



# Full wwPDB X-ray Structure Validation Report ⓘ

May 18, 2020 – 12:09 am BST

PDB ID : 5M7S  
Title : Structure of human O-GlcNAc hydrolase with bound transition state analog ThiametG  
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 : 2.40 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 : 2.11  
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

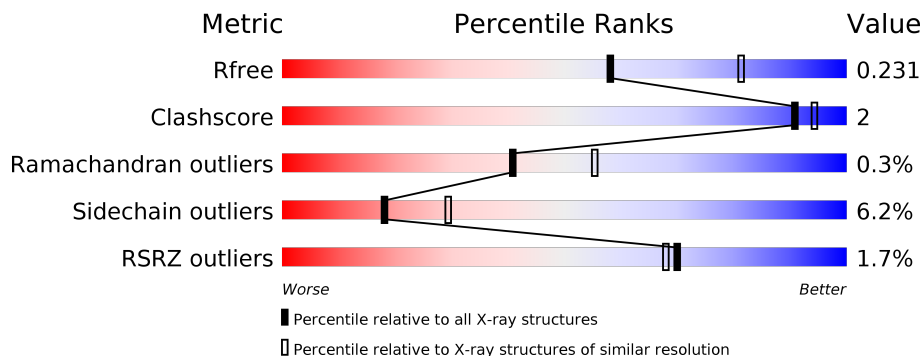
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	<div style="display: flex; align-items: center;"> <div style="width: 10px; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 100%; height: 15px; border: 1px solid black; position: relative;"> <div style="width: 43%; height: 100%; background-color: green;"></div> <div style="width: 5%; height: 100%; background-color: yellow;"></div> <div style="width: 52%; height: 100%; background-color: grey;"></div> </div> <div style="margin-left: 5px;">%</div> </div>
1	B	916	<div style="display: flex; align-items: center;"> <div style="width: 10px; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 100%; height: 15px; border: 1px solid black; position: relative;"> <div style="width: 44%; height: 100%; background-color: green;"></div> <div style="width: 5%; height: 100%; background-color: yellow;"></div> <div style="width: 52%; height: 100%; background-color: grey;"></div> </div> <div style="margin-left: 5px;">%</div> </div>

## 2 Entry composition [i](#)

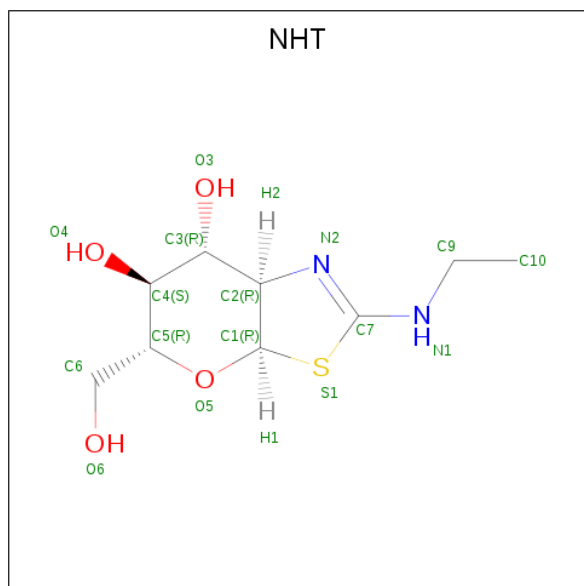
There are 3 unique types of molecules in this entry. The entry contains 7318 atoms, of which 0 are hydrogens and 0 are deuteriums.

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	443	Total 3626	C 2351	N 595	O 656	S 24	0	0	0
1	B	444	Total 3634	C 2355	N 601	O 654	S 24	0	0	0

- Molecule 2 is (3AR,5R,6S,7R,7AR)-2-(ETHYLAMINO)-5-(HYDROXYMETHYL)-5,6,7,7A-TETRAHYDRO-3AH-PYRANO[3,2-D][1,3]THIAZOLE-6,7-DIOL (three-letter code: NHT) (formula: C<sub>9</sub>H<sub>16</sub>N<sub>2</sub>O<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	Total 16	C 9	N 2	O 4	S 1	0	0
2	B	1	Total 16	C 9	N 2	O 4	S 1	0	0

- Molecule 3 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>	<b>ZeroOcc</b>	<b>AltConf</b>
3	A	8	Total O 9 9	0	1
3	B	17	Total O 17 17	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.91Å 101.91Å 282.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	96.06 – 2.40 72.06 – 2.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (96.06-2.40) 100.0 (72.06-2.40)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.68 (at 2.40Å)	Xtrriage
Refinement program	REFMAC 5.8.0155, REFMAC 5.8.0155	Depositor
R, $R_{free}$	0.207 , 0.232 0.210 , 0.231	Depositor DCC
$R_{free}$ test set	1770 reflections (2.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	59.9	Xtrriage
Anisotropy	0.378	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 43.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7318	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	78.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.84% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NHT

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	0.66	0/3725	0.83	4/5046 (0.1%)
1	B	0.69	0/3733	0.82	3/5054 (0.1%)
All	All	0.67	0/7458	0.82	7/10100 (0.1%)

There are no bond length outliers.

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	290	ARG	NE-CZ-NH2	-7.46	116.57	120.30
1	B	290	ARG	NE-CZ-NH1	7.23	123.92	120.30
1	A	634	ARG	NE-CZ-NH2	-6.48	117.06	120.30
1	A	108	ARG	NE-CZ-NH1	6.32	123.46	120.30
1	B	104	ARG	NE-CZ-NH1	5.28	122.94	120.30
1	A	634	ARG	NE-CZ-NH1	5.26	122.93	120.30
1	A	105	MET	CG-SD-CE	5.11	108.38	100.20

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	3626	0	3553	16	0
1	B	3634	0	3569	12	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	16	0	16	1	0
2	B	16	0	16	1	0
3	A	9	0	0	0	0
3	B	17	0	0	0	0
All	All	7318	0	7154	24	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.

All (24) 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	1.25	1.12
1:A:381:LYS:O	1:A:385:THR:HG22	1.81	0.80
1:A:75:MET:CE	1:A:122:LEU:HD22	2.15	0.70
1:A:75:MET:HE1	1:A:122:LEU:CD2	2.14	0.68
1:A:543:ASN:HB2	1:B:182:ALA:HB3	1.83	0.61
1:A:227:ASN:HD21	1:A:230:GLN:HE21	1.49	0.59
1:A:182:ALA:HB3	1:B:543:ASN:HB2	1.86	0.57
1:A:684:GLY:CA	1:B:290:ARG:HD3	2.36	0.55
1:B:215:CYS:SG	2:B:1001:NHT:H11	2.48	0.54
1:B:227:ASN:HD21	1:B:230:GLN:HE21	1.57	0.52
1:B:581:GLU:OE2	1:B:610:ARG:NH1	2.39	0.51
1:A:171:LEU:HD12	1:A:199:THR:HA	1.95	0.49
1:B:59:ARG:NH2	1:B:132:GLU:OE1	2.46	0.49
1:B:97:PRO:HG2	1:B:100:ASP:HB2	1.94	0.47
1:A:296:TYR:CE2	1:A:299:ARG:HG2	2.51	0.45
1:B:320:ALA:HB2	1:B:640:MET:CE	2.46	0.45
1:A:219:TYR:CD1	2:A:1001:NHT:H9C1	2.53	0.43
1:A:97:PRO:HG2	1:A:100:ASP:HB2	2.00	0.42
1:A:304:ILE:HB	1:A:305:PRO:HD3	2.00	0.42
1:A:75:MET:HA	1:A:75:MET:HE3	2.00	0.42
1:A:650:ILE:HD13	1:B:686:ALA:HB1	2.01	0.42
1:B:581:GLU:OE1	1:B:610:ARG:HD2	2.20	0.41
1:B:320:ALA:HB2	1:B:640:MET:HE1	2.02	0.40
1:A:320:ALA:HB2	1:A:640:MET:CE	2.51	0.40

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	433/916 (47%)	417 (96%)	14 (3%)	2 (0%)	29	41
1	B	434/916 (47%)	421 (97%)	12 (3%)	1 (0%)	47	62
All	All	867/1832 (47%)	838 (97%)	26 (3%)	3 (0%)	41	55

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	678	PRO
1	B	314	PRO
1	A	314	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	393/797 (49%)	368 (94%)	25 (6%)	17	28
1	B	393/797 (49%)	369 (94%)	24 (6%)	18	30
All	All	786/1594 (49%)	737 (94%)	49 (6%)	18	29

All (49) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	59	ARG
1	A	61	LEU
1	A	101	TYR

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>
1	A	137	ILE
1	A	142	ASP
1	A	150	GLU
1	A	167	ARG
1	A	177	ASP
1	A	228	VAL
1	A	262	GLU
1	A	265	GLU
1	A	273	ARG
1	A	278	TRP
1	A	299	ARG
1	A	303	LEU
1	A	306	ARG
1	A	385	THR
1	A	558	LEU
1	A	585	LEU
1	A	608	ARG
1	A	613	LYS
1	A	622	MET
1	A	629	SER
1	A	634	ARG
1	A	693	LEU
1	B	83	ARG
1	B	101	TYR
1	B	114	GLU
1	B	122	LEU
1	B	124	SER
1	B	137	ILE
1	B	150	GLU
1	B	171	LEU
1	B	177	ASP
1	B	228	VAL
1	B	247	VAL
1	B	262	GLU
1	B	273	ARG
1	B	278	TRP
1	B	306	ARG
1	B	308	LYS
1	B	339	VAL
1	B	549	THR
1	B	580	ARG
1	B	585	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	602	GLU
1	B	608	ARG
1	B	629	SER
1	B	695	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	103	HIS
1	A	193	HIS
1	A	230	GLN
1	A	326	HIS
1	A	335	ASN
1	B	103	HIS
1	B	230	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](#)

2 ligands 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	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	NHT	A	1001	-	15,17,17	0.32	0	14,24,24	1.17	2 (14%)
2	NHT	B	1001	-	15,17,17	0.42	0	14,24,24	0.99	1 (7%)

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	NHT	A	1001	-	-	0/5/33/33	0/2/2/2
2	NHT	B	1001	-	-	3/5/33/33	0/2/2/2

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1001	NHT	N1-C7-N2	-3.17	120.12	124.26
2	B	1001	NHT	C1-O5-C5	2.70	117.56	112.58
2	A	1001	NHT	C1-O5-C5	2.58	117.33	112.58

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	1001	NHT	N2-C7-N1-C9
2	B	1001	NHT	S1-C7-N1-C9
2	B	1001	NHT	C10-C9-N1-C7

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1001	NHT	1	0
2	B	1001	NHT	1	0

## 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	443/916 (48%)	0.14	9 (2%) 65 63	48, 80, 128, 179	0
1	B	444/916 (48%)	0.10	6 (1%) 75 73	41, 67, 117, 174	0
All	All	887/1832 (48%)	0.12	15 (1%) 70 68	41, 74, 126, 179	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	695	ILE	4.0
1	B	681	PHE	3.5
1	A	225	TYR	3.0
1	B	590	SER	2.9
1	A	325	ILE	2.8
1	A	538	PHE	2.8
1	A	681	PHE	2.7
1	B	664	ARG	2.6
1	B	593	SER	2.6
1	B	679	TRP	2.4
1	B	587	ALA	2.4
1	A	622	MET	2.2
1	A	680	ALA	2.2
1	A	334	SER	2.2
1	A	653	MET	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 carbohydrates 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	NHT	A	1001	16/16	0.97	0.14	49,53,54,55	0
2	NHT	B	1001	16/16	0.97	0.15	42,47,50,53	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.