WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2024 – 09:06 AM EST

PDB ID	:	2ZXQ
Title	:	Crystal structure of endo-alpha-N-acetylgalactosaminidase from Bifidobacterium longum (EngBF)
Authors	:	Suzuki, R.; Katayama, T.; Ashida, H.; Yamamoto, K.; Kitaoka, M.; Fushinobu, S.
Deposited on	:	2009-01-05
Resolution	:	2.00 Å(reported)

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

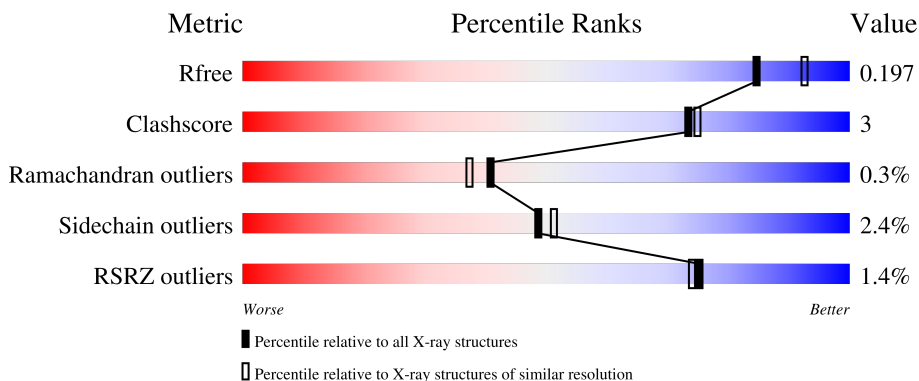|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 2.00 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		0.197
Clashscore		3
Ramachandran outliers		0.3%
Sidechain outliers		2.4%
RSRZ outliers		1.4%

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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	1376	80% 5% • 14%

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:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	MPD	A	1695	X	-	-	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	MPD	A	1697	X	-	X	-

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 10501 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 Endo-alpha-N-acetylgalactosaminidase.

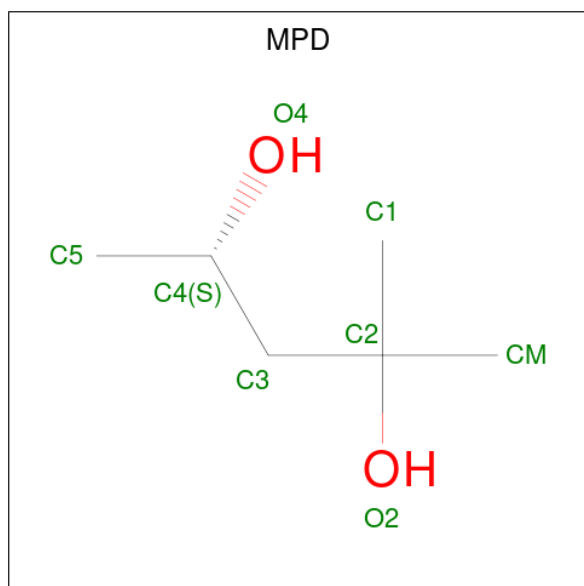| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1178	Total 9070	C 5643	N 1553	O 1850	S 24	0	0	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	319	MET	-	expression tag	UNP Q3T552
A	320	GLY	-	expression tag	UNP Q3T552
A	321	SER	-	expression tag	UNP Q3T552
A	322	SER	-	expression tag	UNP Q3T552
A	323	HIS	-	expression tag	UNP Q3T552
A	324	HIS	-	expression tag	UNP Q3T552
A	325	HIS	-	expression tag	UNP Q3T552
A	326	HIS	-	expression tag	UNP Q3T552
A	327	HIS	-	expression tag	UNP Q3T552
A	328	HIS	-	expression tag	UNP Q3T552
A	329	SER	-	expression tag	UNP Q3T552
A	330	SER	-	expression tag	UNP Q3T552
A	331	GLY	-	expression tag	UNP Q3T552
A	332	LEU	-	expression tag	UNP Q3T552
A	333	VAL	-	expression tag	UNP Q3T552
A	334	PRO	-	expression tag	UNP Q3T552
A	335	ARG	-	expression tag	UNP Q3T552
A	336	GLY	-	expression tag	UNP Q3T552
A	337	SER	-	expression tag	UNP Q3T552
A	338	HIS	-	expression tag	UNP Q3T552
A	339	MET	-	expression tag	UNP Q3T552
A	1123	SER	GLY	SEE REMARK 999	UNP Q3T552

- Molecule 2 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	4	Total Mn 4 4	0	0

- Molecule 3 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: $C_6H_{14}O_2$).

MPD

O4
OH

C1

C5

C4(S)

C2 CM

C3

OH

O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 8	C 6	O 2	0	0
3	A	1	Total 8	C 6	O 2	0	0
3	A	1	Total 8	C 6	O 2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1403	Total O 1403 1403	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: Endo-alpha-N-acetylgalactosaminidase

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α, β, γ	192.27Å 192.27Å 123.02Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	36.49 – 2.00 36.48 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.7 (36.49-2.00) 98.7 (36.48-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	3.16 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.172 , 0.196 0.173 , 0.197	Depositor DCC
R_{free} test set	8598 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	26.0	Xtriage
Anisotropy	0.020	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 53.2	EDS
L-test for twinning[2]	$< |L| > = 0.51, < L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.014 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	10501	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.69% 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: MPD, MN

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	0.69	1/9250 (0.0%)	0.69	6/12541 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	390	GLU	CB-CG	5.85	1.63	1.52

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	694	ARG	NE-CZ-NH2	-14.71	112.95	120.30
1	A	694	ARG	NE-CZ-NH1	11.77	126.18	120.30
1	A	694	ARG	CG-CD-NE	-6.84	97.44	111.80
1	A	1211	ASN	C-N-CA	-6.28	109.11	122.30
1	A	1253	VAL	CB-CA-C	-6.04	99.92	111.40

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	9070	0	8655	53	0
2	A	4	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	24	0	42	17	0
4	A	1403	0	0	12	0
All	All	10501	0	8697	58	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.

The worst 5 of 58 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:878:ASN:HD22	3:A:1697:MPD:HM2	0.98	1.08
1:A:878:ASN:HD22	3:A:1697:MPD:CM	1.78	0.95
1:A:878:ASN:ND2	3:A:1697:MPD:HM2	1.81	0.95
1:A:971:THR:HG21	3:A:1697:MPD:HM1	1.49	0.94
3:A:1697:MPD:H11	4:A:2431:HOH:O	1.69	0.92

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	1174/1376 (85%)	1130 (96%)	41 (4%)	3 (0%)	41	37

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	790	VAL
1	A	1474	ARG
1	A	1473	LYS

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	967/1124 (86%)	944 (98%)	23 (2%)	49	51

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1253	VAL
1	A	1322	ASP
1	A	1280	GLN
1	A	1382	SER
1	A	580	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	978	GLN
1	A	1211	ASN
1	A	1375	GLN
1	A	1135	ASN
1	A	861	HIS

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 (i)

Of 7 ligands modelled in this entry, 4 are monoatomic - leaving 3 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	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
3	MPD	A	1696	-	7,7,7	0.31	0	9,10,10	0.67	0
3	MPD	A	1697	-	7,7,7	0.74	0	9,10,10	1.32	1 (11%)
3	MPD	A	1695	-	7,7,7	0.93	0	9,10,10	1.88	3 (33%)

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
3	MPD	A	1696	-	-	2/5/5/5	-
3	MPD	A	1697	-	1/1/2/2	0/5/5/5	-
3	MPD	A	1695	-	1/1/2/2	3/5/5/5	-

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1695	MPD	C1-C2-C3	3.44	125.99	109.96
3	A	1697	MPD	O2-C2-CM	-3.27	97.57	108.08
3	A	1695	MPD	CM-C2-C1	-2.99	104.35	110.57
3	A	1695	MPD	O2-C2-C3	-2.52	100.34	109.80

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	1695	MPD	C4
3	A	1697	MPD	C4

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1695	MPD	C1-C2-C3-C4
3	A	1695	MPD	O2-C2-C3-C4
3	A	1696	MPD	O2-C2-C3-C4
3	A	1695	MPD	CM-C2-C3-C4
3	A	1696	MPD	CM-C2-C3-C4

There are no ring outliers.

2 monomers are involved in 17 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1697	MPD	12	0
3	A	1695	MPD	5	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(Å^2)	Q<0.9
1	A	1178/1376 (85%)	-0.55	17 (1%)	75	74	17, 27, 47, 78	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1472	SER	5.5
1	A	1523	ALA	3.6
1	A	1475	ALA	3.3
1	A	1240	THR	3.3
1	A	1092	LYS	3.2

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	MPD	A	1697	8/8	0.79	0.34	49,52,60,60	0
3	MPD	A	1696	8/8	0.86	0.21	68,69,70,70	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
3	MPD	A	1695	8/8	0.87	0.21	23,29,48,50	0
2	MN	A	4	1/1	0.99	0.03	37,37,37,37	0
2	MN	A	2	1/1	0.99	0.04	42,42,42,42	0
2	MN	A	1	1/1	1.00	0.02	30,30,30,30	0
2	MN	A	3	1/1	1.00	0.04	33,33,33,33	0

6.5 Other polymers ⓘ

There are no such residues in this entry.