Full wwPDB X-ray Structure Validation Report ⓘ

Feb 26, 2024 – 06:08 pm GMT

PDB ID	:	8QYE
Title	:	Catalytic core of endo-alpha-N-acetylgalactosaminidase from Bifidobacterium longum (EngBF) concieved by deep network hallucination: dEngBF4
Authors	:	Aghajari, N.
Deposited on	:	2023-10-25
Resolution	:	2.05 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

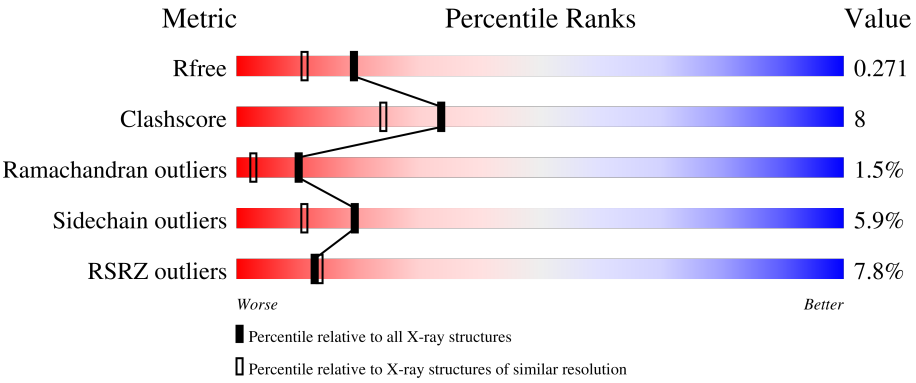
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.05 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	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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	MOLA	299	6% 82% 7% 11%
1	MOLB	299	6% 72% · 26%

2 Entry composition (i)

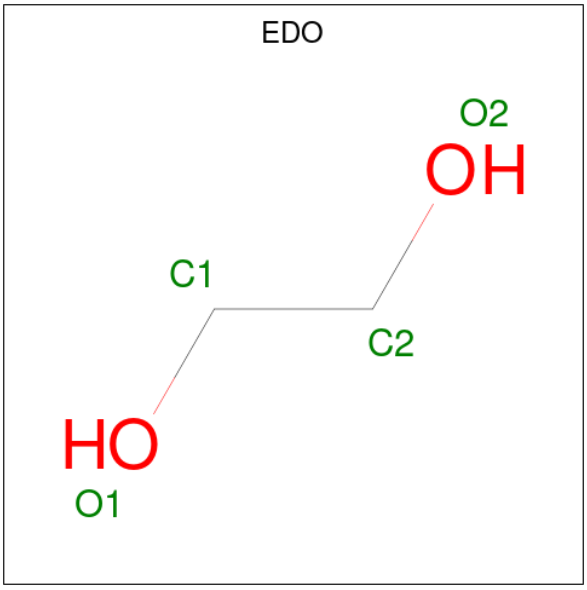
There are 3 unique types of molecules in this entry. The entry contains 3926 atoms, of which 0 are hydrogens and 0 are deuteriums.

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	MOLA	266	Total	C	N	O	S	0	0	0
			2028	1305	327	393	3			
1	MOLB	222	Total	C	N	O	S	0	0	0
			1703	1094	273	334	2			

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	MOLA	1	Total	C	O	0	0
			4	2	2		
2	MOLA	1	Total	C	O	0	0
			4	2	2		
2	MOLA	1	Total	C	O	0	0
			4	2	2		
2	MOLA	1	Total	C	O	0	0
			4	2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	MOLA	1	Total C O 4 2 2	0	0
2	MOLA	1	Total C O 4 2 2	0	0
2	MOLA	1	Total C O 4 2 2	0	0
2	MOLA	1	Total C O 4 2 2	0	0
2	MOLB	1	Total C O 4 2 2	0	0
2	MOLB	1	Total C O 4 2 2	0	0

- Molecule 3 is water.

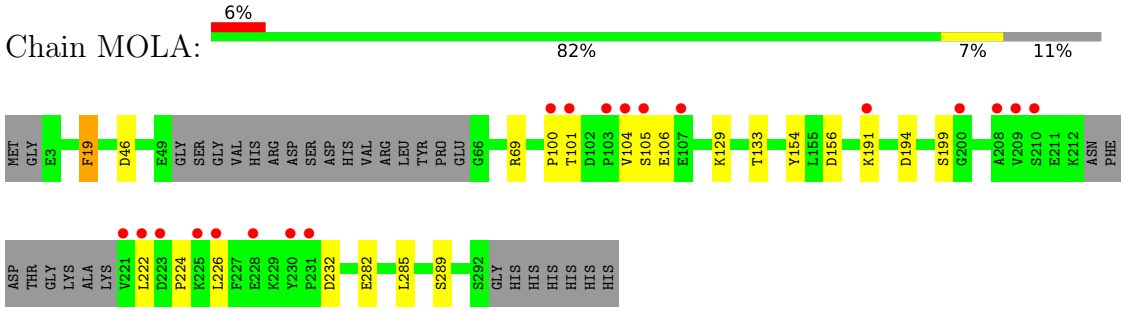
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	MOLA	89	Total O 89 89	0	0
3	MOLB	66	Total O 66 66	0	0

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.

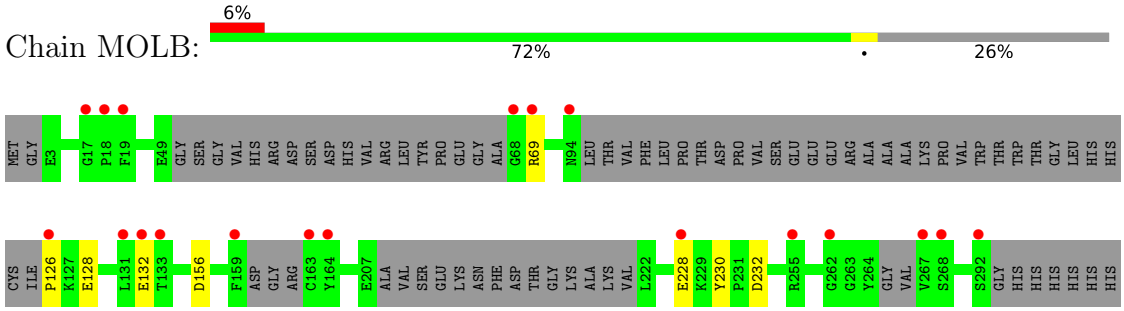
• Molecule 1: ENDO-ALPHA-N-ACETYLGALACTOSAMINIDASE

Chain MOLA:



• Molecule 1: ENDO-ALPHA-N-ACETYLGALACTOSAMINIDASE

Chain MOLB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	35.87Å 70.49Å 197.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	57.35 – 2.05 57.35 – 2.05	Depositor EDS
% Data completeness (in resolution range)	99.3 (57.35-2.05) 99.3 (57.35-2.05)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.31 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.8.0267, PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.229 , 0.273 0.232 , 0.271	Depositor DCC
R_{free} test set	1613 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	46.6	Xtriage
Anisotropy	0.255	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 59.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3926	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.78% 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: EDO

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	MOLA	0.26	0/2064	0.47	0/2809
1	MOLB	0.24	0/1724	0.46	1/2333 (0.0%)
All	All	0.25	0/3788	0.46	1/5142 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	MOLB	126	PRO	N-CA-CB	6.00	110.50	103.30

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	MOLA	2028	0	0	0	0
1	MOLB	1703	0	0	0	0
2	MOLA	32	0	0	0	0
2	MOLB	8	0	0	0	0
3	MOLA	89	0	0	0	0
3	MOLB	66	0	0	0	0
All	All	3926	0	0	0	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 8.

There are no clashes within the asymmetric unit.

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	MOLA	260/299 (87%)	244 (94%)	11 (4%)	5 (2%)	8 2
1	MOLB	210/299 (70%)	203 (97%)	5 (2%)	2 (1%)	15 6
All	All	470/598 (79%)	447 (95%)	16 (3%)	7 (2%)	10 3

All (7) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	MOLA	101	THR
1	MOLA	224	PRO
1	MOLA	106	GLU
1	MOLB	128	GLU
1	MOLB	156	ASP
1	MOLA	19	PHE
1	MOLA	100	PRO

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	MOLA	212/244 (87%)	194 (92%)	18 (8%)	10 4
1	MOLB	179/244 (73%)	174 (97%)	5 (3%)	43 37
All	All	391/488 (80%)	368 (94%)	23 (6%)	19 11

All (23) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	MOLA	19	PHE
1	MOLA	46	ASP
1	MOLA	69	ARG
1	MOLA	104	VAL
1	MOLA	105	SER
1	MOLA	129	LYS
1	MOLA	133	THR
1	MOLA	154	TYR
1	MOLA	156	ASP
1	MOLA	191	LYS
1	MOLA	194	ASP
1	MOLA	199	SER
1	MOLA	222	LEU
1	MOLA	226	LEU
1	MOLA	232	ASP
1	MOLA	282	GLU
1	MOLA	285	LEU
1	MOLA	289	SER
1	MOLB	69	ARG
1	MOLB	132	GLU
1	MOLB	228	GLU
1	MOLB	230	TYR
1	MOLB	232	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

10 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	EDO	MOLB	302	-	3,3,3	0.45	0	2,2,2	0.34	0
2	EDO	MOLA	301	-	3,3,3	0.45	0	2,2,2	0.33	0
2	EDO	MOLA	304	-	3,3,3	0.45	0	2,2,2	0.34	0
2	EDO	MOLA	306	-	3,3,3	0.45	0	2,2,2	0.26	0
2	EDO	MOLB	301	-	3,3,3	0.47	0	2,2,2	0.31	0
2	EDO	MOLA	307	-	3,3,3	0.46	0	2,2,2	0.34	0
2	EDO	MOLA	305	-	3,3,3	0.45	0	2,2,2	0.34	0
2	EDO	MOLA	308	-	3,3,3	0.47	0	2,2,2	0.28	0
2	EDO	MOLA	302	-	3,3,3	0.45	0	2,2,2	0.34	0
2	EDO	MOLA	303	-	3,3,3	0.45	0	2,2,2	0.31	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	EDO	MOLB	302	-	-	0/1/1/1	-
2	EDO	MOLA	301	-	-	0/1/1/1	-
2	EDO	MOLA	304	-	-	0/1/1/1	-
2	EDO	MOLA	306	-	-	0/1/1/1	-
2	EDO	MOLB	301	-	-	0/1/1/1	-
2	EDO	MOLA	307	-	-	0/1/1/1	-
2	EDO	MOLA	305	-	-	0/1/1/1	-
2	EDO	MOLA	308	-	-	0/1/1/1	-
2	EDO	MOLA	302	-	-	0/1/1/1	-
2	EDO	MOLA	303	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

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

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	MOLA	266/299 (88%)	0.60	19 (7%)	16	17	27, 58, 96, 112	0
1	MOLB	222/299 (74%)	0.49	19 (8%)	10	11	31, 58, 94, 111	0
All	All	488/598 (81%)	0.55	38 (7%)	13	14	27, 58, 96, 112	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	MOLB	164	TYR	7.8
1	MOLB	19	PHE	6.8
1	MOLA	223	ASP	6.4
1	MOLB	163	CYS	6.0
1	MOLA	105	SER	5.8
1	MOLA	101	THR	5.7
1	MOLB	68	GLY	5.6
1	MOLB	18	PRO	5.3
1	MOLA	104	VAL	4.8
1	MOLA	103	PRO	4.3
1	MOLA	210	SER	3.8
1	MOLA	209	VAL	3.7
1	MOLB	159	PHE	3.6
1	MOLA	228	GLU	3.6
1	MOLA	222	LEU	3.4
1	MOLA	230	TYR	3.3
1	MOLB	132	GLU	3.3
1	MOLB	255	ARG	3.1
1	MOLA	225	LYS	3.1
1	MOLB	228	GLU	3.0
1	MOLA	231	PRO	3.0
1	MOLB	17	GLY	3.0
1	MOLB	267	VAL	2.9
1	MOLB	133	THR	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	MOLA	100	PRO	2.7
1	MOLB	262	GLY	2.6
1	MOLB	94	ASN	2.5
1	MOLA	191	LYS	2.5
1	MOLA	221	VAL	2.5
1	MOLA	208	ALA	2.4
1	MOLA	107	GLU	2.3
1	MOLB	69	ARG	2.3
1	MOLA	226	LEU	2.2
1	MOLA	200	GLY	2.2
1	MOLB	292	SER	2.1
1	MOLB	131	LEU	2.1
1	MOLB	268	SER	2.0
1	MOLB	126	PRO	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)

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	B-factors(Å^2)	Q<0.9
2	EDO	MOLA	306	4/4	0.68	0.27	46,47,60,63	0
2	EDO	MOLA	302	4/4	0.71	0.26	40,41,49,55	4
2	EDO	MOLA	305	4/4	0.74	0.38	39,48,50,52	0
2	EDO	MOLA	307	4/4	0.75	0.19	65,72,72,76	0
2	EDO	MOLA	301	4/4	0.77	0.35	37,41,42,51	4
2	EDO	MOLA	303	4/4	0.78	0.21	59,64,70,72	0
2	EDO	MOLB	302	4/4	0.78	0.36	40,49,54,65	4
2	EDO	MOLB	301	4/4	0.88	0.25	38,47,54,60	0
2	EDO	MOLA	308	4/4	0.90	0.15	35,42,48,51	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	EDO	MOLA	304	4/4	0.95	0.12	44,47,48,57	0

6.5 Other polymers (i)

There are no such residues in this entry.