Full wwPDB X-ray Structure Validation Report ⓘ

May 21, 2020 – 11:38 pm BST

PDB ID	:	6EZR
Title	:	Crystal structure of GH20 Exo beta-N-Acetylglucosaminidase from Vibrio harveyi
Authors	:	Porfetye, A.T.; Meekrathok, P.; Burger, M.; Vetter, I.R.; Suginta, W.
Deposited on	:	2017-11-16
Resolution	:	2.37 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:

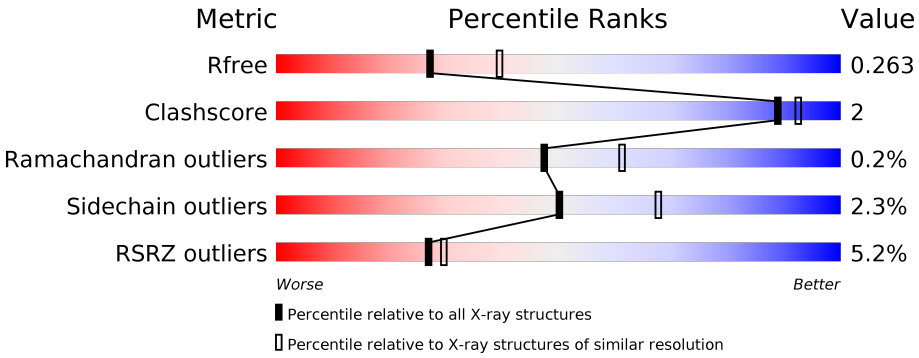
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.37 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	5509 (2.40-2.36)
Clashscore	141614	6082 (2.40-2.36)
Ramachandran outliers	138981	5973 (2.40-2.36)
Sidechain outliers	138945	5975 (2.40-2.36)
RSRZ outliers	127900	5397 (2.40-2.36)

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 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% 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	7% 93%
1	B	652	4% 92% 6%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11303 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	C	N	O	S	0	7	0
			5197	3306	893	976	22			
1	B	639	Total	C	N	O	S	0	7	0
			5183	3298	885	978	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

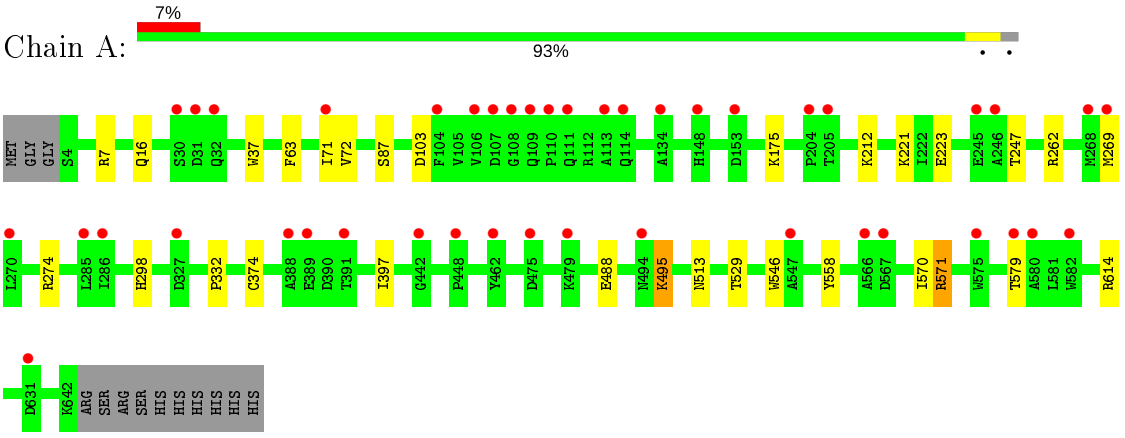
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	435	Total O 435 435	0	0
2	B	488	Total O 488 488	0	0

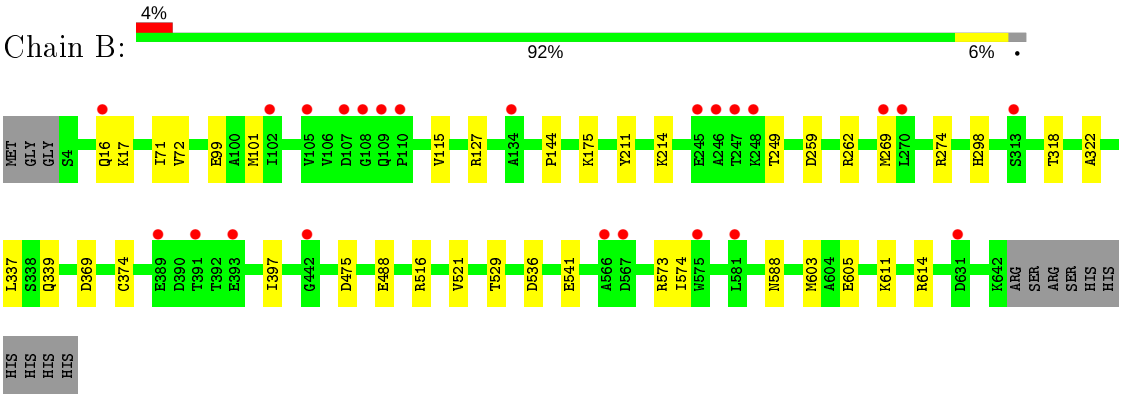
3 Residue-property plots (i)

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 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 (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	90.15Å 130.71Å 98.48Å 90.00° 112.99° 90.00°	Depositor
Resolution (Å)	90.66 – 2.37 48.55 – 2.37	Depositor EDS
% Data completeness (in resolution range)	99.6 (90.66-2.37) 99.6 (48.55-2.37)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.26 (at 2.37Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.214 , 0.259 0.222 , 0.263	Depositor DCC
R_{free} test set	4215 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	39.2	Xtriage
Anisotropy	0.437	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 37.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11303	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.83% 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

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	0.40	0/5350	0.65	0/7272
1	B	0.41	0/5341	0.66	0/7262
All	All	0.40	0/10691	0.66	0/14534

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	5197	0	5042	20	0
1	B	5183	0	5027	20	0
2	A	435	0	0	1	0
2	B	488	0	0	0	0
All	All	11303	0	10069	38	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.

All (38) 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:516[B]:ARG:CG	1:B:516[B]:ARG:HH21	1.84	0.89

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:516[B]:ARG:HH21	1:B:516[B]:ARG:HG3	1.39	0.86
1:A:7[B]:ARG:HH11	1:A:7[B]:ARG:HG2	1.66	0.60
1:B:516[B]:ARG:NH2	1:B:516[B]:ARG:HG3	2.09	0.60
1:A:7[B]:ARG:HG2	1:A:7[B]:ARG:NH1	2.21	0.56
1:B:605:GLU:HG2	1:B:614:ARG:HD3	1.91	0.52
1:A:175:LYS:NZ	2:A:702:HOH:O	2.42	0.52
1:A:571[B]:ARG:CG	1:A:571[B]:ARG:HH11	2.22	0.52
1:A:212:LYS:HB2	1:A:223:GLU:HB2	1.93	0.51
1:A:571[B]:ARG:HB3	1:A:571[B]:ARG:HH11	1.75	0.51
1:B:516[B]:ARG:CG	1:B:516[B]:ARG:NH2	2.55	0.50
1:B:529:THR:HG22	1:B:603:MET:CE	2.42	0.50
1:B:603:MET:HA	1:B:603:MET:HE3	1.95	0.49
1:B:529:THR:HG22	1:B:603:MET:HE1	1.93	0.49
1:B:516[B]:ARG:HH21	1:B:516[B]:ARG:HG2	1.72	0.49
1:A:16[B]:GLN:HG3	1:B:397:ILE:HD13	1.96	0.48
1:A:63:PHE:CD1	1:A:332:PRO:HG2	2.49	0.46
1:B:101:MET:HA	1:B:115:VAL:HG23	1.97	0.46
1:B:269:MET:HG3	1:B:298:HIS:CD2	2.51	0.45
1:A:397:ILE:HD13	1:B:16[B]:GLN:HG3	1.99	0.45
1:B:269:MET:HG3	1:B:298:HIS:CG	2.52	0.44
1:B:211:TYR:CE2	1:B:259:ASP:HB3	2.54	0.43
1:A:37:TRP:HA	1:A:103:ASP:O	2.20	0.42
1:B:536:ASP:HB3	1:B:541:GLU:HG3	2.01	0.42
1:B:144:PRO:HB3	1:B:605:GLU:HA	2.01	0.42
1:A:571[B]:ARG:CG	1:A:571[B]:ARG:NH1	2.82	0.41
1:A:571[B]:ARG:HG2	1:A:571[B]:ARG:NH1	2.35	0.41
1:A:571[B]:ARG:HH11	1:A:571[B]:ARG:CB	2.32	0.41
1:B:71:ILE:HG23	1:B:72:VAL:HG23	2.02	0.40
1:A:558:TYR:O	1:A:614:ARG:NH2	2.52	0.40
1:A:7[B]:ARG:CG	1:A:7[B]:ARG:HH11	2.29	0.40
1:B:318:THR:O	1:B:322:ALA:HB3	2.21	0.40
1:A:71:ILE:HG23	1:A:72:VAL:HG23	2.04	0.40
1:B:521:VAL:CG2	1:B:574:ILE:HD13	2.51	0.40
1:A:269:MET:HG3	1:A:298:HIS:CG	2.57	0.40
1:A:529:THR:HB	1:A:579:THR:HA	2.04	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	644/652 (99%)	628 (98%)	14 (2%)	2 (0%)	41	53
1	B	644/652 (99%)	626 (97%)	17 (3%)	1 (0%)	47	61
All	All	1288/1304 (99%)	1254 (97%)	31 (2%)	3 (0%)	47	61

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	488	GLU
1	B	488	GLU
1	A	495	LYS

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	563/567 (99%)	552 (98%)	11 (2%)	55	72
1	B	563/567 (99%)	547 (97%)	16 (3%)	43	61
All	All	1126/1134 (99%)	1099 (98%)	27 (2%)	50	66

All (27) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	87	SER
1	A	221	LYS
1	A	247	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	262	ARG
1	A	274	ARG
1	A	374	CYS
1	A	495	LYS
1	A	513	ASN
1	A	570	ILE
1	A	571[A]	ARG
1	A	571[B]	ARG
1	B	17	LYS
1	B	99	GLU
1	B	127	ARG
1	B	175	LYS
1	B	214	LYS
1	B	249	THR
1	B	262	ARG
1	B	274	ARG
1	B	337	LEU
1	B	339	GLN
1	B	369	ASP
1	B	374	CYS
1	B	475	ASP
1	B	573	ARG
1	B	588	ASN
1	B	611	LYS

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

Mol	Chain	Res	Type
1	A	588	ASN
1	B	588	ASN

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

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

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, 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	639/652 (98%)	0.47	43 (6%)	17	19	33, 48, 73, 109	0
1	B	639/652 (98%)	0.38	24 (3%)	40	43	33, 48, 73, 104	0
All	All	1278/1304 (98%)	0.43	67 (5%)	27	30	33, 48, 73, 109	0

All (67) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	246	ALA	5.5
1	A	108	GLY	5.3
1	B	109	GLN	5.0
1	A	246	ALA	5.0
1	A	245	GLU	4.8
1	B	567	ASP	4.6
1	A	107	ASP	4.5
1	A	567	ASP	4.3
1	A	109	GLN	4.0
1	A	106	VAL	4.0
1	B	107	ASP	3.8
1	B	391	THR	3.6
1	B	108	GLY	3.6
1	A	566	ALA	3.5
1	A	113	ALA	3.4
1	B	245	GLU	3.4
1	A	270	LEU	3.4
1	A	31	ASP	3.4
1	B	270	LEU	3.3
1	A	475	ASP	3.3
1	A	148	HIS	3.2
1	B	442	GLY	3.2
1	A	153	ASP	3.2
1	A	30	SER	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	110	PRO	3.1
1	B	389	GLU	3.0
1	A	269	MET	3.0
1	A	204	PRO	2.9
1	A	442	GLY	2.9
1	A	389	GLU	2.8
1	A	580	ALA	2.8
1	B	566	ALA	2.8
1	A	110	PRO	2.7
1	A	111	GLN	2.6
1	B	102	ILE	2.6
1	B	105	VAL	2.5
1	B	248	LYS	2.5
1	B	631	ASP	2.5
1	A	114	GLN	2.4
1	A	579	THR	2.4
1	A	388	ALA	2.4
1	A	285	LEU	2.3
1	A	391	THR	2.3
1	A	494	ASN	2.3
1	A	205	THR	2.3
1	B	247	THR	2.3
1	B	313	SER	2.3
1	B	16[A]	GLN	2.3
1	A	104	PHE	2.2
1	A	575	TRP	2.2
1	B	134	ALA	2.2
1	B	269	MET	2.2
1	A	286	ILE	2.2
1	A	32	GLN	2.2
1	A	479	LYS	2.2
1	A	582	TRP	2.2
1	A	327	ASP	2.2
1	A	448	PRO	2.1
1	A	268	MET	2.1
1	A	134	ALA	2.1
1	A	547	ALA	2.1
1	A	631	ASP	2.1
1	A	71	ILE	2.1
1	B	575	TRP	2.1
1	B	581	LEU	2.1
1	A	462	TYR	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	393	GLU	2.0

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 carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.