W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Jan 8, 2024 – 01:47 am GMT

PDB ID	:	6EZS
Title	:	Crystal structure of GH20 Exo beta-N-Acetylglucosaminidase from Vibrio harveyi in complex with N-acetylglucosamine
Authors	:	Porfetye, A.T.; Meekrathok, P.; Burger, M.; Vetter, I.R.; Suginta, W.
Deposited on	:	2017-11-16
Resolution	:	2.50 Å(reported)

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

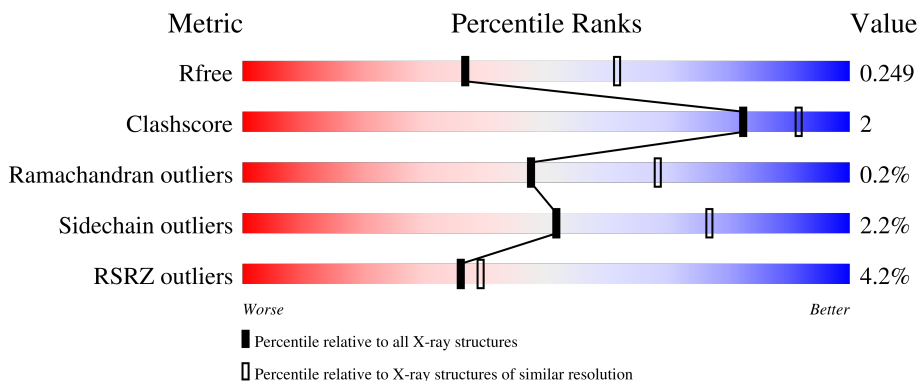|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	Percentile Ranks	Value
Rfree		0.249
Clashscore		2
Ramachandran outliers		0.2%
Sidechain outliers		2.2%
RSRZ outliers		4.2%

Worse — *Better*

▮ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	4% / 89% / 9% ·
1	B	652	4% / 92% / 6% ·

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	MLI	A	702	-	-	X	-

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 11254 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 Beta-N-acetylglucosaminidase Nag2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	639	Total 5161	C 3285	N 880	O 974	S 22	0	3	0
1	B	639	Total 5164	C 3287	N 882	O 973	S 22	0	3	0

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 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total 15	C 8	N 1	O 6	0	0
2	B	1	Total 15	C 8	N 1	O 6	0	0

• Molecule 3 is MALONATE ION (three-letter code: MLI) (formula: $C_3H_2O_4$).
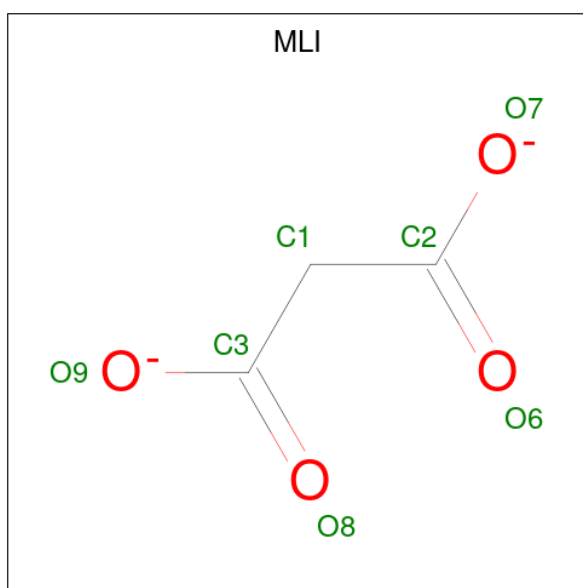
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 7	C 3	O 4	0	0
3	B	1	Total 7	C 3	O 4	0	0

• Molecule 4 is water.

| Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	419	Total O 419 419	0	0
4	B	466	Total O 466 466	0	0

3 Residue-property plots ⓘ

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
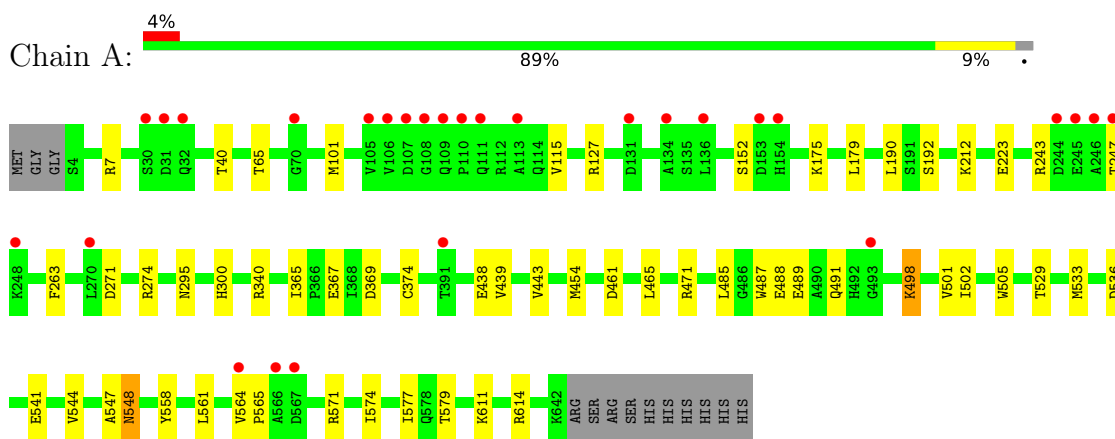
• Molecule 1: Beta-N-acetylglucosaminidase Nag2
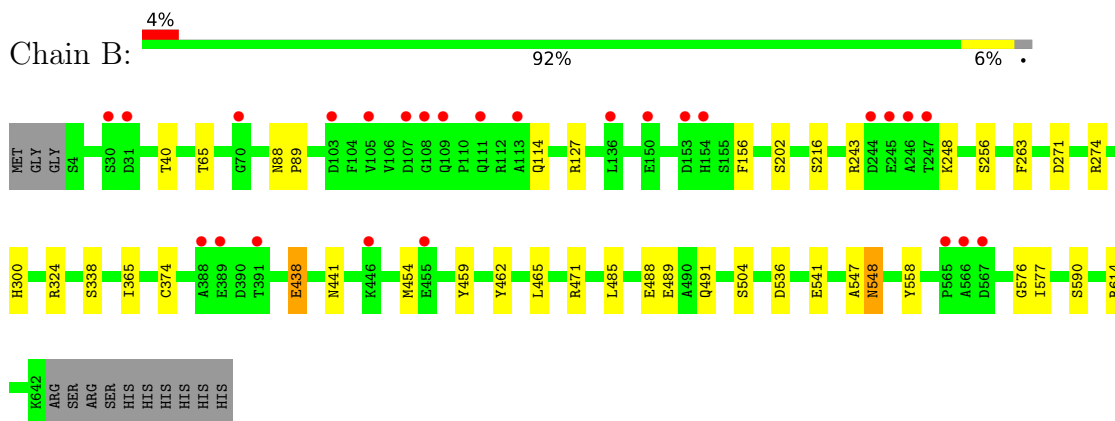
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	91.25Å 129.63Å 100.00Å 90.00° 114.36° 90.00°	Depositor
Resolution (Å)	48.02 – 2.50 48.02 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.6 (48.02-2.50) 99.6 (48.02-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	3.29 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.200 , 0.248 0.205 , 0.249	Depositor DCC
R_{free} test set	3648 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	33.1	Xtriage
Anisotropy	0.210	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 39.8	EDS
L-test for twinning [2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.006 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	11254	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 20.48 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.6025e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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: NAG, MLI

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	0.47	0/5307	0.67	1/7216 (0.0%)
1	B	0.48	0/5307	0.67	1/7215 (0.0%)
All	All	0.47	0/10614	0.67	2/14431 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	340	ARG	NE-CZ-NH2	5.15	122.87	120.30
1	B	324	ARG	NE-CZ-NH1	5.00	122.80	120.30

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	5161	0	5000	30	0
1	B	5164	0	5006	20	0
2	A	15	0	15	0	0
2	B	15	0	15	0	0
3	A	7	0	2	3	0
3	B	7	0	2	0	0
4	A	419	0	0	5	0

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	466	0	0	4	0
All	All	11254	0	10040	50	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.

The worst 5 of 50 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:B:40:THR:HG22	1:B:65:THR:HG22	1.70	0.74
1:A:536:ASP:HB3	1:A:541:GLU:HG3	1.84	0.58
1:A:485:LEU:HD22	1:A:501:VAL:HB	1.86	0.57
1:A:7:ARG:HD2	4:A:1074:HOH:O	2.05	0.56
1:A:243:ARG:NH1	4:A:802:HOH:O	2.30	0.56

There are no symmetry-related clashes.

5.3   Torsion angles ⓘ

5.3.1   Protein backbone ⓘ

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	640/652 (98%)	631 (99%)	8 (1%)	1 (0%)	47	68
1	B	640/652 (98%)	630 (98%)	9 (1%)	1 (0%)	47	68
All	All	1280/1304 (98%)	1261 (98%)	17 (1%)	2 (0%)	47	68

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	548	ASN
1	A	548	ASN

5.3.2　Protein sidechains ⓘ

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	559/567 (99%)	543 (97%)	16 (3%)	42	69
1	B	559/567 (99%)	548 (98%)	11 (2%)	55	79
All	All	1118/1134 (99%)	1091 (98%)	27 (2%)	52	74

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	564	VAL
1	B	243	ARG
1	B	465	LEU
1	B	202	SER
1	B	248	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	114	GLN
1	B	441	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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

4 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	NAG	A	701	-	15,15,15	0.80	1 (6%)	21,21,21	2.32	4 (19%)
3	MLI	B	702	-	6,6,6	1.08	0	7,7,7	0.73	0
2	NAG	B	701	-	15,15,15	0.61	0	21,21,21	1.25	4 (19%)
3	MLI	A	702	-	6,6,6	1.16	0	7,7,7	0.91	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	NAG	A	701	-	-	0/6/26/26	0/1/1/1
3	MLI	B	702	-	-	0/4/4/4	-
2	NAG	B	701	-	-	0/6/26/26	0/1/1/1
3	MLI	A	702	-	-	4/4/4/4	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	701	NAG	C1-C2	2.07	1.55	1.52

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	701	NAG	O5-C1-C2	-8.39	101.08	109.52
2	A	701	NAG	C1-C2-N2	-2.80	107.49	110.73
2	A	701	NAG	C1-C2-C3	2.77	114.32	110.54
2	B	701	NAG	C1-C2-N2	-2.52	107.81	110.73
2	B	701	NAG	C8-C7-N2	-2.27	112.26	116.10

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	702	MLI	C3-C1-C2-O7
3	A	702	MLI	C3-C1-C2-O6
3	A	702	MLI	C2-C1-C3-O8
3	A	702	MLI	C2-C1-C3-O9

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	702	MLI	3	0

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 ⓘ

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, 95th 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(Å²)	Q<0.9
1	A	639/652 (98%)	0.18	28 (4%)	34	37	20, 34, 58, 97	0
1	B	639/652 (98%)	0.16	26 (4%)	37	40	20, 33, 58, 97	0
All	All	1278/1304 (98%)	0.17	54 (4%)	36	39	20, 34, 58, 97	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	246	ALA	9.2
1	A	107	ASP	6.0
1	B	245	GLU	5.6
1	A	108	GLY	5.3
1	B	246	ALA	5.2

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, 95th 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
3	MLI	A	702	7/7	0.68	0.36	34,36,37,38	7
3	MLI	B	702	7/7	0.94	0.15	42,43,44,44	0
2	NAG	B	701	15/15	0.96	0.21	23,25,27,28	0
2	NAG	A	701	15/15	0.97	0.23	26,27,30,30	0

6.5 Other polymers (i)

There are no such residues in this entry.