



# Full wwPDB X-ray Structure Validation Report ⓘ

Dec 10, 2023 – 12:56 AM JST

PDB ID : 8IBR  
Title : Crystal structure of GH42 beta-galactosidase BiBga42A from Bifidobacterium longum subspecies infantis in complex with glycerol  
Authors : Hidaka, M.; Fushinobu, S.; Gotoh, A.; Katayama, T.  
Deposited on : 2023-02-10  
Resolution : 1.70 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](mailto: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 : 1.8.5 (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

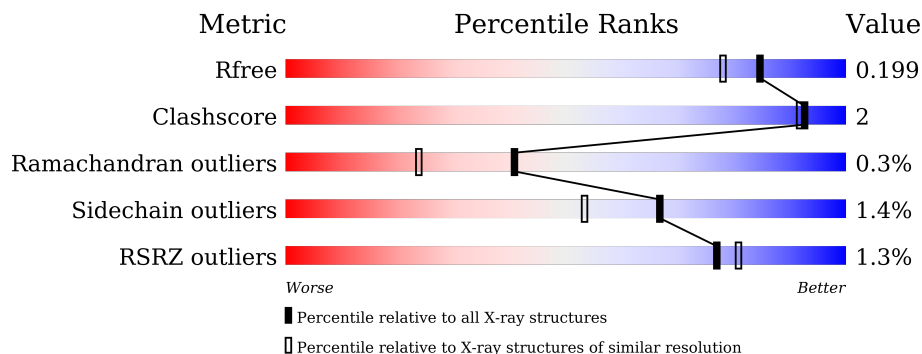
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.70 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	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	702	

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	PEG	A	802	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6160 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-galactosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	691	5477	3482	934	1039	22	0	1	0

There are 11 discrepancies between the modelled and reference sequences:

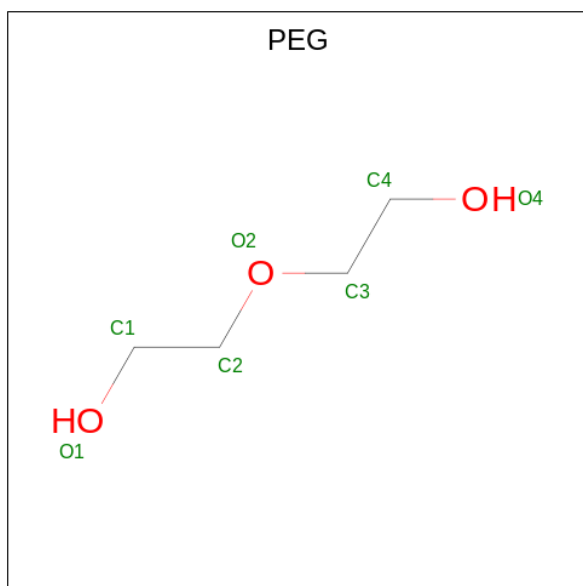
Chain	Residue	Modelled	Actual	Comment	Reference
A	692	ALA	-	expression tag	UNP B7GUD7
A	693	ALA	-	expression tag	UNP B7GUD7
A	694	ALA	-	expression tag	UNP B7GUD7
A	695	LEU	-	expression tag	UNP B7GUD7
A	696	GLU	-	expression tag	UNP B7GUD7
A	697	HIS	-	expression tag	UNP B7GUD7
A	698	HIS	-	expression tag	UNP B7GUD7
A	699	HIS	-	expression tag	UNP B7GUD7
A	700	HIS	-	expression tag	UNP B7GUD7
A	701	HIS	-	expression tag	UNP B7GUD7
A	702	HIS	-	expression tag	UNP B7GUD7

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



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

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula:  $C_4H_{10}O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			7	4	3		

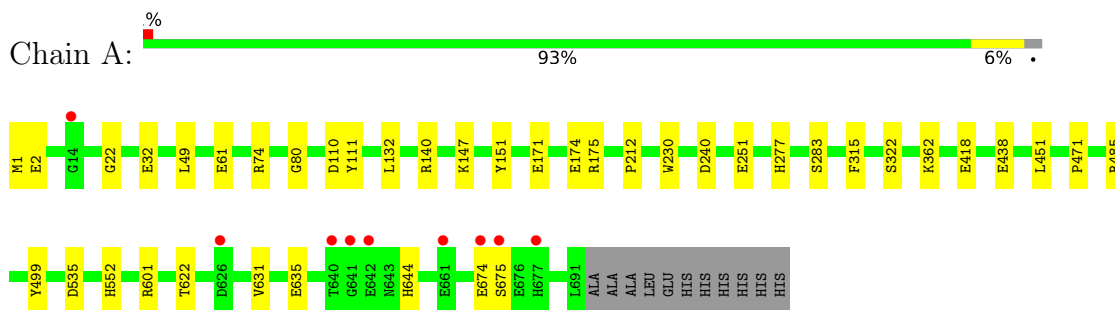
- Molecule 4 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	A	670	Total 670	O 670	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: Beta-galactosidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 3 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	145.64Å 145.64Å 66.84Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	45.91 – 1.70 45.87 – 1.70	Depositor EDS
% Data completeness (in resolution range)	95.5 (45.91-1.70) 95.4 (45.87-1.70)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.43 (at 1.70Å)	Xtrriage
Refinement program	REFMAC 5.8.0403	Depositor
R, $R_{free}$	0.161 , 0.186 0.175 , 0.199	Depositor DCC
$R_{free}$ test set	4302 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	12.4	Xtrriage
Anisotropy	0.273	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 47.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.023 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	6160	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.36% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 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: GOL, PEG

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	0.58	5/5639 (0.1%)	0.80	1/7691 (0.0%)

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 mainchain group or atoms of a sidechain that are expected to be planar.

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

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	251	GLU	CD-OE2	6.16	1.32	1.25
1	A	174	GLU	CD-OE1	-6.11	1.19	1.25
1	A	61	GLU	CD-OE2	5.24	1.31	1.25
1	A	418	GLU	CD-OE2	-5.07	1.20	1.25
1	A	438	GLU	CD-OE1	5.03	1.31	1.25

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	535	ASP	CB-CA-C	-5.46	99.48	110.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	485	ARG	Sidechain



## 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	5477	0	5175	15	0
2	A	6	0	8	0	0
3	A	7	0	10	4	0
4	A	670	0	0	2	0
All	All	6160	0	5193	16	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 (16) 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:132:LEU:HD22	3:A:802:PEG:H21	1.33	1.10
1:A:240:ASP:HB3	3:A:802:PEG:H42	1.59	0.84
1:A:132:LEU:HD22	3:A:802:PEG:C2	2.17	0.66
3:A:802:PEG:H41	4:A:1230:HOH:O	1.98	0.62
1:A:1:MET:HG2	1:A:277:HIS:CD2	2.37	0.60
1:A:32:GLU:OE1	1:A:74:ARG:NH1	2.33	0.58
1:A:140:ARG:NH1	4:A:903:HOH:O	2.26	0.57
1:A:635:GLU:OE1	1:A:644:HIS:NE2	2.35	0.47
1:A:2:GLU:HA	1:A:2:GLU:OE1	2.15	0.45
1:A:22:GLY:HA3	1:A:49:LEU:O	2.17	0.44
1:A:283:SER:HA	1:A:315:PHE:O	2.18	0.43
1:A:171:GLU:CD	1:A:175:ARG:HH22	2.21	0.43
1:A:212:PRO:HB3	1:A:230:TRP:CH2	2.54	0.43
1:A:80:GLY:HA3	1:A:151:TYR:CG	2.55	0.41
1:A:552:HIS:O	1:A:601:ARG:HB3	2.21	0.40
1:A:451:LEU:HD12	1:A:622:THR:HG22	2.03	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	690/702 (98%)	670 (97%)	18 (3%)	2 (0%)	41 24

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	322	SER
1	A	471	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	A	569/576 (99%)	561 (99%)	8 (1%)	67 53

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	110	ASP
1	A	111	TYR
1	A	147	LYS
1	A	362	LYS
1	A	499	TYR
1	A	631	VAL
1	A	674	GLU
1	A	675	SER

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](#)

2 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	GOL	A	801	-	5,5,5	0.24	0	5,5,5	0.38	0
3	PEG	A	802	-	6,6,6	0.32	0	5,5,5	0.55	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. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	A	801	-	-	0/4/4/4	-
3	PEG	A	802	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	802	PEG	O1-C1-C2-O2
3	A	802	PEG	O2-C3-C4-O4
3	A	802	PEG	C4-C3-O2-C2

There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	802	PEG	4	0

## 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	691/702 (98%)	-0.39	9 (1%) 77 81	7, 12, 30, 68	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	675	SER	4.0
1	A	640	THR	3.4
1	A	626	ASP	3.2
1	A	677	HIS	3.2
1	A	674	GLU	2.8
1	A	14	GLY	2.7
1	A	641	GLY	2.7
1	A	642	GLU	2.7
1	A	661	GLU	2.2

### 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<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
3	PEG	A	802	7/7	0.80	0.14	28,30,32,37	0
2	GOL	A	801	6/6	0.94	0.14	9,9,10,11	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.