



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 17, 2021 – 04:07 PM EDT

PDB ID : 7KMQ  
Title : Crystal structure of the GH95 alpha-L-1,2-fucosidase (Xac1774) from Xanthomonas citri  
Authors : Vieira, P.S.; Murakami, M.T.  
Deposited on : 2020-11-03  
Resolution : 2.04 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 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	:	2.18
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)	:	2.18

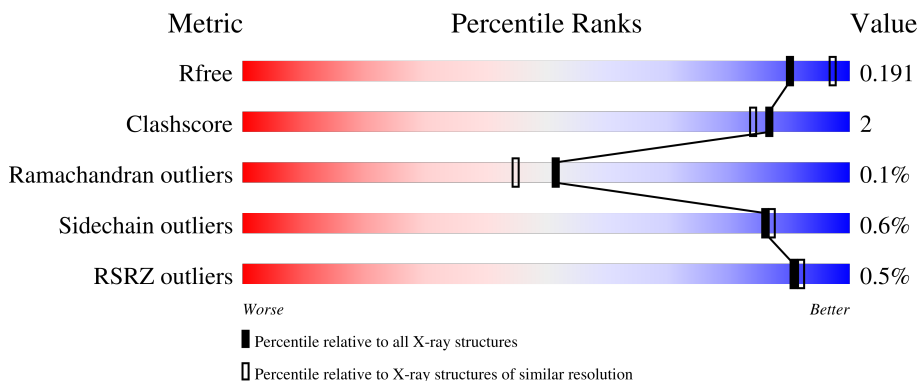
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*



The reported resolution of this entry is 2.04 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  $\geq 3$ , 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	786	 89% 6% •
1	B	786	 90% 6% •

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 11990 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 Glyco\_hyd\_65N\_2 domain-containing protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	753	Total	C	N	O	S	0	0	0
			5873	3705	1073	1075	20			
1	B	754	Total	C	N	O	S	0	0	0
			5877	3708	1074	1075	20			

There are 42 discrepancies between the modelled and reference sequences:

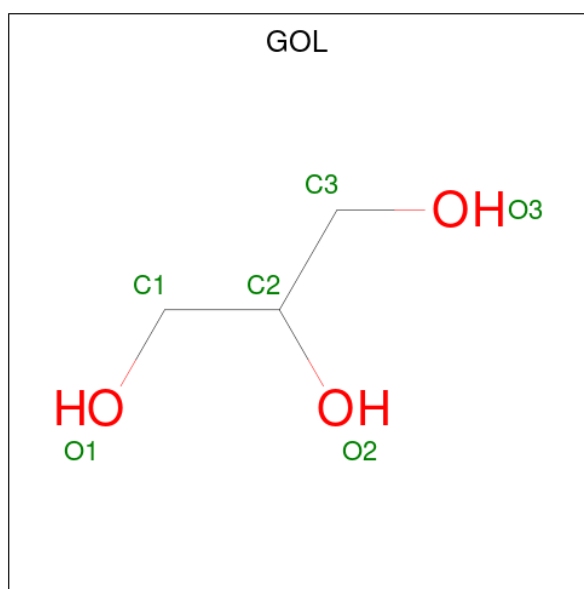
Chain	Residue	Modelled	Actual	Comment	Reference
A	5	MET	-	initiating methionine	UNP Q8PLM1
A	6	GLY	-	expression tag	UNP Q8PLM1
A	7	SER	-	expression tag	UNP Q8PLM1
A	8	SER	-	expression tag	UNP Q8PLM1
A	9	HIS	-	expression tag	UNP Q8PLM1
A	10	HIS	-	expression tag	UNP Q8PLM1
A	11	HIS	-	expression tag	UNP Q8PLM1
A	12	HIS	-	expression tag	UNP Q8PLM1
A	13	HIS	-	expression tag	UNP Q8PLM1
A	14	HIS	-	expression tag	UNP Q8PLM1
A	15	SER	-	expression tag	UNP Q8PLM1
A	16	SER	-	expression tag	UNP Q8PLM1
A	17	GLY	-	expression tag	UNP Q8PLM1
A	18	LEU	-	expression tag	UNP Q8PLM1
A	19	VAL	-	expression tag	UNP Q8PLM1
A	20	PRO	-	expression tag	UNP Q8PLM1
A	21	ARG	-	expression tag	UNP Q8PLM1
A	22	GLY	-	expression tag	UNP Q8PLM1
A	23	SER	-	expression tag	UNP Q8PLM1
A	24	HIS	-	expression tag	UNP Q8PLM1
A	25	MET	-	expression tag	UNP Q8PLM1
B	5	MET	-	initiating methionine	UNP Q8PLM1
B	6	GLY	-	expression tag	UNP Q8PLM1
B	7	SER	-	expression tag	UNP Q8PLM1
B	8	SER	-	expression tag	UNP Q8PLM1

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	9	HIS	-	expression tag	UNP Q8PLM1
B	10	HIS	-	expression tag	UNP Q8PLM1
B	11	HIS	-	expression tag	UNP Q8PLM1
B	12	HIS	-	expression tag	UNP Q8PLM1
B	13	HIS	-	expression tag	UNP Q8PLM1
B	14	HIS	-	expression tag	UNP Q8PLM1
B	15	SER	-	expression tag	UNP Q8PLM1
B	16	SER	-	expression tag	UNP Q8PLM1
B	17	GLY	-	expression tag	UNP Q8PLM1
B	18	LEU	-	expression tag	UNP Q8PLM1
B	19	VAL	-	expression tag	UNP Q8PLM1
B	20	PRO	-	expression tag	UNP Q8PLM1
B	21	ARG	-	expression tag	UNP Q8PLM1
B	22	GLY	-	expression tag	UNP Q8PLM1
B	23	SER	-	expression tag	UNP Q8PLM1
B	24	HIS	-	expression tag	UNP Q8PLM1
B	25	MET	-	expression tag	UNP Q8PLM1

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



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

Continued on next page...

*Continued from previous page...*

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

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Ca	0	0
			1	1		
3	B	1	Total	Ca	0	0
			1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	100	Total	O	0	0
			100	100		
4	B	102	Total	O	0	0
			102	102		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	103.60Å 103.52Å 173.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.78 – 2.04 19.78 – 2.04	Depositor EDS
% Data completeness (in resolution range)	99.5 (19.78-2.04) 99.4 (19.78-2.04)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.45 (at 2.04Å)	Xtriage
Refinement program	PHENIX dev_3139	Depositor
R, $R_{free}$	0.174 , 0.191 0.174 , 0.191	Depositor DCC
$R_{free}$ test set	5832 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.0	Xtriage
Anisotropy	0.030	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 35.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.023 for k,h,-l	Xtriage
Reported twinning fraction	0.480 for k,h,-l	Depositor
Outliers	0 of 117474 reflections	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	11990	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.92% 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: GOL, CA

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	0.24	0/6016	0.44	0/8186
1	B	0.24	0/6021	0.44	0/8196
All	All	0.24	0/12037	0.44	0/16382

There are no bond length outliers.

There are no bond angle outliers.

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	5873	0	5765	25	0
1	B	5877	0	5771	25	0
2	A	18	0	24	2	0
2	B	18	0	24	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	100	0	0	0	0
4	B	102	0	0	0	0
All	All	11990	0	11584	50	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.



The worst 5 of 50 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:89:ALA:HB3	1:A:136:PRO:HD2	1.83	0.60
1:B:124:ASP:HA	1:B:128:LEU:HG	1.85	0.59
1:B:749:ALA:HB3	1:B:780:VAL:HG12	1.84	0.59
1:B:490:LEU:HD11	1:B:556:MET:HG2	1.86	0.57
1:B:565:GLN:O	1:B:566:LEU:HB3	2.05	0.57

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	749/786 (95%)	717 (96%)	31 (4%)	1 (0%)	51	45
1	B	752/786 (96%)	719 (96%)	32 (4%)	1 (0%)	51	45
All	All	1501/1572 (96%)	1436 (96%)	63 (4%)	2 (0%)	51	45

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	398	ILE
1	B	398	ILE

### 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	597/623 (96%)	593 (99%)	4 (1%)	84	84
1	B	597/623 (96%)	594 (100%)	3 (0%)	88	89
All	All	1194/1246 (96%)	1187 (99%)	7 (1%)	86	87

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	723	ARG
1	B	36	LEU
1	B	299	ARG
1	B	149	ARG
1	A	299	ARG

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](#)

Of 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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
2	GOL	A	801	-	5,5,5	0.92	0	5,5,5	0.98	0
2	GOL	B	802	-	5,5,5	0.95	0	5,5,5	0.91	0
2	GOL	A	802	-	5,5,5	1.03	0	5,5,5	0.97	0
2	GOL	B	801	-	5,5,5	0.94	0	5,5,5	0.99	0
2	GOL	B	803	-	5,5,5	0.90	0	5,5,5	1.01	0
2	GOL	A	803	-	5,5,5	0.90	0	5,5,5	0.97	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	GOL	A	801	-	-	2/4/4/4	-
2	GOL	B	802	-	-	0/4/4/4	-
2	GOL	A	802	-	-	4/4/4/4	-
2	GOL	B	801	-	-	2/4/4/4	-
2	GOL	B	803	-	-	2/4/4/4	-
2	GOL	A	803	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	802	GOL	C1-C2-C3-O3
2	B	803	GOL	O1-C1-C2-O2
2	B	803	GOL	O1-C1-C2-C3
2	A	802	GOL	O2-C2-C3-O3
2	A	801	GOL	C1-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	801	GOL	1	0
2	A	802	GOL	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	753/786 (95%)	-0.14	3 (0%) 92 93	28, 37, 47, 59	0
1	B	754/786 (95%)	-0.17	4 (0%) 91 92	28, 36, 48, 68	0
All	All	1507/1572 (95%)	-0.15	7 (0%) 91 92	28, 37, 48, 68	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	174	GLY	3.7
1	B	566	LEU	3.7
1	B	304	LEU	2.5
1	B	36	LEU	2.2
1	B	332	GLU	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( $\text{\AA}^2$ )	Q<0.9
2	GOL	B	802	6/6	0.82	0.25	35,37,45,49	0
2	GOL	A	802	6/6	0.85	0.20	32,38,44,52	0
2	GOL	B	803	6/6	0.86	0.16	38,43,43,54	0
2	GOL	A	801	6/6	0.94	0.11	33,37,38,48	0
2	GOL	B	801	6/6	0.94	0.12	36,39,41,45	0
3	CA	B	804	1/1	0.94	0.12	43,43,43,43	0
2	GOL	A	803	6/6	0.95	0.13	37,40,42,43	0
3	CA	A	804	1/1	0.98	0.04	47,47,47,47	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.