	2.18	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	A	579/576 (100%)	574 (99%)	5 (1%)	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	446/442 (101%)	435 (98%)	11 (2%)	47 31

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	196[A]	GLU
1	A	196[B]	GLU
1	A	260	TYR
1	A	266	ARG
1	A	309	LYS
1	A	319	ARG
1	A	407	PHE
1	A	425[A]	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	425[B]	THR
1	A	467	ARG
1	A	503	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	235	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	$\#\|Z\| > 2$
4	NHE	A	604	-	11,11,13	1.28	2 (18%)	13,14,17	2.03	4 (30%)
2	BLG	A	601	-	36,36,36	3.16	14 (38%)	41,53,53	1.88	13 (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
4	NHE	A	604	-	-	5/10/10/15	-
2	BLG	A	601	-	-	0/28/60/60	0/2/2/2

All (16) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	BLG	CB-CA	-8.39	1.36	1.53
2	A	601	BLG	CG-CD	-7.75	1.37	1.54
2	A	601	BLG	C10-N3	6.59	1.48	1.33
2	A	601	BLG	CA-N	6.33	1.57	1.48
2	A	601	BLG	CG-CB	5.23	1.64	1.52
2	A	601	BLG	CD-N	4.57	1.56	1.46
2	A	601	BLG	C12-S	3.81	1.82	1.77
2	A	601	BLG	O4-S4	3.74	1.68	1.57
2	A	601	BLG	C7-N2	3.38	1.46	1.34
4	A	604	NHE	O2-S	2.90	1.53	1.45
2	A	601	BLG	OS2-S	2.76	1.53	1.45
4	A	604	NHE	O1-S	2.58	1.52	1.45
2	A	601	BLG	O5-C1	2.48	1.48	1.41
2	A	601	BLG	C8-C7	2.13	1.55	1.50
2	A	601	BLG	OS1-S	2.10	1.51	1.45
2	A	601	BLG	C2-N2	2.03	1.49	1.45

All (17) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	604	NHE	O3-S-C2	4.55	113.12	105.77
2	A	601	BLG	C1-C2-N2	-4.30	103.60	111.00
2	A	601	BLG	OS2-S-C12	4.15	111.91	106.92
2	A	601	BLG	CG-CD-C10	3.42	116.31	111.19
4	A	604	NHE	O2-S-O1	-3.37	102.29	113.95
2	A	601	BLG	OS2-S-OS1	-3.33	102.41	113.95
2	A	601	BLG	C10-CD-N	-3.05	106.22	111.74
2	A	601	BLG	C2-N2-C7	-2.96	115.97	123.18
2	A	601	BLG	C9-CA-N	-2.84	106.37	111.46
2	A	601	BLG	C12-C11-N3	-2.78	103.17	111.30
4	A	604	NHE	O2-S-C2	2.72	110.19	106.92
2	A	601	BLG	OS1-S-C12	2.59	110.04	106.92
2	A	601	BLG	O42-S4-O41	-2.51	102.15	112.22

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	604	NHE	C5'-C6'-C1'	-2.36	110.39	114.19
2	A	601	BLG	C1-O5-C5	-2.28	109.20	113.69
2	A	601	BLG	C1-C2-C3	2.11	116.26	110.06
2	A	601	BLG	CG-CB-CA	-2.07	100.86	103.80

There are no chirality outliers.

All (5) torsion outliers are listed below:

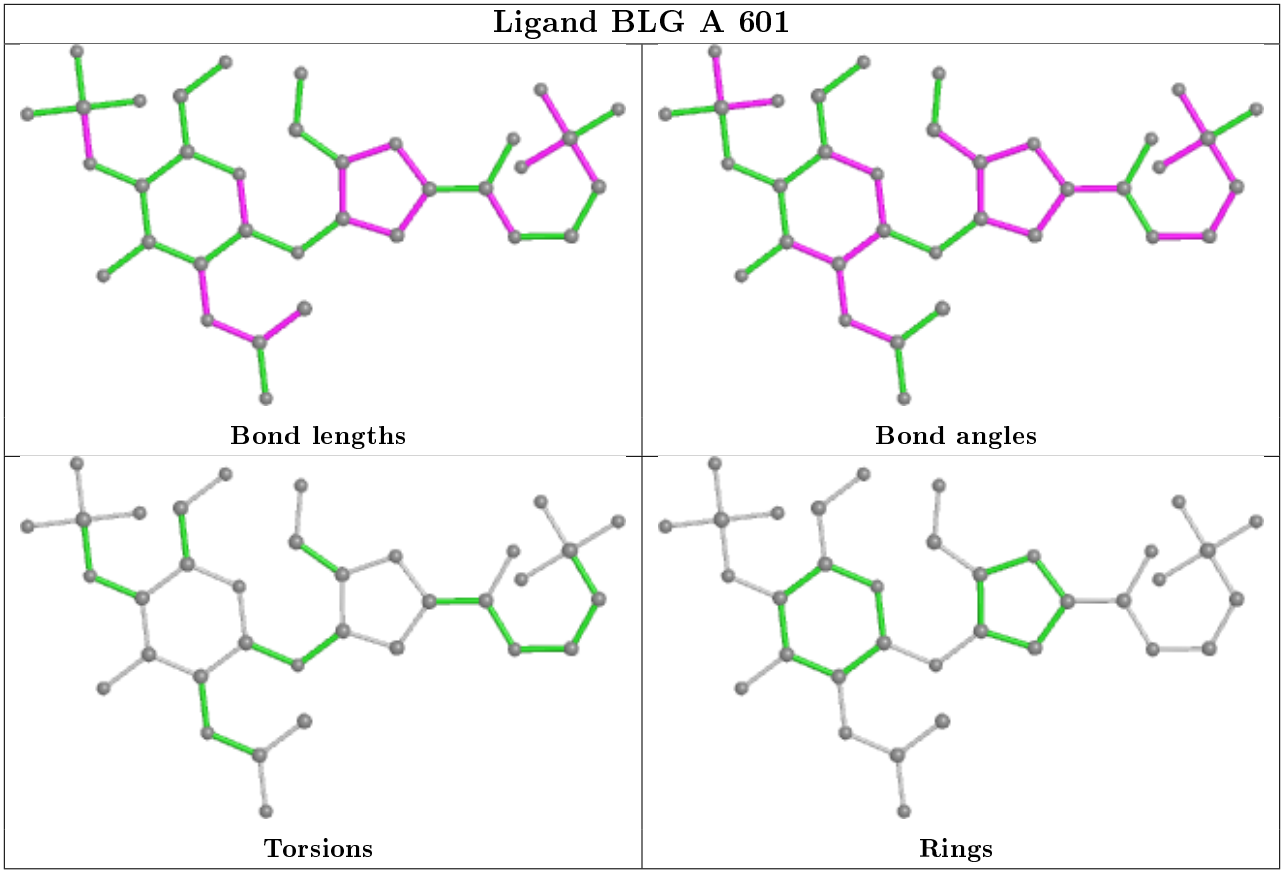
Mol	Chain	Res	Type	Atoms
4	A	604	NHE	C2'-C1'-N-C1
4	A	604	NHE	C6'-C1'-N-C1
4	A	604	NHE	C4'-C5'-C6'-C1'
4	A	604	NHE	C2-C1-N-C1'
4	A	604	NHE	N-C1'-C6'-C5'

There are no ring outliers.

2 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	604	NHE	6	0
2	A	601	BLG	3	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 then 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

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.

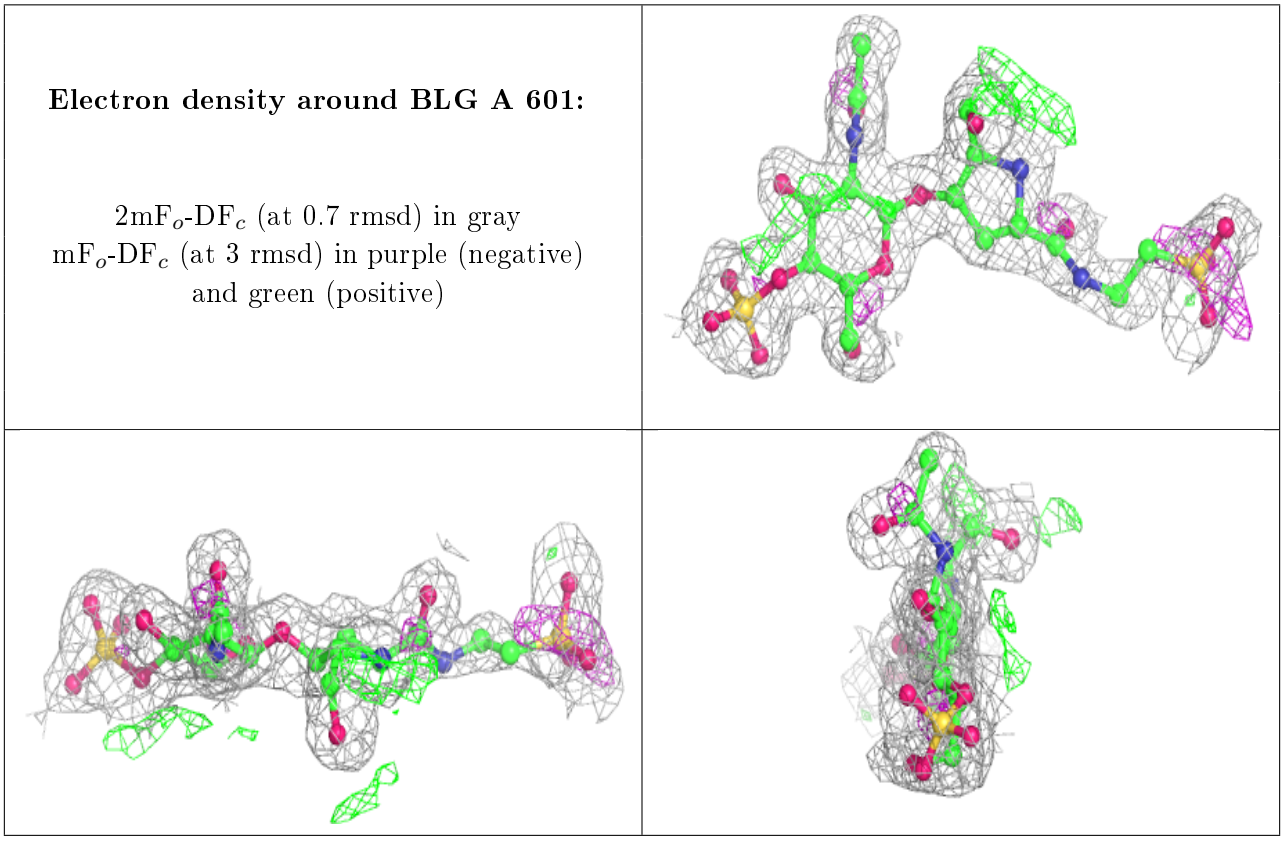
6.4 Ligands

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.

Electron density around BLG A 601:

$2mF_o$-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)

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