



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 20, 2017 – 01:51 PM EDT

PDB ID : 2CHN  
Title : Bacteroides thetaiotaomicron hexosaminidase with O-GlcNAcase activity-NAG-thiazoline complex  
Authors : Dennis, R.J.; Taylor, E.J.; Macauley, M.S.; Stubbs, K.A.; Turkenburg, J.P.; Hart, S.J.; Black, G.N.; Vocadlo, D.J.; Davies, G.J.  
Deposited on : 2006-03-15  
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

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20029077  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20029077

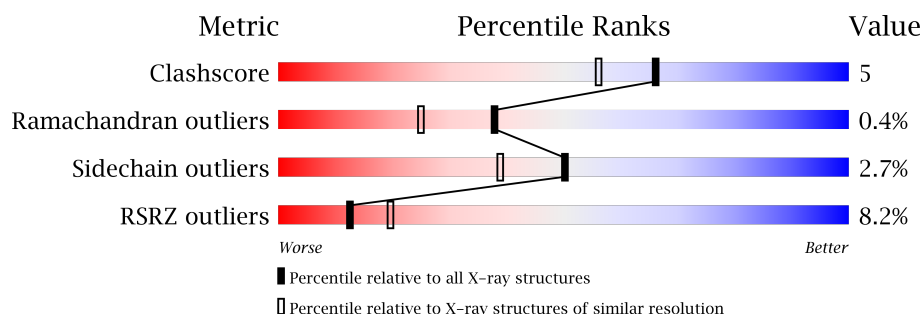
# 1 Overall quality at a glance

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))
Clashscore	112137	2136 (1.96-1.96)
Ramachandran outliers	110173	2117 (1.96-1.96)
Sidechain outliers	110143	2117 (1.96-1.96)
RSRZ outliers	101464	2018 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	716	<div> <div>6%</div> <div>79%</div> <div>10%</div> <div>10%</div> </div>
1	B	716	<div> <div>9%</div> <div>78%</div> <div>10%</div> <div>11%</div> </div>
2	C	12	<div> <div>92%</div> <div>8%</div> </div>
2	D	12	<div> <div>100%</div> </div>

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
5	GOL	A	1719	-	-	-	X
5	GOL	B	1718	-	-	-	X
5	GOL	D	1720	-	-	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 11738 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 GLUCOSAMINIDASE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	642	Total	C	N	O	S	Se	0	4	0
			5235	3361	880	976	2	16			
1	B	639	Total	C	N	O	S	Se	0	7	0
			5227	3358	876	973	2	18			

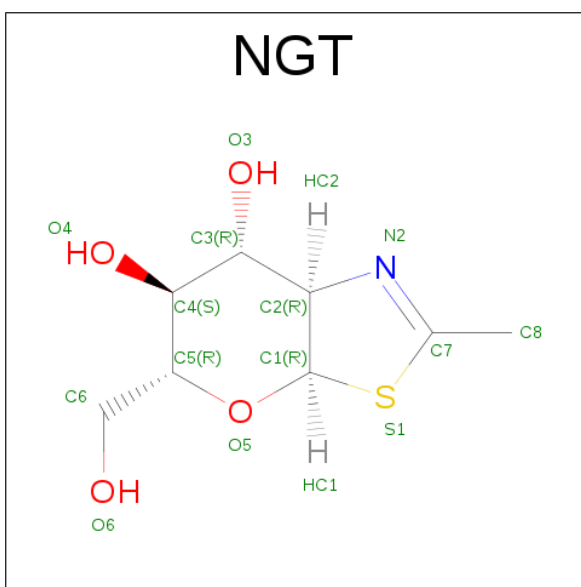
- Molecule 2 is a protein called GLUCOSAMINIDASE.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	12	Total	C	N	O	0	0	0
			60	36	12	12			
2	D	12	Total	C	N	O	0	0	0
			59	36	12	11			

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

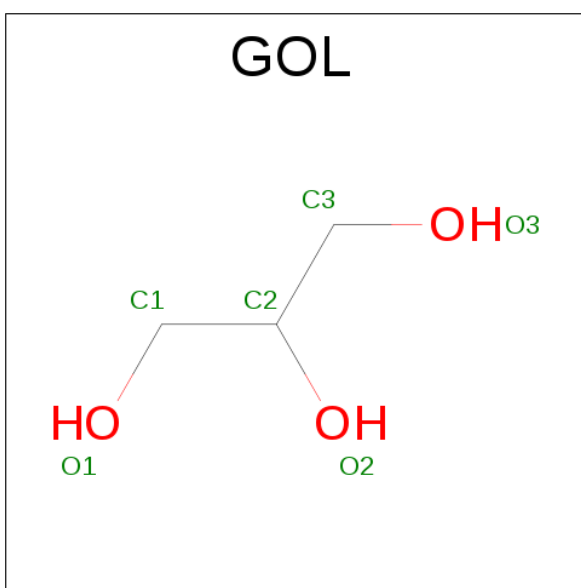
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Ca	0	0
			1	1		
3	A	1	Total	Ca	0	0
			1	1		

- Molecule 4 is 3AR,5R,6S,7R,7AR-5-HYDROXYMETHYL-2-METHYL-5,6,7,7A-TETRAHYDRO-3AH-PYRANO[3,2-D]THIAZOLE-6,7-DIOL (three-letter code: NGT) (formula: C<sub>8</sub>H<sub>13</sub>NO<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	S	0	0
			14	8	1	4	1		
4	B	1	Total	C	N	O	S	0	0
			14	8	1	4	1		

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			6	3	3		
5	B	1	Total	C	O	0	0
			6	3	3		

*Continued on next page...*

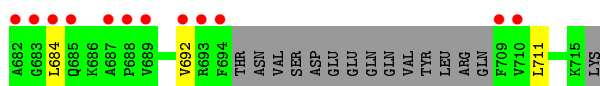
*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	C	1	Total	C	O	0	0
			6	3	3		
5	D	1	Total	C	O	0	0
			6	3	3		

- Molecule 6 is water.

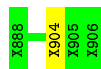
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	536	Total	O	0	0
			536	536		
6	B	557	Total	O	0	0
			557	557		
6	C	6	Total	O	0	0
			6	6		
6	D	4	Total	O	0	0
			4	4		





- Molecule 2: GLUCOSAMINIDASE

Chain C: 92% 8%



- Molecule 2: GLUCOSAMINIDASE

Chain D: 100%

There are no outlier residues recorded for this chain.

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	51.71Å 94.54Å 99.48Å 75.34° 93.86° 77.07°	Depositor
Resolution (Å)	39.84 – 1.95 39.83 – 1.95	Depositor EDS
% Data completeness (in resolution range)	96.7 (39.84-1.95) 88.1 (39.83-1.95)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.18 (at 1.95Å)	Xtriage
Refinement program	REFMAC	Depositor
R, $R_{free}$	0.185 , 0.226 0.190 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	25.2	Xtriage
Anisotropy	0.090	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 56.9	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	0.95	EDS
Total number of atoms	11738	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.53% 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: NGT, GOL, CA

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	0.78	1/5359 (0.0%)	0.75	6/7236 (0.1%)
1	B	0.80	1/5357 (0.0%)	0.77	5/7233 (0.1%)
All	All	0.79	2/10716 (0.0%)	0.76	11/14469 (0.1%)

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	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	345	TYR	CD1-CE1	-5.26	1.31	1.39
1	B	263	GLU	CB-CG	5.14	1.61	1.52

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	104	LEU	CA-CB-CG	-6.61	100.11	115.30
1	A	104	LEU	CA-CB-CG	-6.06	101.36	115.30
1	A	442	ASP	CB-CG-OD1	5.99	123.69	118.30
1	A	427	ASP	CB-CG-OD1	5.42	123.18	118.30
1	B	329	ARG	NE-CZ-NH1	5.32	122.96	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	614	VAL	Peptide

## 5.2 Too-close contacts ⓘ

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	5235	0	5152	45	0
1	B	5227	0	5148	53	0
2	C	60	0	16	1	0
2	D	59	0	16	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	14	0	13	0	0
4	B	14	0	13	0	0
5	A	6	0	8	0	0
5	B	6	0	8	0	0
5	C	6	0	8	1	0
5	D	6	0	8	2	0
6	A	536	0	0	7	0
6	B	557	0	0	12	0
6	C	6	0	0	0	0
6	D	4	0	0	0	0
All	All	11738	0	10390	96	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 96 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:459:ASN:H	1:B:459:ASN:HD22	1.05	1.00
1:B:555:LYS:HE3	6:B:2521:HOH:O	1.81	0.78
1:A:540:TYR:O	1:A:544:THR:HG23	1.84	0.77
1:B:459:ASN:N	1:B:459:ASN:HD22	1.83	0.77
1:A:22:PRO:HB3	1:A:55:MSE:HE3	1.68	0.75

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	634/716 (88%)	612 (96%)	21 (3%)	1 (0%)	51	41
1	B	634/716 (88%)	611 (96%)	19 (3%)	4 (1%)	28	15
All	All	1268/1432 (88%)	1223 (96%)	40 (3%)	5 (0%)	38	25

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	518	GLY
1	B	518	GLY
1	B	519	ARG
1	B	455	LYS
1	B	453	ALA

### 5.3.2 Protein sidechains

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	568/614 (92%)	552 (97%)	16 (3%)	49	37
1	B	568/614 (92%)	554 (98%)	14 (2%)	53	42
All	All	1136/1228 (92%)	1106 (97%)	30 (3%)	50	40

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	614	VAL

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	12	LEU
1	B	642	GLU
1	A	715	LYS
1	B	21	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	529	ASN
1	A	647	ASN
1	B	529	ASN
1	A	349	HIS
1	B	459	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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	NGT	A	1718	-	13,15,15	3.51	1 (7%)	11,22,22	1.60	1 (9%)
5	GOL	A	1719	-	5,5,5	0.46	0	5,5,5	0.54	0
4	NGT	B	1717	-	13,15,15	2.81	1 (7%)	11,22,22	1.50	1 (9%)
5	GOL	B	1718	-	5,5,5	0.42	0	5,5,5	0.57	0
5	GOL	C	1720	-	5,5,5	0.40	0	5,5,5	0.73	0
5	GOL	D	1720	-	5,5,5	0.30	0	5,5,5	1.29	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
4	NGT	A	1718	-	-	0/2/30/30	0/2/2/2
5	GOL	A	1719	-	-	0/4/4/4	0/0/0/0
4	NGT	B	1717	-	-	0/2/30/30	0/2/2/2
5	GOL	B	1718	-	-	0/4/4/4	0/0/0/0
5	GOL	C	1720	-	-	0/4/4/4	0/0/0/0
5	GOL	D	1720	-	-	0/4/4/4	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1718	NGT	C7-S1	-12.39	1.66	1.77
4	B	1717	NGT	C7-S1	-10.01	1.68	1.77

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	1717	NGT	C8-C7-S1	3.81	124.69	118.96
4	A	1718	NGT	C8-C7-S1	3.85	124.76	118.96

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	C	1720	GOL	1	0
5	D	1720	GOL	2	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	D	1
2	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	891:UNK	C	899:UNK	N	5.34
1	D	891:UNK	C	899:UNK	N	5.17

## 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	626/716 (87%)	0.22	41 (6%) 20 29	12, 26, 63, 80	0
1	B	623/716 (87%)	0.32	61 (9%) 8 13	14, 27, 63, 80	0
2	C	0/12	-	-	-	-
2	D	0/12	-	-	-	-
All	All	1249/1456 (85%)	0.27	102 (8%) 12 20	12, 27, 63, 80	0

The worst 5 of 102 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	648	PHE	7.1
1	B	605	LEU	6.7
1	A	694	PHE	6.7
1	B	694	PHE	5.9
1	B	646	ILE	5.7

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	GOL	B	1718	6/6	0.88	0.15	5.24	42,48,50,51	0
5	GOL	A	1719	6/6	0.92	0.16	4.39	32,42,44,47	0
5	GOL	D	1720	6/6	0.91	0.16	3.84	42,44,45,45	0
4	NGT	B	1717	14/14	0.98	0.13	0.93	17,19,20,21	0
4	NGT	A	1718	14/14	0.98	0.12	0.31	15,17,19,20	0
5	GOL	C	1720	6/6	0.94	0.10	-0.25	30,33,34,35	0
3	CA	B	1716	1/1	0.99	0.09	-	44,44,44,44	0
3	CA	A	1717	1/1	0.98	0.07	-	42,42,42,42	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.