



# wwPDB X-ray Structure Validation Summary Report

Apr 11, 2023 – 02:52 pm BST

PDB ID : 7ZEQ  
Title : Apo crystal structure of beta-xylosidase from Thermotoga maritima  
Authors : Gloster, T.M.; Foltanyi, F.  
Deposited on : 2022-03-31  
Resolution : 1.89 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 : 1.8.4, CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.32.2  
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) : 2.32.2

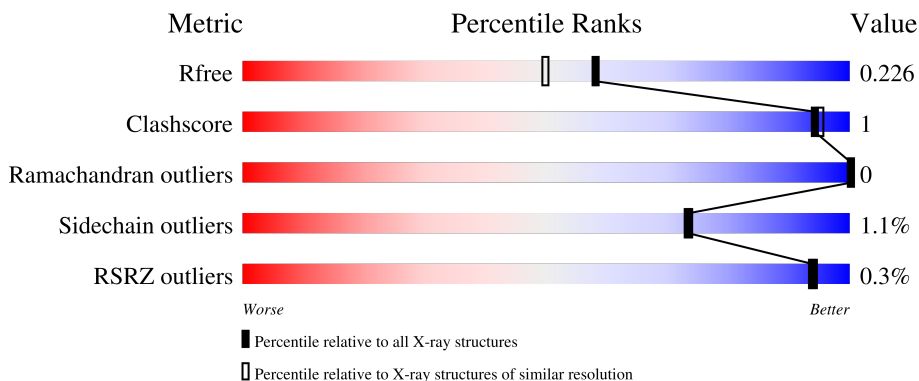
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.89 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	X	801	 90% 6%
1	Y	801	 91% 5%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 11961 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-xylosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	X	754	5868	3740	1001	1104	23	0	2	0
1	Y	759	5895	3757	1010	1105	23	0	2	0

There are 46 discrepancies between the modelled and reference sequences:

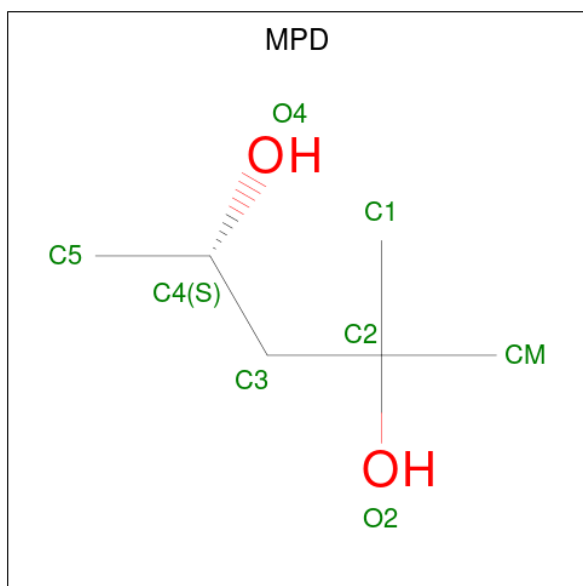
Chain	Residue	Modelled	Actual	Comment	Reference
X	779	ALA	-	expression tag	UNP R4NX63
X	780	ALA	-	expression tag	UNP R4NX63
X	781	ALA	-	expression tag	UNP R4NX63
X	782	LEU	-	expression tag	UNP R4NX63
X	783	GLU	-	expression tag	UNP R4NX63
X	784	GLU	-	expression tag	UNP R4NX63
X	785	ASN	-	expression tag	UNP R4NX63
X	786	LEU	-	expression tag	UNP R4NX63
X	787	TYR	-	expression tag	UNP R4NX63
X	788	PHE	-	expression tag	UNP R4NX63
X	789	GLN	-	expression tag	UNP R4NX63
X	790	GLY	-	expression tag	UNP R4NX63
X	791	ALA	-	expression tag	UNP R4NX63
X	792	HIS	-	expression tag	UNP R4NX63
X	793	HIS	-	expression tag	UNP R4NX63
X	794	HIS	-	expression tag	UNP R4NX63
X	795	HIS	-	expression tag	UNP R4NX63
X	796	HIS	-	expression tag	UNP R4NX63
X	797	HIS	-	expression tag	UNP R4NX63
X	798	HIS	-	expression tag	UNP R4NX63
X	799	HIS	-	expression tag	UNP R4NX63
X	800	HIS	-	expression tag	UNP R4NX63
X	801	HIS	-	expression tag	UNP R4NX63
Y	779	ALA	-	expression tag	UNP R4NX63
Y	780	ALA	-	expression tag	UNP R4NX63

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
Y	781	ALA	-	expression tag	UNP R4NX63
Y	782	LEU	-	expression tag	UNP R4NX63
Y	783	GLU	-	expression tag	UNP R4NX63
Y	784	GLU	-	expression tag	UNP R4NX63
Y	785	ASN	-	expression tag	UNP R4NX63
Y	786	LEU	-	expression tag	UNP R4NX63
Y	787	TYR	-	expression tag	UNP R4NX63
Y	788	PHE	-	expression tag	UNP R4NX63
Y	789	GLN	-	expression tag	UNP R4NX63
Y	790	GLY	-	expression tag	UNP R4NX63
Y	791	ALA	-	expression tag	UNP R4NX63
Y	792	HIS	-	expression tag	UNP R4NX63
Y	793	HIS	-	expression tag	UNP R4NX63
Y	794	HIS	-	expression tag	UNP R4NX63
Y	795	HIS	-	expression tag	UNP R4NX63
Y	796	HIS	-	expression tag	UNP R4NX63
Y	797	HIS	-	expression tag	UNP R4NX63
Y	798	HIS	-	expression tag	UNP R4NX63
Y	799	HIS	-	expression tag	UNP R4NX63
Y	800	HIS	-	expression tag	UNP R4NX63
Y	801	HIS	-	expression tag	UNP R4NX63

- Molecule 2 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	X	1	Total	C	O	0	0
			8	6	2		
2	Y	1	Total	C	O	0	0
			8	6	2		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	X	79	Total	O	0	0
			79	79		
3	Y	103	Total	O	0	0
			103	103		



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	179.08Å 98.59Å 100.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	55.40 – 1.89 55.40 – 1.89	Depositor EDS
% Data completeness (in resolution range)	99.9 (55.40-1.89) 99.9 (55.40-1.89)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.31 (at 1.90Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.194 , 0.226 0.194 , 0.226	Depositor DCC
$R_{free}$ test set	7250 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	31.2	Xtrriage
Anisotropy	0.561	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 34.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.004 for -h,l,k	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	11961	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.32% 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: MPD

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	X	0.44	1/5979 (0.0%)	0.75	7/8081 (0.1%)
1	Y	0.44	0/6006	0.75	7/8120 (0.1%)
All	All	0.44	1/11985 (0.0%)	0.75	14/16201 (0.1%)

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	X	0	2
1	Y	0	1
All	All	0	3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	X	210	GLU	CD-OE2	5.10	1.31	1.25

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	X	650	ARG	NE-CZ-NH2	-7.58	116.51	120.30
1	Y	269	ARG	NE-CZ-NH2	-6.92	116.84	120.30
1	Y	164	ARG	NE-CZ-NH2	-6.76	116.92	120.30
1	X	650	ARG	NE-CZ-NH1	6.36	123.48	120.30
1	Y	650	ARG	NE-CZ-NH1	6.34	123.47	120.30

There are no chirality outliers.

All (3) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	X	189	GLY	Peptide
1	X	573	GLY	Peptide
1	Y	505	ASP	Peptide

## 5.2 Too-close contacts [i](#)

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	X	5868	0	5898	16	0
1	Y	5895	0	5915	15	0
2	X	8	0	14	0	0
2	Y	8	0	14	0	0
3	X	79	0	0	0	0
3	Y	103	0	0	1	0
All	All	11961	0	11841	30	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.

The worst 5 of 30 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:X:656:VAL:HG11	1:X:662:VAL:HG22	1.53	0.89
1:Y:164:ARG:HD3	1:Y:167:GLU:OE2	1.89	0.72
1:Y:370[B]:ARG:HH12	1:Y:457:PRO:HD2	1.55	0.71
1:X:653:PRO:HG2	1:X:656:VAL:HG12	1.73	0.70
1:X:130:ILE:HG12	1:X:179:MET:CE	2.31	0.60

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	X	752/801 (94%)	737 (98%)	15 (2%)	0	100	100
1	Y	757/801 (94%)	748 (99%)	9 (1%)	0	100	100
All	All	1509/1602 (94%)	1485 (98%)	24 (2%)	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	X	624/678 (92%)	618 (99%)	6 (1%)	76	76
1	Y	624/678 (92%)	616 (99%)	8 (1%)	69	68
All	All	1248/1356 (92%)	1234 (99%)	14 (1%)	73	73

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	Y	325	ASP
1	Y	355	GLU
1	Y	770	LYS
1	Y	607	VAL
1	Y	656	VAL

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	MPD	X	901	-	7,7,7	0.14	0	9,10,10	0.51	0
2	MPD	Y	901	-	7,7,7	0.11	0	9,10,10	0.76	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	MPD	X	901	-	-	0/5/5/5	-
2	MPD	Y	901	-	-	0/5/5/5	-

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 [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	X	754/801 (94%)	-0.37	1 (0%) 95   95	25, 35, 54, 77	0
1	Y	759/801 (94%)	-0.28	3 (0%) 92   93	24, 34, 54, 72	0
All	All	1513/1602 (94%)	-0.32	4 (0%) 94   94	24, 34, 54, 77	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	X	449	ILE	4.6
1	Y	445	LEU	3.3
1	Y	449	ILE	3.0
1	Y	408	VAL	2.1

### 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(Å <sup>2</sup> )	Q<0.9
2	MPD	Y	901	8/8	0.92	0.20	43,47,57,58	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	MPD	X	901	8/8	0.95	0.23	41,53,61,65	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.