



# wwPDB X-ray Structure Validation Summary Report

Sep 1, 2022 – 06:08 pm BST


PDB ID : 7ZGN  
Title : Plant/insect N-glycan active PNGase  
Authors : Basle, A.; Crouch, L.; Bolam, D.  
Deposited on : 2022-04-04  
Resolution : 1.95 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  
Xtriage (Phenix) : 1.13  
EDS : 2.30  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0267  
CCP4 : 7.1.010 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.30

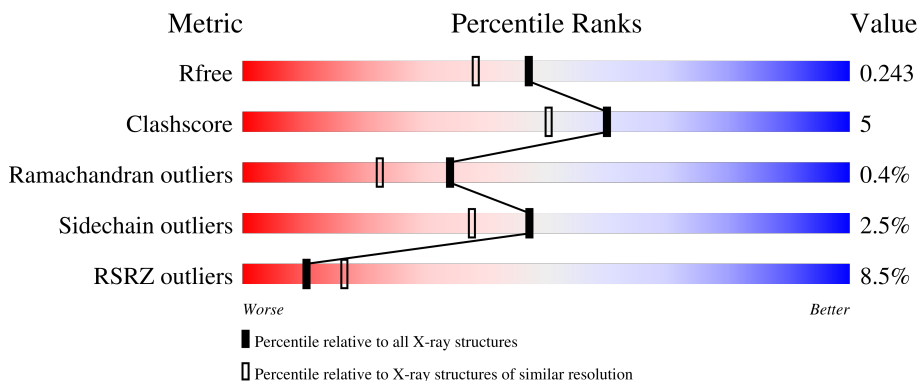
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.95 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	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	559	 4% 86% 8% • 5%
1	B	559	 12% 79% 13% • 5%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 16641 atoms, of which 7984 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 Glpgli family protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	A	531	8149	2652	3987	710	789	11	0	0	0
1	B	529	8156	2652	3997	712	784	11	0	2	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	MET	-	initiating methionine	UNP U6RE59
A	571	LEU	-	expression tag	UNP U6RE59
A	572	GLU	-	expression tag	UNP U6RE59
A	573	HIS	-	expression tag	UNP U6RE59
A	574	HIS	-	expression tag	UNP U6RE59
A	575	HIS	-	expression tag	UNP U6RE59
A	576	HIS	-	expression tag	UNP U6RE59
A	577	HIS	-	expression tag	UNP U6RE59
A	578	HIS	-	expression tag	UNP U6RE59
B	20	MET	-	initiating methionine	UNP U6RE59
B	571	LEU	-	expression tag	UNP U6RE59
B	572	GLU	-	expression tag	UNP U6RE59
B	573	HIS	-	expression tag	UNP U6RE59
B	574	HIS	-	expression tag	UNP U6RE59
B	575	HIS	-	expression tag	UNP U6RE59
B	576	HIS	-	expression tag	UNP U6RE59
B	577	HIS	-	expression tag	UNP U6RE59
B	578	HIS	-	expression tag	UNP U6RE59

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	149	Total	O	0	0
			149	149		

*Continued on next page...*

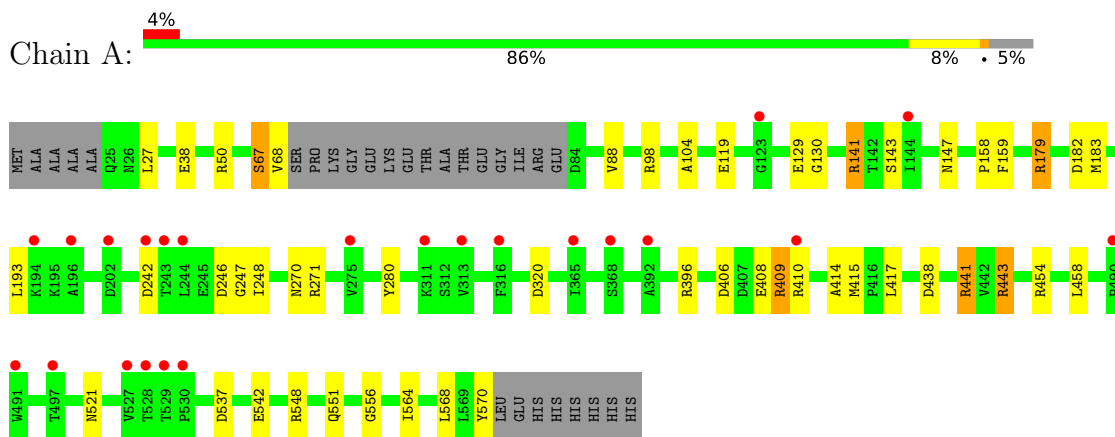
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
2	B	187	Total 187	O 187	0	0

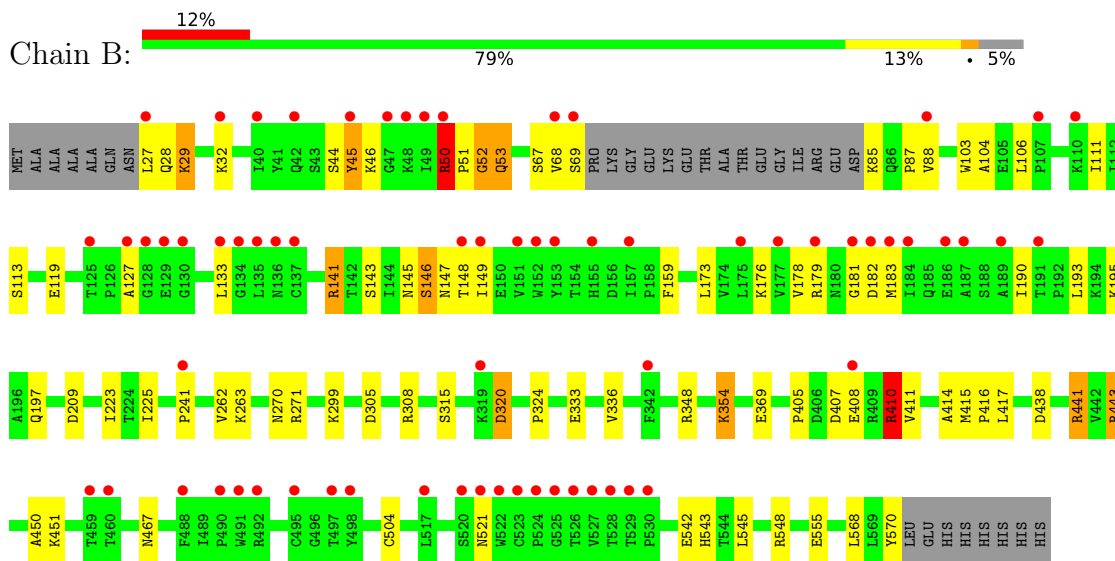
### 3 Residue-property plots [i](#)

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: Glpgli family protein



- Molecule 1: Glpgli family protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	59.05Å 100.55Å 180.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.38 – 1.95 44.34 – 1.95	Depositor EDS
% Data completeness (in resolution range)	99.8 (44.38-1.95) 99.8 (44.34-1.95)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.78 (at 1.95Å)	Xtriage
Refinement program	REFMAC 5.8.0326	Depositor
R, $R_{free}$	0.189 , 0.239 0.198 , 0.243	Depositor DCC
$R_{free}$ test set	3990 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	36.1	Xtriage
Anisotropy	0.120	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	(Not available) , (Not available)	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	16641	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.87% 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](#)

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	0.60	0/4264	0.93	3/5784 (0.1%)
1	B	0.61	2/4267 (0.0%)	0.92	1/5786 (0.0%)
All	All	0.60	2/8531 (0.0%)	0.92	4/11570 (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	9
1	B	0	11
All	All	0	20

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	119	GLU	CD-OE1	6.08	1.32	1.25
1	B	369	GLU	CD-OE2	6.04	1.32	1.25

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	570	TYR	CA-C-O	-10.27	98.54	120.10
1	A	182	ASP	CB-CA-C	-7.69	95.03	110.40
1	A	454	ARG	NE-CZ-NH2	-6.71	116.94	120.30
1	B	50	ARG	NE-CZ-NH1	6.32	123.46	120.30

There are no chirality outliers.

5 of 20 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	141	ARG	Sidechain
1	A	179	ARG	Sidechain
1	A	271	ARG	Sidechain
1	A	409	ARG	Sidechain
1	A	50	ARG	Sidechain

## 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	A	4162	3987	4089	26	0
1	B	4159	3997	4102	52	0
2	A	149	0	0	3	0
2	B	187	0	0	5	0
All	All	8657	7984	8191	76	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 5.

The worst 5 of 76 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:B:354:LYS:HD3	2:B:667:HOH:O	1.50	1.09
1:B:28:GLN:O	1:B:29:LYS:HG3	1.83	0.78
1:B:53:GLN:HG3	1:B:68:VAL:HG11	1.65	0.77
1:B:113:SER:OG	2:B:601:HOH:O	2.09	0.70
1:A:443:ARG:HH11	1:A:542:GLU:CD	1.96	0.68

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	527/559 (94%)	508 (96%)	18 (3%)	1 (0%)	47	38
1	B	527/559 (94%)	504 (96%)	20 (4%)	3 (1%)	25	14
All	All	1054/1118 (94%)	1012 (96%)	38 (4%)	4 (0%)	34	22

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	521	ASN
1	A	521	ASN
1	B	52	GLY
1	B	146	SER

### 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	448/469 (96%)	443 (99%)	5 (1%)	73	71
1	B	448/469 (96%)	431 (96%)	17 (4%)	33	21
All	All	896/938 (96%)	874 (98%)	22 (2%)	47	38

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	263	LYS
1	B	410	ARG
1	B	354	LYS
1	B	443	ARG
1	B	32	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	476	ASN
1	B	86	GLN
1	B	197	GLN
1	B	412	ASN
1	B	476	ASN

### 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](#)

There are no ligands in this entry.

### 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	531/559 (94%)	0.47	23 (4%) 35 45	26, 44, 73, 95	0
1	B	529/559 (94%)	0.81	67 (12%) 3 6	21, 43, 86, 120	0
All	All	1060/1118 (94%)	0.64	90 (8%) 10 17	21, 43, 80, 120	0

The worst 5 of 90 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	27	LEU	7.9
1	B	45	TYR	7.7
1	B	49	ILE	6.5
1	B	69	SER	5.7
1	B	135	LEU	5.5

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.