

# wwPDB X-ray Structure Validation Summary Report (i)

#### Nov 19, 2023 – 10:07 PM JST

PDB ID	:	7C25
Title	:	Glycosidase Wild Type at pH8.0
Authors	:	Tagami, T.; Kikuchi, A.; Okuyama, M.; Kimura, A.
Deposited on		
Resolution	:	1.50  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 A user guide is available at https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

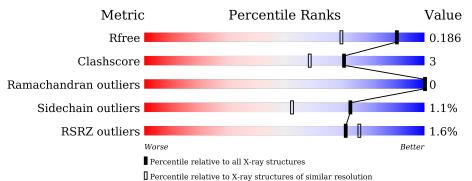
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal 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:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 1.50 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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
R <sub>free</sub>	130704	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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	А	405	85%	6% •	8%			

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
4	NH4	А	406	-	-	Х	-



## 2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3384 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 Isomaltose glucohydrolase.

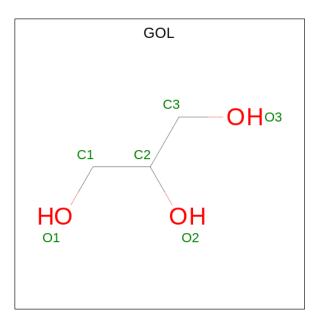
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	А	372	Total 2931	C 1856	N 533	O 533	S 9	0	8	0

Chain	Residue	Modelled	Actual	Comment	Reference
А	-19	MET	-	expression tag	UNP D2PPM8
А	-18	GLY	-	expression tag	UNP D2PPM8
А	-17	SER	-	expression tag	UNP D2PPM8
А	-16	SER	-	expression tag	UNP D2PPM8
A	-15	HIS	-	expression tag	UNP D2PPM8
А	-14	HIS	-	expression tag	UNP D2PPM8
А	-13	HIS	-	expression tag	UNP D2PPM8
А	-12	HIS	-	expression tag	UNP D2PPM8
А	-11	HIS	-	expression tag	UNP D2PPM8
A	-10	HIS	-	expression tag	UNP D2PPM8
А	-9	SER	-	expression tag	UNP D2PPM8
A	-8	SER	-	expression tag	UNP D2PPM8
А	-7	GLY	-	expression tag	UNP D2PPM8
А	-6	LEU	-	expression tag	UNP D2PPM8
А	-5	VAL	-	expression tag	UNP D2PPM8
A	-4	PRO	-	expression tag	UNP D2PPM8
А	-3	ARG	-	expression tag	UNP D2PPM8
А	-2	GLY	-	expression tag	UNP D2PPM8
А	-1	SER	-	expression tag	UNP D2PPM8
А	0	HIS	-	expression tag	UNP D2PPM8

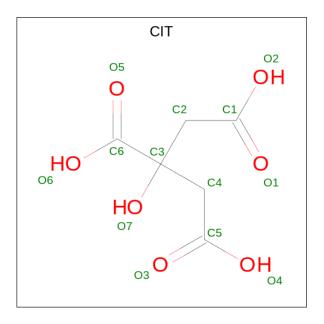
There are 20 discrepancies between the modelled and reference sequences:

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





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 6 & 3 & 3 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total}  \text{C}  \text{O} \\ 6  3  3 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 6 & 3 & 3 \end{array}$	0	0
2	А	1	$\begin{array}{ccc} \text{Total}  \text{C}  \text{O} \\ 6  3  3 \end{array}$	0	0





$\mathbf{N}$	Iol	Chain	Residues	Atoms		ZeroOcc	AltConf	
	3	А	1	Total 13	С 6	O 7	0	0

• Molecule 4 is AMMONIUM ION (three-letter code: NH4) (formula:  $H_4N$ ).

NH4	
NH <sup>+</sup> N	

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	1	Total N 1 1	0	0

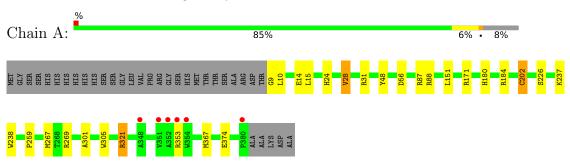
• Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	415	Total O 415 415	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: Isomaltose glucohydrolase



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 4 21 2	Depositor
Cell constants	105.46Å $105.46$ Å $89.95$ Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	47.21 - 1.50	Depositor
Resolution (A)	47.16 - 1.50	EDS
% Data completeness	92.6 (47.21-1.50)	Depositor
(in resolution range)	92.6 (47.16-1.50)	EDS
R <sub>merge</sub>	(Not available)	Depositor
R <sub>sym</sub>	0.08	Depositor
$< I/\sigma(I) > 1$	$1.37 (at 1.50 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
D D.	0.136 , $0.173$	Depositor
$R, R_{free}$	0.151 , $0.186$	DCC
$R_{free}$ test set	3721 reflections $(4.99%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	19.7	Xtriage
Anisotropy	0.037	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.38 , $44.5$	EDS
L-test for twinning <sup>2</sup>	$ < L >=0.51, < L^2>=0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	3384	wwPDB-VP
Average B, all atoms $(Å^2)$	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.09% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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, CSX, CIT, NH4  $\,$ 

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).

Mal	Chain	Bo	nd lengths	Bond angles		
Mol Chair		RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.90	3/3006~(0.1%)	1.03	8/4108~(0.2%)	

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	А	9	GLY	N-CA	8.50	1.58	1.46
1	А	9	GLY	CA-C	6.40	1.62	1.51
1	А	171	ARG	CD-NE	-5.64	1.36	1.46

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
1	А	171	ARG	NE-CZ-NH2	-10.84	114.88	120.30
1	А	321	ARG	NE-CZ-NH2	-8.12	116.24	120.30
1	А	88	ARG	NE-CZ-NH2	-5.99	117.31	120.30
1	А	88	ARG	NE-CZ-NH1	5.30	122.95	120.30
1	А	184	ARG	NE-CZ-NH2	-5.30	117.65	120.30

There are no chirality outliers.

There are no planarity outliers.

#### 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	А	2931	0	2766	20	0
2	А	24	0	32	1	0
3	А	13	0	5	3	0
4	А	1	0	0	2	0
5	А	415	0	0	3	0
All	All	3384	0	2803	20	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 3.

The worst 5 of 20 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:374:GLU:OE1	4:A:406:NH4:N	1.93	1.02	
1:A:24:HIS:CE1	3:A:405:CIT:H21	2.08	0.89	
1:A:31:ARG:HD2	3:A:405:CIT:O4	1.77	0.85	
1:A:24:HIS:HE1	3:A:405:CIT:H21	1.43	0.82	
1:A:226[B]:SER:OG	2:A:402:GOL:O2	2.07	0.71	

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	376/405~(93%)	365~(97%)	11 (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	А	286/306~(94%)	283~(99%)	3 (1%)	76 57	

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	28	VAL
1	А	56	ASP
1	А	353	ARG

Sometimes side chains can be flipped to improve hydrogen bonding and reduce clashes. There are no such side chains 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)

2 non-standard protein/DNA/RNA residues 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	Iol Type Chain Res	Link	B	ond leng	$_{ m gths}$	I	Bond an	gles		
IVIOI	туре	Chain	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
1	CSX	А	202[B]	1	$3,\!6,\!7$	1.03	0	$1,\!6,\!8$	2.42	1 (100%)
1	CSX	А	202[A]	1	3,6,7	0.90	0	1,6,8	2.39	1 (100%)

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
1	CSX	А	202[B]	1	-	0/1/5/7	-
1	CSX	А	202[A]	1	-	0/1/5/7	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$Observed(^{o})$	$Ideal(^{o})$
1	А	202[B]	CSX	CA-CB-SG	-2.42	108.08	113.36
1	А	202[A]	CSX	CA-CB-SG	2.39	118.56	113.36

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	А	202[A]	CSX	1	0

#### 5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

#### 5.6 Ligand geometry (i)

Of 6 ligands modelled in this entry, 1 is modelled with single atom - leaving 5 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	Turne	Chain	Res	Link	Bond lengths			Bond angles		
MOI	Type	Unam	nes	LIIIK	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2
3	CIT	А	405	-	12,12,12	1.87	3 (25%)	$17,\!17,\!17$	1.47	3 (17%)
2	GOL	А	402	-	$5,\!5,\!5$	0.24	0	$5,\!5,\!5$	0.87	0
2	GOL	А	403	-	$5,\!5,\!5$	0.37	0	$5,\!5,\!5$	0.82	0
2	GOL	А	404	-	$5,\!5,\!5$	0.25	0	$5,\!5,\!5$	0.60	0



Mol	Type	Chain	Res	Link	Bond lengths Bond angles			les		
	Type	Unam	nes		Counts	RMSZ	# Z >2	Counts	RMSZ	# Z  > 2
2	GOL	А	401	-	$5,\!5,\!5$	0.21	0	$5,\!5,\!5$	0.30	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
3	CIT	А	405	-	-	6/16/16/16	-
2	GOL	А	402	-	-	0/4/4/4	-
2	GOL	А	403	-	-	3/4/4/4	-
2	GOL	А	404	-	-	4/4/4/4	-
2	GOL	А	401	-	-	0/4/4/4	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	А	405	CIT	C3-C6	4.73	1.58	1.53
3	А	405	CIT	O1-C1	2.68	1.31	1.22
3	А	405	CIT	O4-C5	-2.04	1.23	1.30

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^{o})$	$Ideal(^{o})$
3	А	405	CIT	C3-C4-C5	2.86	120.75	113.81
3	А	405	CIT	C4-C3-C6	-2.52	104.68	110.11
3	А	405	CIT	O7-C3-C2	-2.06	104.58	109.40

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	А	404	GOL	C1-C2-C3-O3
3	А	405	CIT	C2-C3-C6-O5
3	А	405	CIT	C2-C3-C6-O6
3	А	405	CIT	O7-C3-C6-O5
3	А	405	CIT	O7-C3-C6-O6

There are no ring outliers.

2 monomers are involved in 4 short contacts:



Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	А	405	CIT	3	0
2	А	402	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,  $95^{th}$  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	$\langle RSRZ \rangle$	#RSRZ>2	$OWAB(Å^2)$	Q < 0.9
1	А	371/405~(91%)	-0.31	6 (1%) 72 77	15, 21, 38, 83	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	380	PRO	5.6
1	А	354	TRP	3.5
1	А	352	ALA	3.4
1	А	353	ARG	2.6
1	А	351	VAL	2.6

#### 6.2 Non-standard residues in protein, DNA, RNA chains (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	$\mathbf{B} ext{-factors}(\mathrm{\AA}^2)$	Q < 0.9
1	CSX	А	202[A]	7/8	0.99	0.08	18,18,19,23	7
1	CSX	А	202[B]	7/8	0.99	0.08	18,19,22,24	7

#### 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,



7C25
------

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\operatorname{B-factors}(\operatorname{\AA}^2)$	$\mathbf{Q} \!\!<\!\! 0.9$
2	GOL	А	404	6/6	0.70	0.15	$53,\!60,\!68,\!72$	0
3	CIT	А	405	13/13	0.88	0.33	38,47,84,93	0
2	GOL	А	402	6/6	0.90	0.11	$26,\!28,\!41,\!43$	0
2	GOL	А	403	6/6	0.92	0.12	37,46,49,50	0
2	GOL	А	401	6/6	0.96	0.09	23,28,39,40	0
4	NH4	А	406	1/1	1.00	0.18	14,14,14,14	0

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.

#### 6.5 Other polymers (i)

There are no such residues in this entry.

