

# Full wwPDB X-ray Structure Validation Report (i)

### Oct 23, 2021 – 04:34 PM EDT

PDB ID	:	1D0M
Title	:	THE ESCHERICHIA COLI LYTIC TRANSGLYCOSYLASE SLT35 IN
		COMPLEX WITH BULGECIN A AND (GLCNAC)2
Authors	:	van Asselt, E.J.; Kalk, K.H.; Dijkstra, B.W.
Deposited on		
Resolution	:	2.47  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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

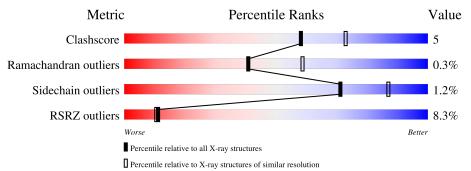
MolProbity Mogul Xtriage (Phenix)	:	4.02b-467 1.8.5 (274361), CSD as541be (2020) 1.13
EDS	:	2.23.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber $(2001)$
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 2.47 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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
Clashscore	141614	6594 (2.50-2.46)
Ramachandran outliers	138981	6469 (2.50-2.46)
Sidechain outliers	138945	6471 (2.50-2.46)
RSRZ outliers	127900	5738 (2.50-2.46)

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	А	322	8%	12%	·				
2	В	2	100%						



#### 1D0M

# 2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2772 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 35KD SOLUBLE LYTIC TRANSGLYCOSYLASE.

Mol	Chain	Residues		At	oms			ZeroOcc	AltConf	Trace
1	А	312	Total 2507	C 1596	N 439	0 461	S 11	0	7	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	40	MET	LEU	engineered mutation	UNP P41052
А	41	VAL	LEU	engineered mutation	UNP P41052

• Molecule 2 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-a cetamido-2-deoxy-alpha-D-glucopyranose.



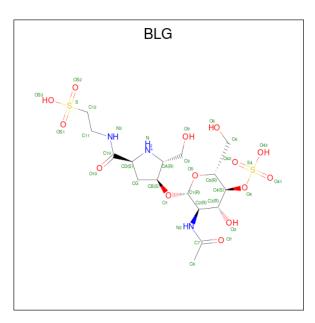
Mol	Chain	Residues	A	Aton	ns		ZeroOcc	AltConf	Trace
2	В	2	Total 29	C 16	N 2	0 11	0	0	0

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	А	1	Total Ca 1 1	0	0

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





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
4	А	1	Total 35	C 16	N 3	0 14	$\begin{array}{c} \mathrm{S} \\ \mathrm{2} \end{array}$	0	0

• Molecule 5 is water.

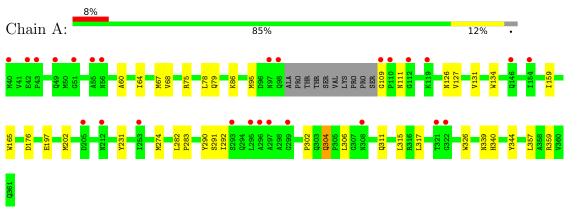
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	200	Total         O           200         200	0	0



# 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: 35KD SOLUBLE LYTIC TRANSGLYCOSYLASE



• Molecule 2: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-alpha-D-glucopyranose

Chain B:

100%

NDG 1 NAG 2



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	58.63Å 67.75Å 97.76Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	20.00 - 2.47	Depositor
Resolution (A)	23.57 - 2.48	EDS
% Data completeness	89.1 (20.00-2.47)	Depositor
(in resolution range)	89.7(23.57-2.48)	EDS
R <sub>merge</sub>	0.04	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$5.93 (at 2.47 \text{\AA})$	Xtriage
Refinement program	X-PLOR 3.843	Depositor
D D.	0.174 , $0.241$	Depositor
$R, R_{free}$	0.178 , (Not available)	DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor $(Å^2)$	31.0	Xtriage
Anisotropy	0.308	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.39, 60.8	EDS
L-test for twinning <sup>2</sup>	$ \langle L  \rangle = 0.47, \langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2772	wwPDB-VP
Average B, all atoms $(Å^2)$	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 6.28% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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: NDG, CA, BLG, NAG

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	Bond	angles	
	loi Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5
1	А	0.35	0/2601	0.58	0/3528

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 (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	А	2507	0	2397	23	0
2	В	29	0	24	0	0
3	А	1	0	0	0	0
4	А	35	0	28	1	0
5	А	200	0	0	2	0
All	All	2772	0	2449	24	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 5.

All (24) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.



1D0M
------

Atom 1	Atom 2	Interatomic	Clash
Atom-1	Atom-2	distance $(\text{\AA})$	overlap (Å)
1:A:315:LEU:HB2	1:A:326:TRP:HB2	1.68	0.75
1:A:109:GLY:HA2	1:A:111:ASN:HD22	1.63	0.64
4:A:401:BLG:H121	5:A:596:HOH:O	2.05	0.56
1:A:134:TRP:CD2	1:A:359:ARG:HD2	2.41	0.56
1:A:64:ILE:HG23	1:A:78:LEU:HD23	1.87	0.56
1:A:304:GLN:HA	1:A:304:GLN:HE21	1.71	0.55
1:A:60:ALA:O	1:A:64:ILE:HG13	2.11	0.51
1:A:339:ASN:HB3	1:A:344:TYR:CD1	2.46	0.51
1:A:64:ILE:O	1:A:68:VAL:HG23	2.11	0.51
1:A:283:PRO:HD2	1:A:290:TYR:OH	2.10	0.50
1:A:95:MET:HE2	5:A:539:HOH:O	2.12	0.50
1:A:67:MET:SD	1:A:197:GLU:HG3	2.53	0.47
1:A:131:VAL:HG13	1:A:359:ARG:HG2	1.98	0.45
1:A:202[A]:MET:HG3	1:A:231:TYR:CZ	2.52	0.44
1:A:86:LYS:HE2	1:A:176:ASP:OD2	2.17	0.43
1:A:317:LEU:HD12	1:A:326:TRP:HE1	1.83	0.43
1:A:282:LEU:HA	1:A:283:PRO:HD3	1.85	0.42
1:A:292:ILE:HD12	1:A:306:LEU:HD13	2.00	0.42
1:A:75:ARG:O	1:A:79:GLN:HG3	2.21	0.41
1:A:291:SER:HA	1:A:311[A]:GLN:HG2	2.01	0.41
1:A:159:ILE:HG22	1:A:165:TRP:HA	2.02	0.41
1:A:127:VAL:O	1:A:131:VAL:HG23	2.20	0.41
1:A:302:PRO:HB2	1:A:304:GLN:O	2.20	0.41
1:A:357:LEU:HD23	1:A:357:LEU:HA	1.96	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	А	315/322~(98%)	308~(98%)	6~(2%)	1 (0%)	41	59

All (1) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	А	340	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	Analysed Rotameric		Percentiles	
1	А	260/263~(99%)	257~(99%)	3~(1%)	71 87	

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	126	ASN
1	А	274	MET
1	А	304	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	98	GLN
1	А	212	ASN
1	А	304	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)

2 monosaccharides 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).

Mal True		pe Chain Res Lin		Link Bond lengths			Bond angles			
	Mol Type Chain R	nes		Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2	
2	NDG	В	1	2	$15,\!15,\!15$	0.65	0	21,21,21	1.67	7 (33%)
2	NAG	В	2	2	14,14,15	1.18	1 (7%)	17,19,21	1.32	2 (11%)

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	NDG	В	1	2	-	0/6/26/26	0/1/1/1
2	NAG	В	2	2	-	0/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	В	2	NAG	C4-C5	3.17	1.59	1.53

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^{o})$	$Ideal(^{o})$
2	В	2	NAG	C1-O5-C5	3.34	116.72	112.19
2	В	1	NDG	C6-C5-C4	-3.04	105.88	113.00
2	В	1	NDG	C3-C2-N2	-2.91	105.12	110.62
2	В	1	NDG	C1-C2-N2	2.65	113.80	110.73
2	В	1	NDG	O4-C4-C3	-2.50	104.57	110.35
2	В	1	NDG	C1-C2-C3	2.37	113.77	110.54
2	В	1	NDG	O1-C1-C2	2.22	113.83	109.22
2	В	1	NDG	C4-C3-C2	2.11	113.43	110.34
2	В	2	NAG	O7-C7-C8	-2.05	118.25	122.06

There are no chirality outliers.

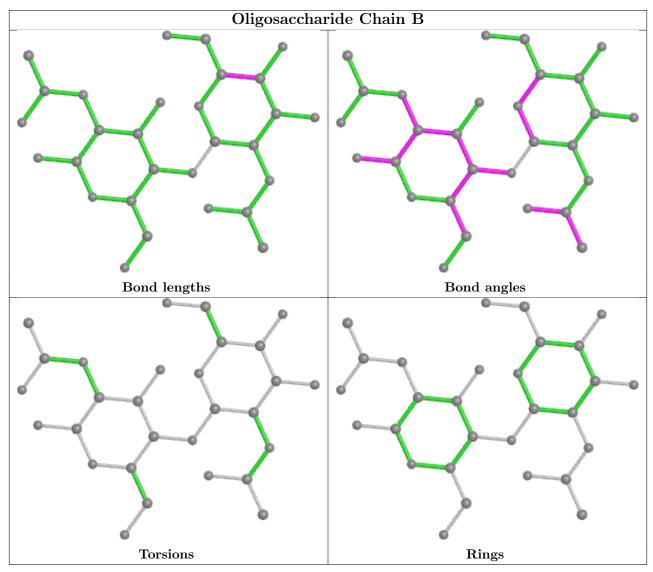
There are no torsion outliers.

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 oligosaccharide.



## 5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 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	Bo	ond leng	ths	Bond angles		
WIOI			nes	Link	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z  > 2
4	BLG	А	401	-	36,36,36	1.11	4 (11%)	41,53,53	1.36	5 (12%)

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	BLG	А	401	-	-	4/28/60/60	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$\operatorname{Observed}(\operatorname{\AA})$	Ideal(Å)
4	А	401	BLG	C12-S	3.21	1.82	1.77
4	А	401	BLG	O4-C4	-2.73	1.40	1.46
4	А	401	BLG	O1-CB	-2.71	1.40	1.44
4	А	401	BLG	CD-N	2.42	1.51	1.46

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^{o})$	$Ideal(^{o})$
4	А	401	BLG	OS2-S-C12	5.50	113.54	106.92
4	А	401	BLG	CG-CD-C10	2.77	115.33	111.19
4	А	401	BLG	OS2-S-OS1	-2.55	105.11	113.95
4	А	401	BLG	OS1-S-C12	2.24	109.61	106.92
4	А	401	BLG	C1-C2-N2	-2.08	107.43	111.00

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	А	401	BLG	N3-C11-C12-S
4	А	401	BLG	C11-C12-S-OS1
4	А	401	BLG	C11-C12-S-OS3
4	А	401	BLG	C11-C12-S-OS2

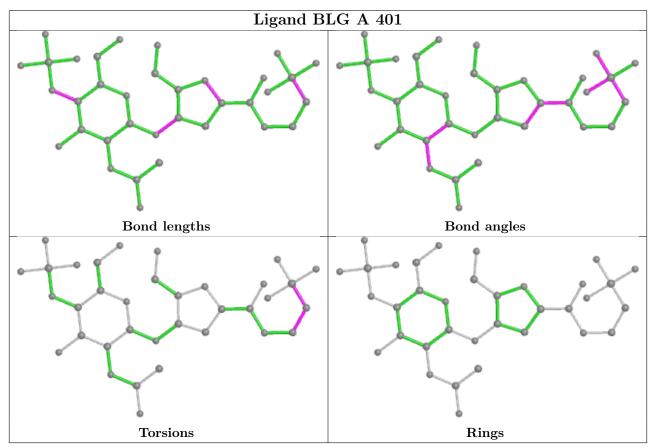
There are no ring outliers.

1 monomer is involved in 1 short contact:



Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	А	401	BLG	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 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 sufficient must be 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,  $95^{th}$  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	$\begin{tabular}{lllllllllllllllllllllllllllllllllll$		$\mathbf{OWAB}(\mathrm{\AA}^2)$	Q<0.9	
1	А	312/322~(96%)	0.22	26 (8%) 11 10	12, 27, 59, 77	0

All (26) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	296	ALA	5.2
1	А	42	GLU	4.6
1	А	321	THR	4.6
1	А	49	GLN	4.5
1	А	297	ALA	4.4
1	А	97	ASN	4.4
1	А	110	PRO	4.0
1	А	205	ASP	4.0
1	А	109	GLY	4.0
1	А	299	GLY	4.0
1	А	40	MET	3.8
1	А	308	ASN	3.3
1	А	146	GLN	3.3
1	А	212	ASN	3.1
1	А	112	GLY	2.8
1	А	322	GLY	2.6
1	А	154	ILE	2.5
1	А	119	LYS	2.4
1	А	295	LEU	2.3
1	А	293	SER	2.3
1	А	56	ASN	2.2
1	А	51	GLY	2.2
1	А	55	ALA	2.1
1	А	98	GLN	2.1
1	А	43	PRO	2.0
1	А	253	ILE	2.0



### 6.2 Non-standard residues in protein, DNA, RNA chains (i)

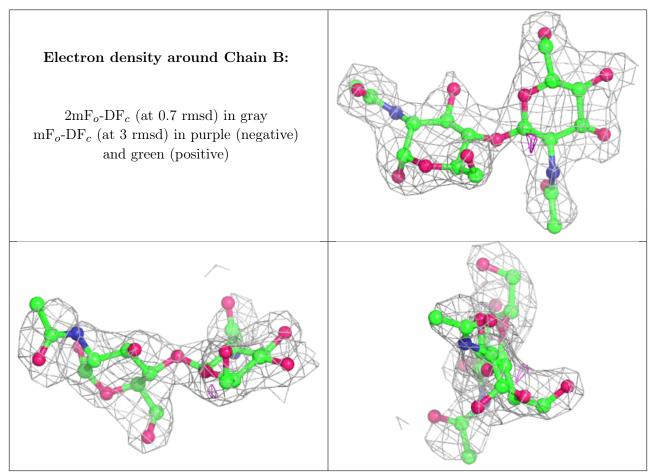
There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates (i)

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	$\mathbf{B} ext{-factors}(\mathrm{\AA}^2)$	Q < 0.9
2	NAG	В	2	14/15	0.91	0.19	31,42,48,52	0
2	NDG	В	1	15/15	0.92	0.19	40,49,59,60	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.



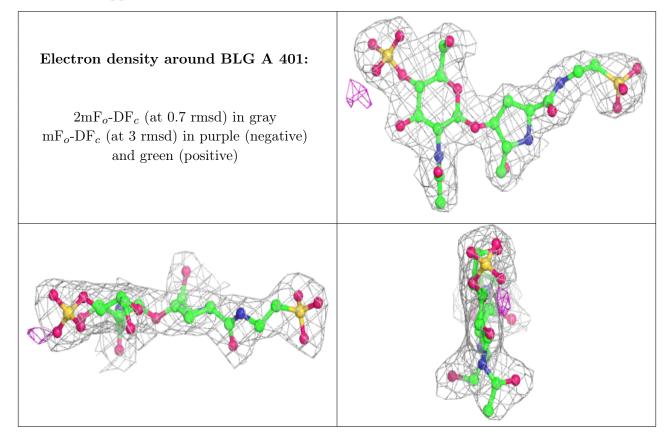


## 6.4 Ligands (i)

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,  $95^{th}$  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	$\mathbf{B} ext{-factors}(\mathbf{A}^2)$	Q<0.9
4	BLG	А	401	35/35	0.95	0.12	$21,\!27,\!57,\!59$	0
3	CA	А	400	1/1	0.99	0.07	22,22,22,22	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.



## 6.5 Other polymers (i)

There are no such residues in this entry.

