



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 20, 2022 – 12:19 PM JST

PDB ID : 7XTL  
Title : Crystal structure of the C-terminal domain of human N-acetylglucosaminyltransferase IVa  
Authors : Miyazaki, T.; Oka, N.; Mori, S.  
Deposited on : 2022-05-17  
Resolution : 1.97 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  
Xtrriage (Phenix) : 1.13  
EDS : 2.30  
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) : 2.30

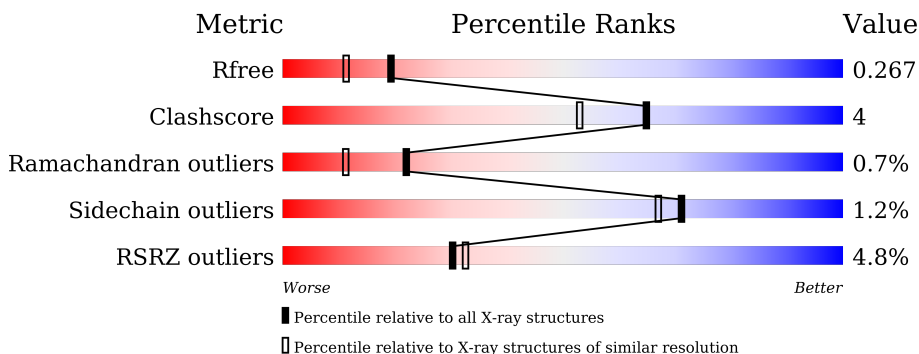
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.97 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	11647 (2.00-1.96)
Clashscore	141614	1014 (1.98-1.98)
Ramachandran outliers	138981	1006 (1.98-1.98)
Sidechain outliers	138945	1006 (1.98-1.98)
RSRZ outliers	127900	11410 (2.00-1.96)

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	172	 3% 77% 9% 14%
1	B	172	 5% 72% 12% 16%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 2359 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 Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A soluble form.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	148	1179	763	192	222	2	0	0	0
1	B	145	1158	748	187	221	2	0	1	0

There are 46 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	364	MET	-	initiating methionine	UNP Q9UM21
A	365	GLY	-	expression tag	UNP Q9UM21
A	366	SER	-	expression tag	UNP Q9UM21
A	367	SER	-	expression tag	UNP Q9UM21
A	368	HIS	-	expression tag	UNP Q9UM21
A	369	HIS	-	expression tag	UNP Q9UM21
A	370	HIS	-	expression tag	UNP Q9UM21
A	371	HIS	-	expression tag	UNP Q9UM21
A	372	HIS	-	expression tag	UNP Q9UM21
A	373	HIS	-	expression tag	UNP Q9UM21
A	374	SER	-	expression tag	UNP Q9UM21
A	375	SER	-	expression tag	UNP Q9UM21
A	376	GLY	-	expression tag	UNP Q9UM21
A	377	LEU	-	expression tag	UNP Q9UM21
A	378	VAL	-	expression tag	UNP Q9UM21
A	379	PRO	-	expression tag	UNP Q9UM21
A	380	ARG	-	expression tag	UNP Q9UM21
A	381	GLY	-	expression tag	UNP Q9UM21
A	382	SER	-	expression tag	UNP Q9UM21
A	383	HIS	-	expression tag	UNP Q9UM21
A	384	MET	-	expression tag	UNP Q9UM21
A	385	ALA	-	expression tag	UNP Q9UM21
A	386	SER	-	expression tag	UNP Q9UM21
B	364	MET	-	initiating methionine	UNP Q9UM21

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	365	GLY	-	expression tag	UNP Q9UM21
B	366	SER	-	expression tag	UNP Q9UM21
B	367	SER	-	expression tag	UNP Q9UM21
B	368	HIS	-	expression tag	UNP Q9UM21
B	369	HIS	-	expression tag	UNP Q9UM21
B	370	HIS	-	expression tag	UNP Q9UM21
B	371	HIS	-	expression tag	UNP Q9UM21
B	372	HIS	-	expression tag	UNP Q9UM21
B	373	HIS	-	expression tag	UNP Q9UM21
B	374	SER	-	expression tag	UNP Q9UM21
B	375	SER	-	expression tag	UNP Q9UM21
B	376	GLY	-	expression tag	UNP Q9UM21
B	377	LEU	-	expression tag	UNP Q9UM21
B	378	VAL	-	expression tag	UNP Q9UM21
B	379	PRO	-	expression tag	UNP Q9UM21
B	380	ARG	-	expression tag	UNP Q9UM21
B	381	GLY	-	expression tag	UNP Q9UM21
B	382	SER	-	expression tag	UNP Q9UM21
B	383	HIS	-	expression tag	UNP Q9UM21
B	384	MET	-	expression tag	UNP Q9UM21
B	385	ALA	-	expression tag	UNP Q9UM21
B	386	SER	-	expression tag	UNP Q9UM21

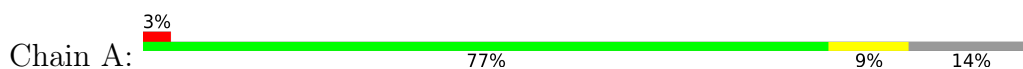
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	17	Total O 17 17	0	0
2	B	5	Total O 5 5	0	0

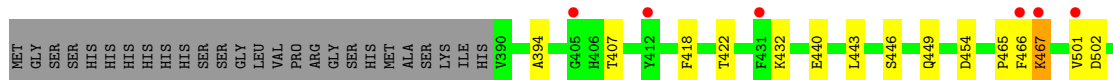
### 3 Residue-property plots [i](#)

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: Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A soluble form



- Molecule 1: Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A soluble form



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	50.70Å 32.49Å 85.63Å 90.00° 94.71° 90.00°	Depositor
Resolution (Å)	50.58 – 1.97 50.53 – 1.97	Depositor EDS
% Data completeness (in resolution range)	99.9 (50.58-1.97) 100.0 (50.53-1.97)	Depositor EDS
$R_{merge}$	0.21	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.25 (at 1.97Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, $R_{free}$	0.220 , 0.266 0.228 , 0.267	Depositor DCC
$R_{free}$ test set	959 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	36.3	Xtrriage
Anisotropy	0.256	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 55.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2359	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.70% 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	0.87	1/1209 (0.1%)	0.94	0/1639
1	B	1.04	4/1187 (0.3%)	0.88	0/1610
All	All	0.96	5/2396 (0.2%)	0.91	0/3249

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	534	THR	C-O	15.96	1.53	1.23
1	B	432	LYS	C-O	11.12	1.44	1.23
1	B	394	ALA	C-O	8.10	1.38	1.23
1	A	479	ASP	CB-CG	5.98	1.64	1.51
1	B	407	THR	C-O	5.84	1.34	1.23

There are no bond angle outliers.

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	A	1179	0	1166	11	0
1	B	1158	0	1139	9	0
2	A	17	0	0	0	0
2	B	5	0	0	0	0
All	All	2359	0	2305	19	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 4.

The worst 5 of 19 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:501:VAL:HG11	1:A:508:ILE:HD11	1.58	0.84
1:A:501:VAL:HG11	1:A:508:ILE:CD1	2.19	0.73
1:B:506:ASN:OD1	1:B:508:ILE:HD12	1.98	0.63
1:A:473:ILE:HG23	1:A:477:THR:HB	1.84	0.59
1:A:501:VAL:CG1	1:A:508:ILE:HD11	2.33	0.56

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	146/172 (85%)	143 (98%)	3 (2%)	0	100	100
1	B	144/172 (84%)	139 (96%)	3 (2%)	2 (1%)	11	3
All	All	290/344 (84%)	282 (97%)	6 (2%)	2 (1%)	22	11

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	467	LYS
1	B	422	THR

### 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	130/150 (87%)	129 (99%)	1 (1%)	81	80
1	B	128/150 (85%)	126 (98%)	2 (2%)	62	56
All	All	258/300 (86%)	255 (99%)	3 (1%)	71	67

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

Mol	Chain	Res	Type
1	A	475	LYS
1	B	449	GLN
1	B	502	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	449	GLN

### 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

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	148/172 (86%)	0.30	5 (3%) 45 48	44, 55, 73, 122	0
1	B	145/172 (84%)	0.58	9 (6%) 20 22	48, 71, 103, 123	0
All	All	293/344 (85%)	0.44	14 (4%) 30 32	44, 61, 100, 123	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	520	ALA	3.7
1	A	470	GLY	3.7
1	B	412	TYR	3.6
1	B	466	PHE	3.5
1	B	405	GLY	3.3

### 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.