



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 06:09 am BST

PDB ID : 5M7R  
Title : Structure of human O-GlcNAc hydrolase  
Authors : Roth, C.; Chan, S.; Offen, W.A.; Hemsworth, G.R.; Willems, L.I.; King, D.;  
Varghese, V.; Britton, R.; Vocadlo, D.J.; Davies, G.J.  
Deposited on : 2016-10-28  
Resolution : 2.35 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  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

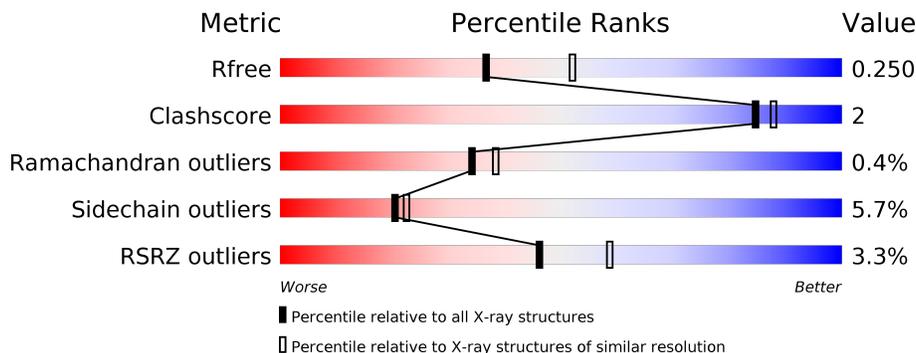
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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 on 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	916	
1	B	916	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7704 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 Protein O-GlcNAcase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	466	3813	2468	628	693	24	0	0	0
1	B	470	3839	2481	634	699	25	0	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	21	Total 21	O 21	0	0
2	B	31	Total 31	O 31	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	102.23Å 102.23Å 285.49Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	96.25 – 2.35 96.25 – 2.35	Depositor EDS
% Data completeness (in resolution range)	100.0 (96.25-2.35) 100.0 (96.25-2.35)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.51 (at 2.34Å)	Xtrriage
Refinement program	REFMAC 5.8.0155, REFMAC 5.8.0155	Depositor
R, $R_{free}$	0.216 , 0.249 0.219 , 0.250	Depositor DCC
$R_{free}$ test set	1909 reflections (2.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	59.3	Xtrriage
Anisotropy	0.630	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 39.5	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	0.96	EDS
Total number of atoms	7704	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	80.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.08% 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	0.85	1/3915 (0.0%)	0.97	15/5302 (0.3%)
1	B	0.66	0/3942	0.84	8/5339 (0.1%)
All	All	0.76	1/7857 (0.0%)	0.91	23/10641 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	677	GLU	CG-CD	6.59	1.61	1.51

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	634	ARG	NE-CZ-NH1	9.74	125.17	120.30
1	A	634	ARG	NE-CZ-NH2	-7.99	116.30	120.30
1	A	167	ARG	CG-CD-NE	7.63	127.82	111.80
1	B	290	ARG	NE-CZ-NH2	-7.62	116.49	120.30
1	A	167	ARG	NE-CZ-NH1	7.53	124.06	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	A	3813	0	3741	22	0
1	B	3839	0	3759	15	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	21	0	0	0	0
2	B	31	0	0	1	0
All	All	7704	0	7500	35	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 35 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:75:MET:HE1	1:A:122:LEU:HD22	1.33	1.06
1:B:596:CYS:HG	1:B:600:ASP:N	1.61	0.98
1:A:75:MET:CE	1:A:122:LEU:HD22	2.16	0.73
1:A:75:MET:HA	1:A:75:MET:HE3	1.77	0.67
1:A:595:ASN:O	1:A:599:LYS:N	2.32	0.62

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	454/916 (50%)	437 (96%)	15 (3%)	2 (0%)	34 38
1	B	458/916 (50%)	440 (96%)	16 (4%)	2 (0%)	34 38
All	All	912/1832 (50%)	877 (96%)	31 (3%)	4 (0%)	34 38

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	314	PRO
1	B	314	PRO

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	712	SER
1	A	710	PRO

### 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	416/797 (52%)	388 (93%)	28 (7%)	16	17
1	B	419/797 (53%)	399 (95%)	20 (5%)	25	30
All	All	835/1594 (52%)	787 (94%)	48 (6%)	20	22

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	648	LYS
1	A	682	ARG
1	B	656	SER
1	A	656	SER
1	A	677	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	630	ASN
1	B	193	HIS
1	B	103	HIS
1	A	326	HIS
1	A	676	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 carbohydrates 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	466/916 (50%)	0.20	12 (2%) 56 65	57, 83, 124, 164	0
1	B	470/916 (51%)	0.21	19 (4%) 38 51	45, 68, 127, 160	0
All	All	936/1832 (51%)	0.21	31 (3%) 46 59	45, 78, 126, 164	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	673	ILE	6.1
1	B	669	ALA	5.2
1	B	665	SER	4.0
1	B	679	TRP	4.0
1	B	371	ASP	3.9

### 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 carbohydrates 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.