



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 07:20 am BST

PDB ID : 5AYB  
Title : Crystal structure of GH1 Beta-Glucosidase TD2F2 N223G mutant  
Authors : Jo, T.; Manninen, J.A.; Matsuzawa, T.; Uchiyama, T.; Yaoi, K.; Arakawa, T.; Fushinobu, S.  
Deposited on : 2015-08-12  
Resolution : 1.80 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 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 : 2.11  
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

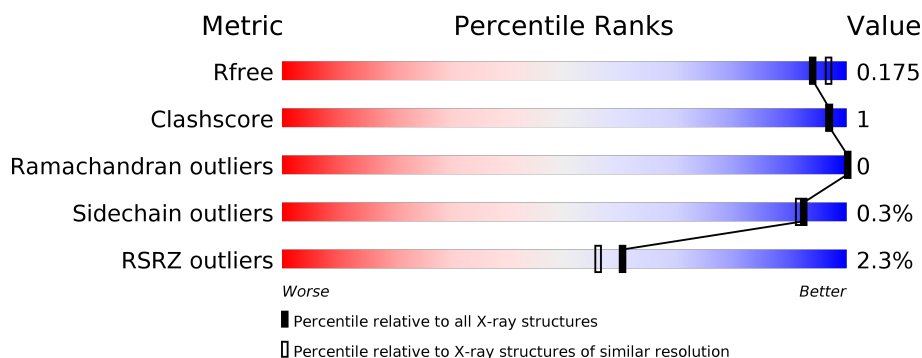
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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  $\geq 3$ , 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	457	<div> <div>2%</div> <div>93%</div> <div>.</div> <div>.</div> </div>

## 2 Entry composition [i](#)

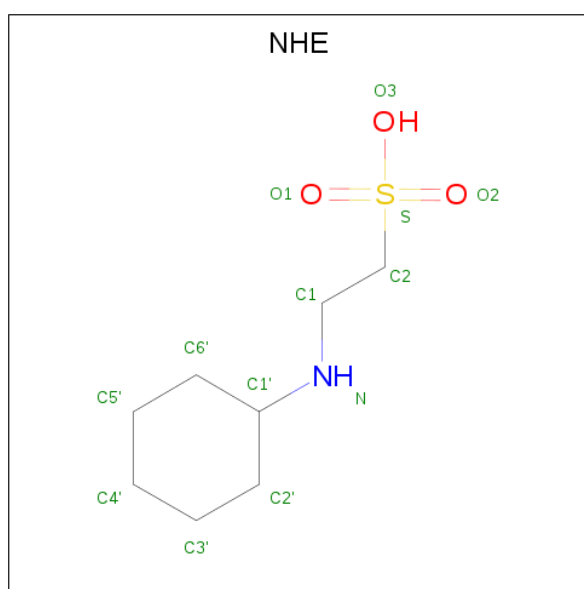
There are 5 unique types of molecules in this entry. The entry contains 3764 atoms, of which 0 are hydrogens and 0 are deuteriums.

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-GLUCOSIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	440	Total	C	N	O	S	0	1	0
			3459	2208	601	637	13			

- Molecule 2 is 2-[N-CYCLOHEXYLAMINO]ETHANE SULFONIC ACID (three-letter code: NHE) (formula:  $C_8H_{17}NO_3S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			13	8	1	3	1		

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Na	0	0
			1	1		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	283	Total	O	0	0
			283	283		



- Molecule 1: BETA-GLUCOSIDASE



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	68.90Å 69.84Å 96.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 1.80 39.41 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.8 (50.00-1.80) 99.9 (39.41-1.80)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.76 (at 1.81Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, $R_{free}$	0.138 , 0.164 0.153 , 0.175	Depositor DCC
$R_{free}$ test set	2190 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.6	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.42 , 51.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.018 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3764	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.21% 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 ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: NHE, NA, 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	A	1.01	3/3566 (0.1%)	0.96	10/4869 (0.2%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	426	GLU	CD-OE2	-6.41	1.18	1.25
1	A	326	GLU	CD-OE2	-5.73	1.19	1.25
1	A	166	GLU	CD-OE2	5.66	1.31	1.25

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	209	ARG	NE-CZ-NH2	-9.36	115.62	120.30
1	A	385	ARG	NE-CZ-NH1	8.68	124.64	120.30
1	A	131	ARG	NE-CZ-NH1	7.81	124.21	120.30
1	A	346	ARG	NE-CZ-NH1	-7.47	116.56	120.30
1	A	385	ARG	NE-CZ-NH2	-6.91	116.85	120.30
1	A	209	ARG	NE-CZ-NH1	5.99	123.30	120.30
1	A	52	ASP	CB-CG-OD1	5.83	123.54	118.30
1	A	131	ARG	NE-CZ-NH2	-5.52	117.54	120.30
1	A	346	ARG	NE-CZ-NH2	5.42	123.01	120.30
1	A	189	ARG	NE-CZ-NH1	5.17	122.89	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	3459	0	3288	6	0
2	A	13	0	17	1	0
3	A	1	0	0	0	0
4	A	8	0	12	0	0
5	A	283	0	0	0	0
All	All	3764	0	3317	6	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 1.

All (6) 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:227:THR:O	2:A:501:NHE:HC'1	2.14	0.46
1:A:187:LEU:HD21	1:A:308:LEU:HD13	1.97	0.46
1:A:5:ARG:C	1:A:384:GLN:HE22	2.20	0.45
1:A:308:LEU:HD23	1:A:308:LEU:C	2.40	0.41
1:A:82:TRP:HB3	1:A:83:PRO:HD3	2.02	0.41
1:A:82:TRP:N	1:A:83:PRO:CD	2.84	0.41

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	439/457 (96%)	427 (97%)	12 (3%)	0	100	100



There are no Ramachandran outliers to report.

### 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	344/355 (97%)	343 (100%)	1 (0%)	92	91

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	189	ARG

Some 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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 1 is 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$
4	EDO	A	504	-	3,3,3	0.92	0	2,2,2	0.45	0
4	EDO	A	503	-	3,3,3	0.32	0	2,2,2	0.55	0
2	NHE	A	501	-	13,13,13	0.96	1 (7%)	16,17,17	2.69	6 (37%)

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
4	EDO	A	504	-	-	0/1/1/1	-
4	EDO	A	503	-	-	1/1/1/1	-
2	NHE	A	501	-	-	1/7/15/15	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	501	NHE	C2-S	-2.44	1.74	1.77

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	501	NHE	C4'-C3'-C2'	5.34	122.30	111.42
2	A	501	NHE	C6'-C1'-C2'	5.28	119.97	110.82
2	A	501	NHE	C5'-C6'-C1'	4.22	119.05	111.11
2	A	501	NHE	C1-N-C1'	3.49	120.98	114.14
2	A	501	NHE	O1-S-C2	-2.37	104.06	106.92
2	A	501	NHE	C5'-C4'-C3'	2.23	118.05	111.18

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	501	NHE	C6'-C1'-N-C1
4	A	503	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	NHE	1	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	440/457 (96%)	-0.33	10 (2%) 60 56	12, 17, 34, 67	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	275	ALA	4.9
1	A	318	ASP	3.1
1	A	276	ASN	3.1
1	A	277	PRO	3.0
1	A	278	GLU	2.9
1	A	315	ASP	2.8
1	A	305	ASP	2.8
1	A	317	PRO	2.4
1	A	274	GLN	2.4
1	A	307	SER	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 carbohydrates 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
4	EDO	A	504	4/4	0.81	0.13	27,29,30,35	0
3	NA	A	502	1/1	0.90	0.25	40,40,40,40	0
4	EDO	A	503	4/4	0.94	0.09	21,25,27,29	0
2	NHE	A	501	13/13	0.97	0.08	16,19,26,27	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.