wwPDB X-ray Structure Validation Summary Report (i)

Nov 15, 2023 – 10:37 PM JST

PDB ID	:	6K35
Title	:	Crystal structure of GH20 exo beta-N-acetylglucosaminidase from Vibrio harveyi in complex with NAG-thiazoline
Authors	:	Meekrathok, P.; Stubbs, K.A.; Bulmer, D.M.; van den Berg, B.; Suginta, W.
Deposited on	:	2019-05-16
Resolution	:	2.36 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (i)) were used in the production of this report:

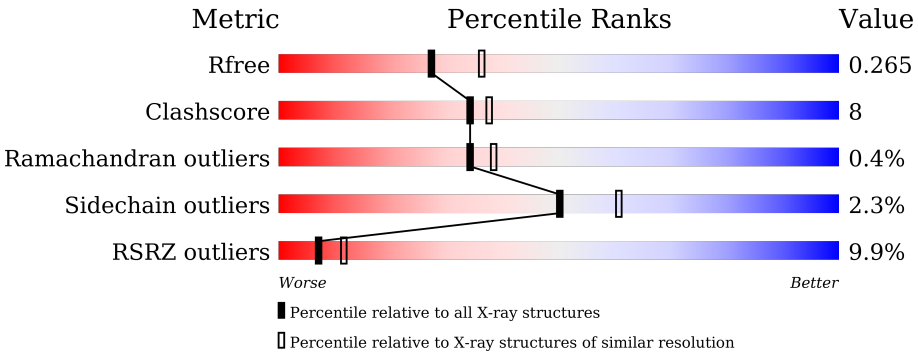
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.36 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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 >=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% 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	652	8% 78% 19%
1	B	652	11% 78% 19%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 20758 atoms, of which 9991 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.

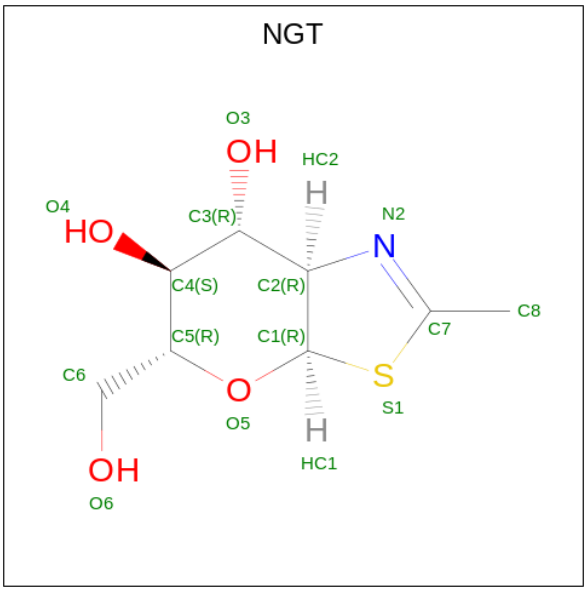
- Molecule 1 is a protein called Beta-N-acetylglucosaminidase Nag2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	640	Total	C	H	N	O	S	0	0	0
			10125	3273	4980	880	970	22			
1	B	640	Total	C	H	N	O	S	0	0	0
			10132	3275	4985	880	970	22			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	643	ARG	-	expression tag	UNP D9ISE0
A	644	SER	-	expression tag	UNP D9ISE0
A	645	ARG	-	expression tag	UNP D9ISE0
A	646	SER	-	expression tag	UNP D9ISE0
A	647	HIS	-	expression tag	UNP D9ISE0
A	648	HIS	-	expression tag	UNP D9ISE0
A	649	HIS	-	expression tag	UNP D9ISE0
A	650	HIS	-	expression tag	UNP D9ISE0
A	651	HIS	-	expression tag	UNP D9ISE0
A	652	HIS	-	expression tag	UNP D9ISE0
B	643	ARG	-	expression tag	UNP D9ISE0
B	644	SER	-	expression tag	UNP D9ISE0
B	645	ARG	-	expression tag	UNP D9ISE0
B	646	SER	-	expression tag	UNP D9ISE0
B	647	HIS	-	expression tag	UNP D9ISE0
B	648	HIS	-	expression tag	UNP D9ISE0
B	649	HIS	-	expression tag	UNP D9ISE0
B	650	HIS	-	expression tag	UNP D9ISE0
B	651	HIS	-	expression tag	UNP D9ISE0
B	652	HIS	-	expression tag	UNP D9ISE0

- Molecule 2 is 3AR,5R,6S,7R,7AR-5-HYDROXYMETHYL-2-METHYL-5,6,7,7A-TETRAHYDRO-3AH-PYRANO[3,2-D]THIAZOLE-6,7-DIOL (three-letter code: NGT) (formula: $C_8H_{13}NO_4S$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	S	0	0
			27	8	13	1	4	1		
2	B	1	Total	C	H	N	O	S	0	0
			27	8	13	1	4	1		

- Molecule 3 is water.

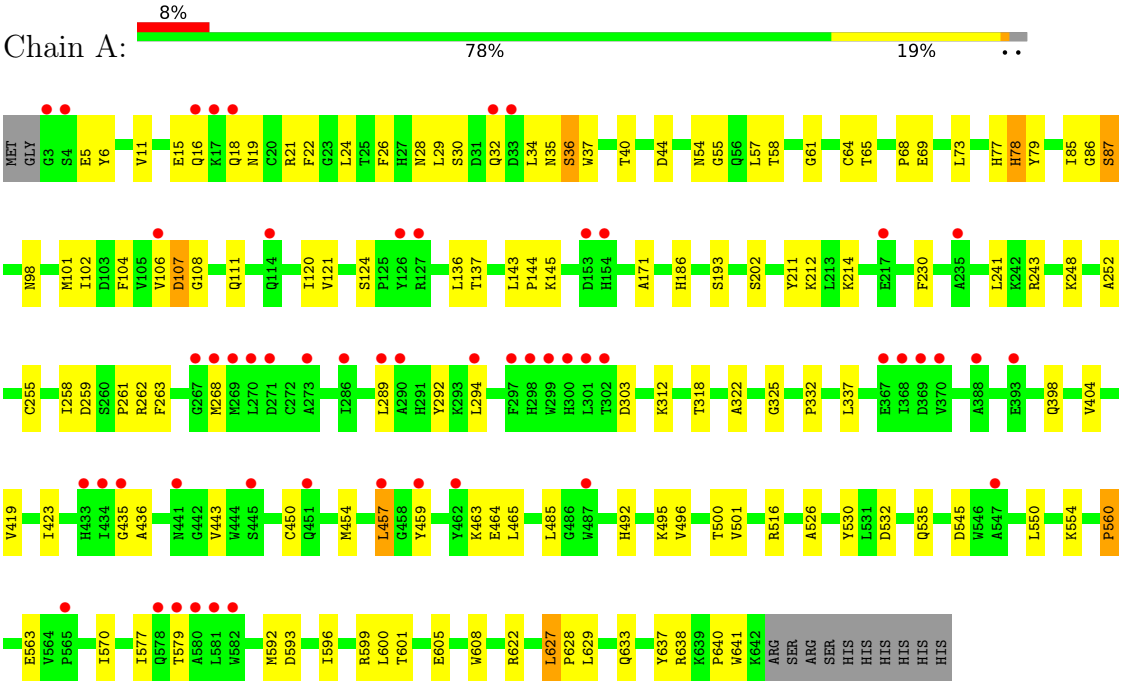
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	237	Total	O	0	0
			237	237		
3	B	210	Total	O	0	0
			210	210		

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.

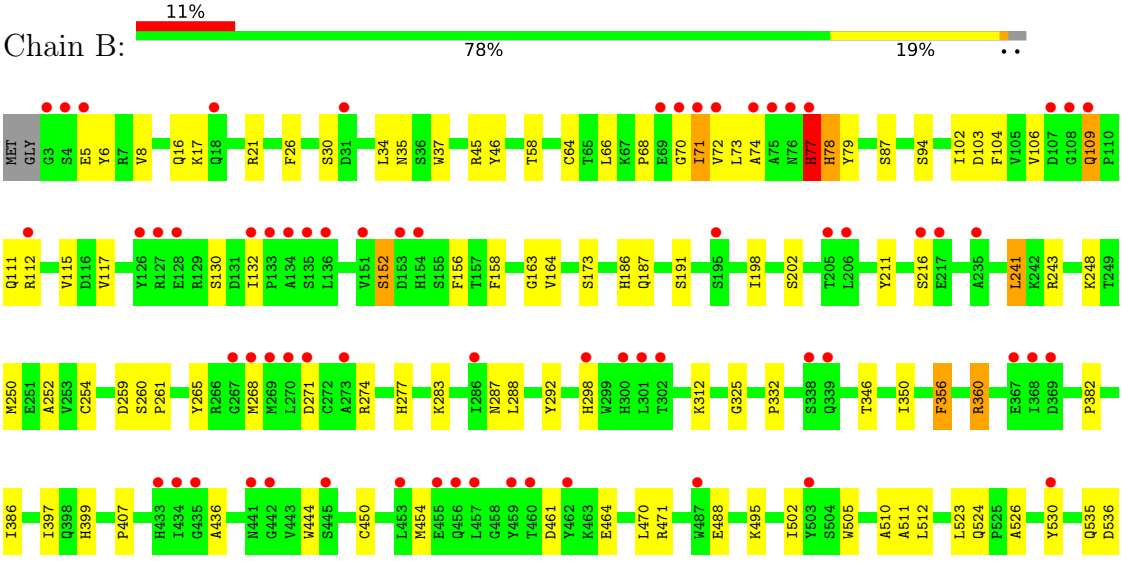
• Molecule 1: Beta-N-acetylglucosaminidase Nag2

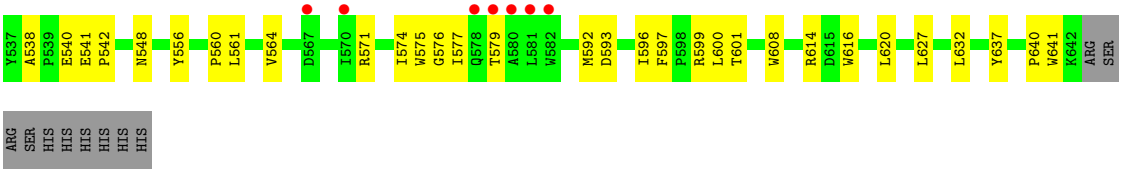
Chain A:



• Molecule 1: Beta-N-acetylglucosaminidase Nag2

Chain B:




Y537
A538
P539
E540
E541
P542
N548
Y556
P560
L561
V564
D567
I570
R571
I574
W575
G576
I577
Q578
T579
A580
L581
W582
M592
D593
I596
F597
P598
R599
L600
T601
W608
R614
D615
W616
L620
L627
L632
Y637
P640
W641
K642
ARG
SER
ARG
SER
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	62.34Å 145.04Å 88.68Å 90.00° 109.66° 90.00°	Depositor
Resolution (Å)	29.35 – 2.36 29.35 – 2.36	Depositor EDS
% Data completeness (in resolution range)	99.7 (29.35-2.36) 99.7 (29.35-2.36)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.58 (at 2.36Å)	Xtriage
Refinement program	PHENIX 1.8.2-1309	Depositor
R, R_{free}	0.194 , 0.265 0.194 , 0.265	Depositor DCC
R_{free} test set	2943 reflections (4.85%)	wwPDB-VP
Wilson B-factor (Å^2)	34.1	Xtriage
Anisotropy	0.696	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 46.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.021 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	20758	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.14% 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: NGT

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	0.63	3/5282 (0.1%)	0.82	8/7182 (0.1%)
1	B	0.64	2/5284 (0.0%)	0.79	4/7185 (0.1%)
All	All	0.64	5/10566 (0.0%)	0.80	12/14367 (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	B	0	2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	356	PHE	CE1-CZ	-8.33	1.21	1.37
1	A	563	GLU	CD-OE1	-6.91	1.18	1.25
1	A	563	GLU	CG-CD	6.84	1.62	1.51
1	B	356	PHE	CD2-CE2	-6.30	1.26	1.39
1	A	450	CYS	CB-SG	-5.37	1.73	1.81

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	563	GLU	OE1-CD-OE2	-10.82	110.31	123.30
1	A	545	ASP	CB-CG-OD1	6.78	124.40	118.30
1	B	360	ARG	NE-CZ-NH1	6.66	123.63	120.30
1	A	593	ASP	CB-CG-OD1	6.43	124.09	118.30
1	A	457	LEU	CB-CG-CD1	-5.75	101.22	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	77	HIS	Mainchain,Sidechain

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	5145	4980	4976	84	0
1	B	5147	4985	4983	87	0
2	A	14	13	11	0	0
2	B	14	13	11	0	0
3	A	237	0	0	0	0
3	B	210	0	0	0	0
All	All	10767	9991	9981	171	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 8.

The worst 5 of 171 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:30:SER:OG	1:A:32:GLN:OE1	1.67	1.10
1:A:26:PHE:CZ	1:A:102:ILE:HD12	1.91	1.03
1:A:26:PHE:HZ	1:A:102:ILE:HD12	1.38	0.84
1:B:70:GLY:O	1:B:71:ILE:HG13	1.78	0.83
1:A:37:TRP:CE2	1:A:68:PRO:HG3	2.16	0.81

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	638/652 (98%)	613 (96%)	23 (4%)	2 (0%)	41	47
1	B	638/652 (98%)	612 (96%)	23 (4%)	3 (0%)	29	32
All	All	1276/1304 (98%)	1225 (96%)	46 (4%)	5 (0%)	34	38

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	107	ASP
1	B	71	ILE
1	A	36	SER
1	B	87	SER
1	B	77	HIS

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	555/567 (98%)	544 (98%)	11 (2%)	55	66
1	B	556/567 (98%)	542 (98%)	14 (2%)	47	58
All	All	1111/1134 (98%)	1086 (98%)	25 (2%)	50	61

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	78	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	173	SER
1	B	632	LEU
1	B	152	SER
1	B	202	SER

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

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	NGT	A	701	-	13,15,15	13.44	8 (61%)	12,22,22	3.08	6 (50%)
2	NGT	B	701	-	13,15,15	12.67	9 (69%)	12,22,22	2.68	6 (50%)

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	NGT	A	701	-	-	0/2/30/30	0/2/2/2
2	NGT	B	701	-	-	0/2/30/30	0/2/2/2

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	701	NGT	C7-S1	-31.79	1.50	1.77
2	A	701	NGT	C7-N2	31.44	1.56	1.27
2	B	701	NGT	C7-N2	31.22	1.55	1.27
2	B	701	NGT	C7-S1	-28.26	1.53	1.77
2	A	701	NGT	O5-C1	13.18	1.62	1.42

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	701	NGT	O5-C1-C2	-7.86	97.99	115.27
2	B	701	NGT	O5-C1-C2	-5.49	103.20	115.27
2	B	701	NGT	O3-C3-C4	-4.31	100.38	110.35
2	A	701	NGT	O5-C5-C6	4.03	116.44	106.44
2	A	701	NGT	O3-C3-C2	3.72	117.65	109.14

There are no chirality outliers.

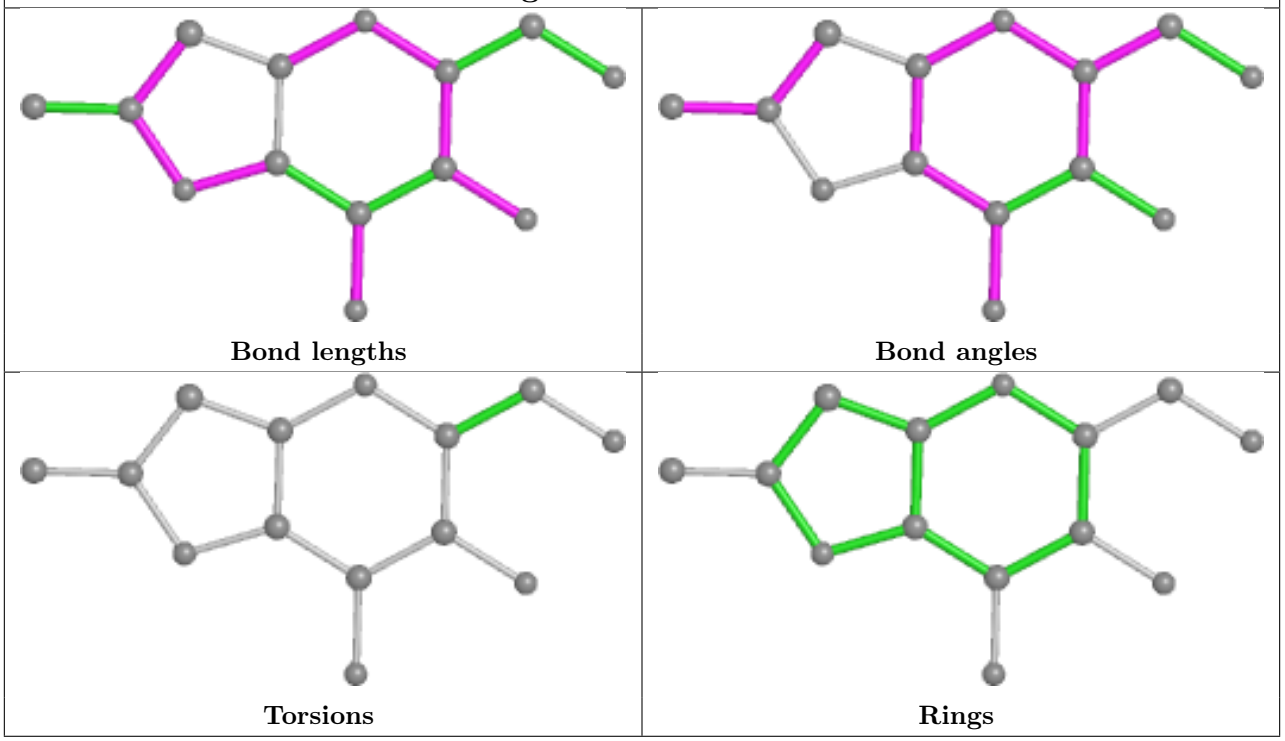
There are no torsion outliers.

There are no ring outliers.

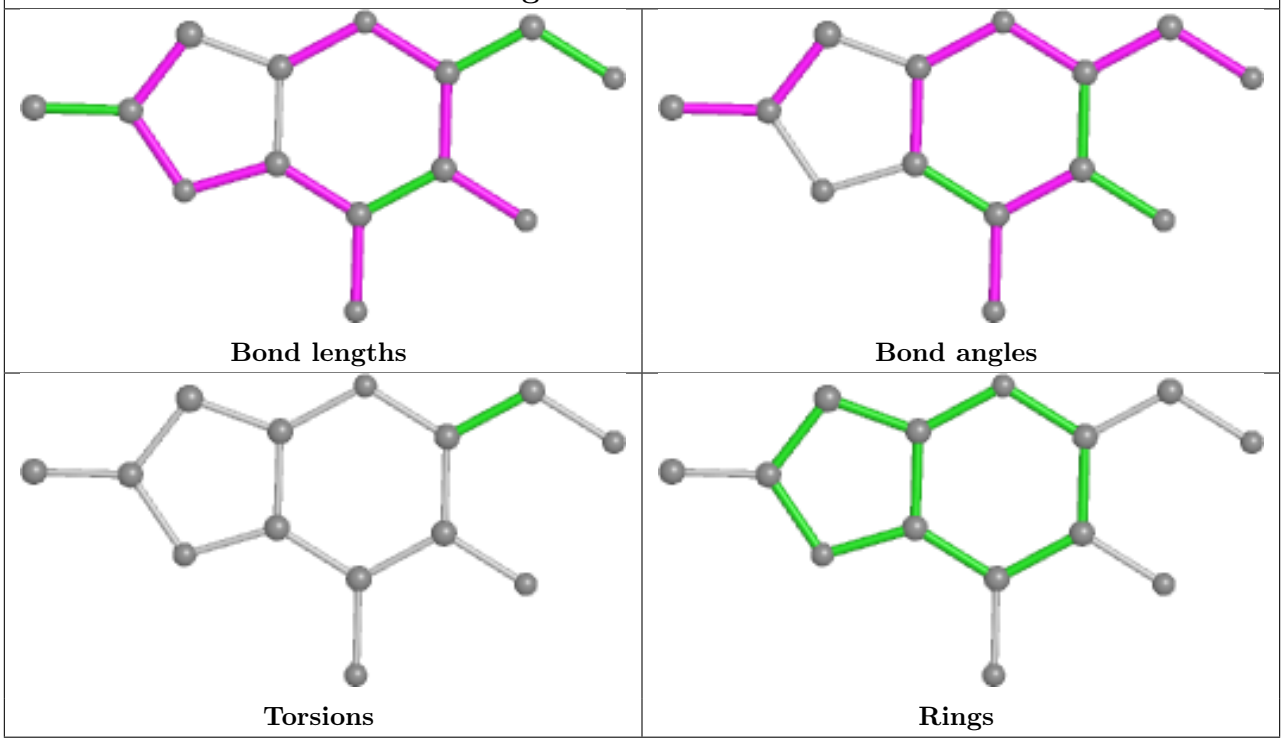
No monomer is involved in short contacts.

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 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.

Ligand NGT A 701



Ligand NGT B 701



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^{th} 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^2)	Q<0.9
1	A	640/652 (98%)	0.54	54 (8%) 11 16	24, 38, 57, 99	0
1	B	640/652 (98%)	0.66	73 (11%) 5 7	25, 39, 64, 115	0
All	All	1280/1304 (98%)	0.60	127 (9%) 7 11	24, 39, 60, 115	0

The worst 5 of 127 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	71	ILE	7.1
1	B	3	GLY	6.1
1	B	69	GLU	5.1
1	A	17	LYS	5.0
1	B	108	GLY	4.7

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

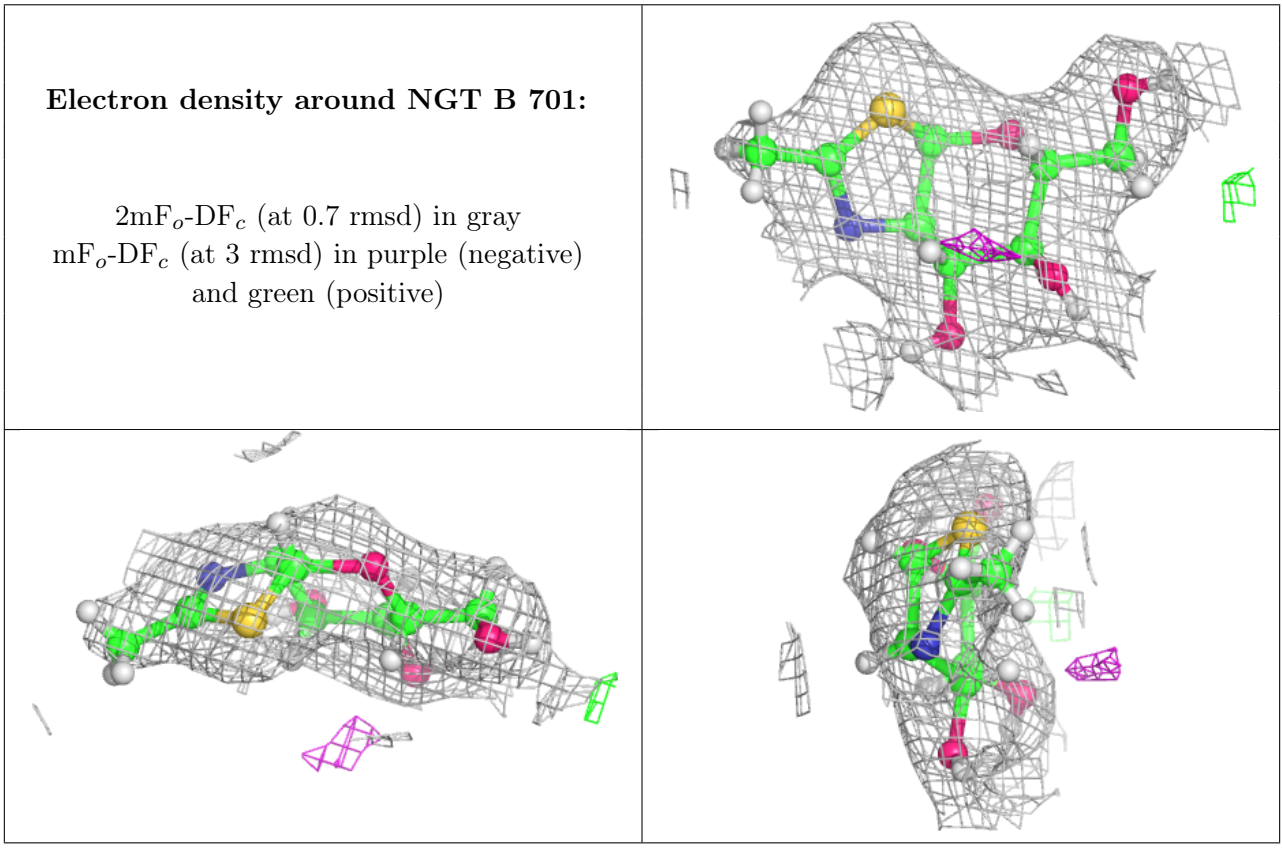
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	NGT	B	701	14/14	0.96	0.23	29,35,42,43	0
2	NGT	A	701	14/14	0.97	0.20	25,32,45,45	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.

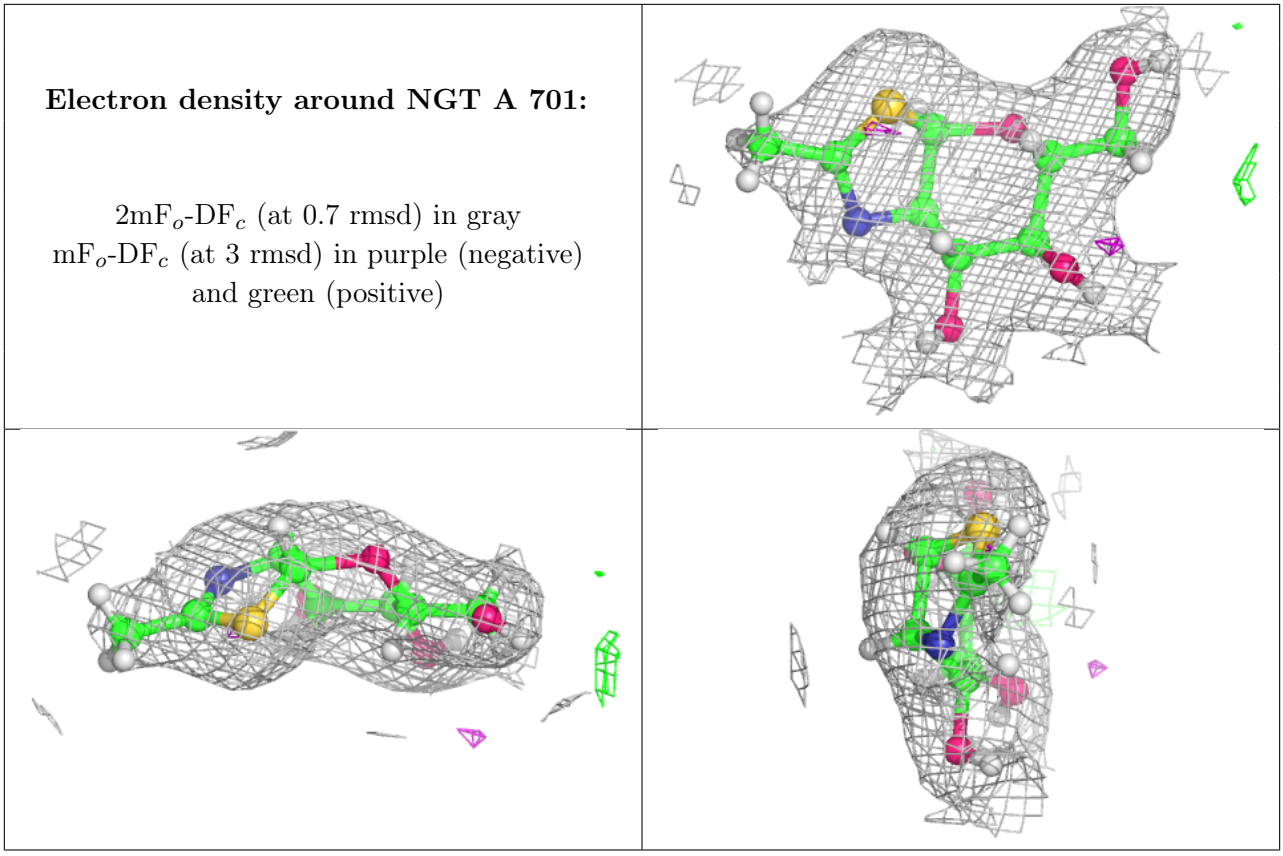
Electron density around NGT B 701:

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



Electron density around NGT A 701:

$2mF_o$-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.