

Full wwPDB X-ray Structure Validation Report (i)

Apr 17, 2024 – 04:09 PM JST

PDB ID : 8K2K

Title : Crystal structure of Group 3 Oligosaccharide/Monosaccharide-releasing beta-

N-acetylgalactosaminidase NgaDssm in complex with GalNAc-thiazoline

Authors : Sumida, T.; Fushinobu, S.

Deposited on : 2023-07-12

Resolution : 1.75 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 types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

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

 $Mol Probity \quad : \quad 4.02b\text{--}467$

Mogul: 1.8.5 (274361), CSD as541be (2020)

Xtriage (Phenix) : 1.13 EDS : 2.36

Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)

 $Refmac \quad : \quad 5.8.0158$

CCP4 : 7.0.044 (Gargrove) roteins) : Engh & Huber (2001)

Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

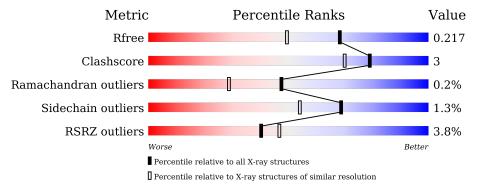
Validation Pipeline (wwPDB-VP) : 2.36

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 1.75 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 $(\# \mathrm{Entries})$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries},{\rm resolution\ range}(\mathring{\rm A})) \end{array}$
R_{free}	130704	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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 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	
			4%	
1	A	542	90%	7% ••



2 Entry composition (i)

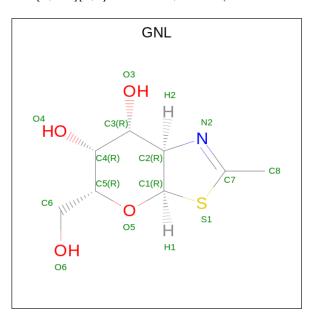
There are 6 unique types of molecules in this entry. The entry contains 4739 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Oligosaccharide/Monosaccharide-releasing beta-N-acetylgalac tosaminidase.

Mo	l Chain	Residues		At	oms			ZeroOcc	AltConf	Trace
1	A	529	Total 4307	C 2751	N 769	O 775	S 12	0	1	0

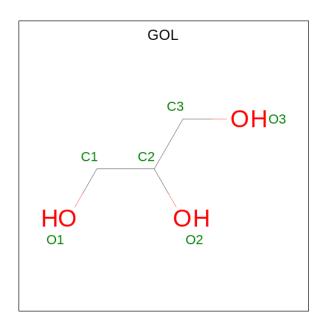
• Molecule 2 is (3aR,5R,6R,7R,7aR)-5-(hydroxymethyl)-2-methyl-5,6,7,7a-tetrahydro-3aH-py rano[3,2-d[1,3]thiazole-6,7-diol (three-letter code: GNL) (formula: $C_8H_{13}NO_4S$).



Mol	Chain	Residues		${f Atoms}$			ZeroOcc	AltConf	
2	Λ	1	Total	С	N	О	S	0	0
2	A	1	14	8	1	4	1	0	0

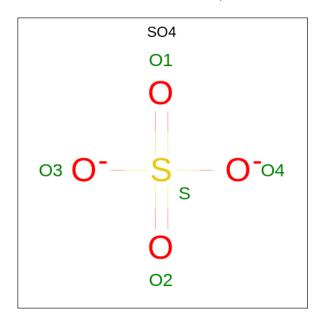
• Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).





Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 6	C 3	O 3	0	0

 \bullet Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: $\mathrm{O_4S}).$



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0

• Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Cl 1 1	0	0

• Molecule 6 is water.

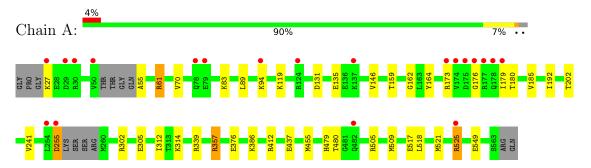
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	401	Total O 401 401	0	0



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: Oligosaccharide/Monosaccharide-releasing beta-N-acetylgalactosaminidase





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants	53.54Å 53.54Å 428.29Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.93 - 1.75	Depositor
rtesolution (A)	47.88 - 1.75	EDS
% Data completeness	100.0 (47.93-1.75)	Depositor
(in resolution range)	100.0 (47.88-1.75)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	2.27 (at 1.75Å)	Xtriage
Refinement program	REFMAC 5.8.0352	Depositor
D.D.	0.176 , 0.206	Depositor
R, R_{free}	0.188 , 0.217	DCC
R_{free} test set	3313 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å ²)	22.9	Xtriage
Anisotropy	0.054	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.36, 43.8	EDS
L-test for twinning ²	$ < L > = 0.46, < L^2> = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4739	wwPDB-VP
Average B, all atoms (Å ²)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 6.49% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical 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.



¹Intensities 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: SO4, GNL, GOL, CL

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	Bo	nd lengths	Bo	ond angles
IVIOI	Chain	RMSZ	# Z > 5	RMSZ	# Z > 5
1	A	0.55	5/4425 (0.1%)	0.84	6/6017 (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 maintain group or atoms of a sidechain that are expected to be planar.

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

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(A)	Ideal(A)
1	A	135	GLU	CD-OE2	7.18	1.33	1.25
1	A	376	GLU	CD-OE2	6.55	1.32	1.25
1	A	549	GLU	CD-OE2	6.37	1.32	1.25
1	A	549	GLU	CD-OE1	5.11	1.31	1.25
1	A	437	GLU	CD-OE2	5.08	1.31	1.25

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$Observed(^o)$	$\operatorname{Ideal}({}^{o})$
1	A	412	ARG	NE-CZ-NH2	-8.68	115.96	120.30
1	A	505	ARG	NE-CZ-NH2	-6.61	117.00	120.30
1	A	61	ARG	NE-CZ-NH2	-5.78	117.41	120.30
1	A	302	ARG	CG-CD-NE	5.32	122.96	111.80
1	A	412	ARG	NE-CZ-NH1	5.19	122.89	120.30
1	A	339	ARG	NE-CZ-NH2	-5.13	117.73	120.30

There are no chirality outliers.



All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	357	ARG	Sidechain
1	A	525	ARG	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	4307	0	4255	24	0
2	A	14	0	13	0	0
3	A	6	0	8	0	0
4	A	10	0	0	0	0
5	A	1	0	0	0	0
6	A	401	0	0	2	0
All	All	4739	0	4276	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 3.

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

Atom-1	Atom-2	$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} \ ({\rm \AA}) \end{array}$	Clash overlap (Å)
1:A:517:GLU:HG3	1:A:521:MET:HE2	1.71	0.72
1:A:517:GLU:HG3	1:A:521:MET:CE	2.23	0.68
1:A:357:ARG:HH21	1:A:386:LYS:H	1.47	0.62
1:A:173:ARG:HH12	1:A:176:GLY:HA2	1.66	0.60
1:A:173:ARG:NH1	1:A:176:GLY:HA2	2.18	0.58
1:A:525:ARG:HH11	1:A:525:ARG:HG3	1.70	0.56
1:A:509:MET:SD	6:A:1000:HOH:O	2.58	0.54
1:A:479:HIS:CG	1:A:480:THR:H	2.27	0.52
1:A:61:ARG:HD2	1:A:159:THR:O	2.10	0.51
1:A:518:LEU:HA	1:A:521:MET:HE3	1.94	0.50
1:A:305:GLU:OE2	1:A:314:LYS:HE2	2.12	0.50
1:A:179:ILE:O	1:A:180:THR:HG22	2.13	0.49
1:A:173:ARG:NH1	1:A:176:GLY:O	2.46	0.48
1:A:89:LEU:HD13	1:A:164:TYR:CZ	2.49	0.47

Continued on next page...



I 'omtamalod	trom	mmonia	maaa
Continued	11 0116	DICUIUUS	Daue
	.,	10	1

Atom-1	Atom-2	$egin{aligned} ext{Interatomic} \ ext{distance} \ (ext{Å}) \end{aligned}$	Clash overlap (Å)
1:A:312:ILE:HD12	1:A:312:ILE:C	2.38	0.44
1:A:255:ASP:C	1:A:255:ASP:OD1	2.57	0.43
1:A:192:ILE:HD12	6:A:936:HOH:O	2.18	0.42
1:A:179:ILE:O	1:A:180:THR:CG2	2.67	0.42
1:A:517:GLU:HG3	1:A:521:MET:HE1	1.99	0.42
1:A:70:VAL:CG1	1:A:146:VAL:HG13	2.49	0.42
1:A:89:LEU:HD11	1:A:162:GLY:HA3	2.02	0.41
1:A:55:ALA:HB3	1:A:185:VAL:HG12	2.02	0.41
1:A:202:THR:HB	1:A:241:VAL:HB	2.02	0.40
1:A:517:GLU:O	1:A:521:MET:HE2	2.20	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	$524/542 \ (97\%)$	511 (98%)	12 (2%)	1 (0%)	47 29	

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	131	ASP

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	Chain Analysed Rotameric Outli		Outliers	Percentiles
1	A	462/471 (98%)	456 (99%)	6 (1%)	69 54

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	27	LYS
1	A	83	LYS
1	A	94	LYS
1	A	119	LYS
1	A	255	ASP
1	A	455	MET

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)

Of 5 ligands modelled in this entry, 1 is monoatomic - leaving 4 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 Ch		Chain	Chain Dag	s Link Bond lengths			Bond angles			
MIOI	Type	Chain	Res	LIIIK	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	GNL	A	601	-	13,15,15	0.78	0	12,22,22	0.91	1 (8%)
3	GOL	A	602	-	5,5,5	0.13	0	5,5,5	0.38	0
4	SO4	A	604	-	4,4,4	0.31	0	6,6,6	0.31	0
4	SO4	A	603	-	4,4,4	0.28	0	6,6,6	0.20	0

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	GNL	A	601	-	-	0/2/30/30	0/2/2/2
3	GOL	A	602	-	-	0/4/4/4	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	\mathbf{Z}	$Observed(^o)$	$\operatorname{Ideal}(^{o})$
2	A	601	GNL	C1-O5-C5	2.37	116.96	112.58

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	Analysed	$\langle { m RSRZ} \rangle$	# RSRZ > 2	$OWAB(A^2)$	Q < 0.9
1	A	529/542 (97%)	-0.03	20 (3%) 40 47	14, 22, 43, 69	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	175	ASP	5.2
1	A	254	LEU	3.9
1	A	174	VAL	3.6
1	A	29	ASP	3.3
1	A	176	GLY	3.3
1	A	94	LYS	3.0
1	A	177	ARG	2.9
1	A	525	ARG	2.7
1	A	30	ARG	2.6
1	A	255	ASP	2.4
1	A	178	GLN	2.3
1	A	27	LYS	2.3
1	A	78	GLN	2.2
1	A	124	ARG	2.2
1	A	482	GLN	2.2
1	A	137	LYS	2.2
1	A	50	VAL	2.2
1	A	79	GLU	2.1
1	A	173	ARG	2.1
1	A	179	ILE	2.1

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^{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	$\mathbf{B} ext{-}\mathbf{factors}(\mathring{\mathbf{A}}^2)$	Q < 0.9
4	SO4	A	603	5/5	0.93	0.18	48,49,51,62	0
3	GOL	A	602	6/6	0.96	0.07	23,25,25,27	0
5	CL	A	605	1/1	0.96	0.11	50,50,50,50	0
4	SO4	A	604	5/5	0.97	0.17	42,50,54,57	0
2	GNL	A	601	14/14	0.97	0.07	16,18,19,19	0

6.5 Other polymers (i)

There are no such residues in this entry.

