



# wwPDB X-ray Structure Validation Summary Report

Apr 3, 2014 – 12:21 PM EDT

PDB ID : 4C7F  
Title : Structure and activity of the GH20 beta-N-acetylhexosaminidase from *Streptomyces coelicolor* A3(2)  
Authors : Nguyen Thi, N.; Offen, W.A.; Davies, G.J.; Doucet, N.  
Deposited on : 2013-09-20  
Resolution : 2.00 Å (reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

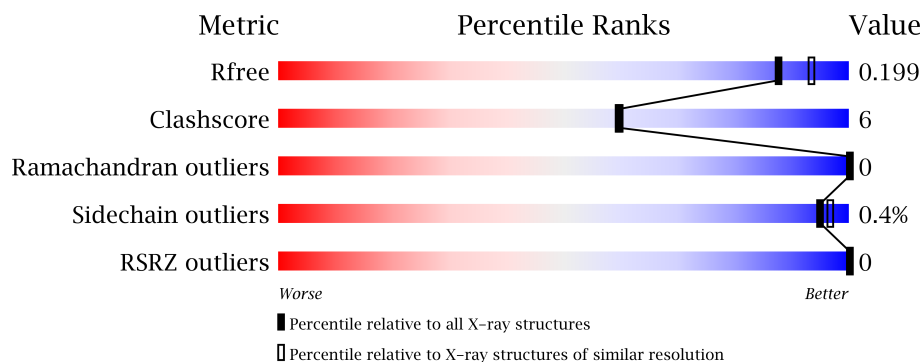
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	<b>FAILED</b>
Xtriage (Phenix)	:	dev-1439
EDS	:	stable22978
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22978

# 1 Overall quality at a glance

The reported resolution of this entry is 2.00 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))
$R_{free}$	66092	4888 (2.00-2.00)
Clashscore	79885	6188 (2.00-2.00)
Ramachandran outliers	78287	6102 (2.00-2.00)
Sidechain outliers	78261	6100 (2.00-2.00)
RSRZ outliers	66119	4890 (2.00-2.00)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	494	<div><div></div><div></div></div>
1	B	494	<div><div></div><div></div></div>

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	EDO	A	1498	-	X
2	EDO	A	1500	-	X
2	EDO	A	1501	-	X
2	EDO	A	1502	-	X
2	EDO	A	1503	-	X
2	EDO	A	1504[A]	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Geometry	Electron density
2	EDO	A	1504[B]	-	X
2	EDO	A	1505	-	X
2	EDO	A	1506	-	X
2	EDO	A	1507	-	X
2	EDO	A	1510	-	X
2	EDO	A	1512	-	X
2	EDO	B	1497	-	X
2	EDO	B	1498	-	X
2	EDO	B	1499	-	X
2	EDO	B	1500	-	X
2	EDO	B	1501	-	X
2	EDO	B	1504	-	X
2	EDO	B	1505	-	X
2	EDO	B	1506	-	X
2	EDO	B	1507	-	X

## 2 Entry composition i

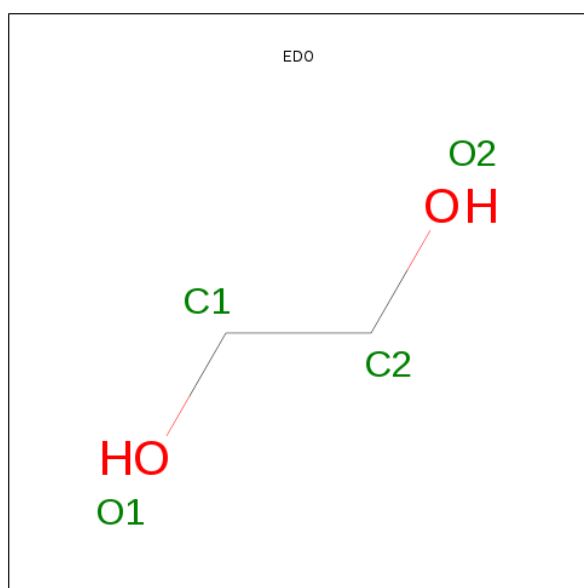
There are 4 unique types of molecules in this entry. The entry contains 9165 atoms, of which 0 are hydrogen and 0 are deuterium.

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 BETA-N-ACETYLHEXOSAMINIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	494	Total	C	N	O	S	0	9	0
			3901	2479	690	724	8			
1	B	493	Total	C	N	O	S	0	14	0
			3946	2497	702	739	8			

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			4	2	2		
2	A	1	Total	C	O	0	0
			4	2	2		
2	A	1	Total	C	O	0	0
			4	2	2		
2	A	1	Total	C	O	0	0
			4	2	2		

*Continued on next page...*

*Continued from previous page...*

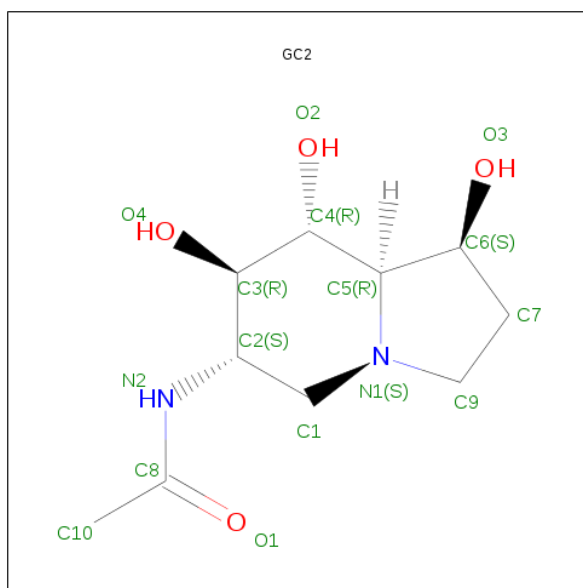
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 8	C 4	O 4	0	1
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0
2	B	1	Total 8	C 4	O 4	0	1
2	B	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0
2	B	1	Total 4	C 2	O 2	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total	C	O	0	0
			4	2	2		
2	B	1	Total	C	O	0	0
			4	2	2		
2	B	1	Total	C	O	0	0
			4	2	2		
2	B	1	Total	C	O	0	0
			4	2	2		
2	B	1	Total	C	O	0	0
			4	2	2		
2	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 3 is 6-ACETAMIDO-6-DEOXY-CASTANOSPERMINE (three-letter code: GC2) (formula:  $C_{10}H_{18}N_2O_4$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			16	10	2	4		
3	B	1	Total	C	N	O	0	0
			16	10	2	4		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	554	Total 557	O 557	0	3
4	B	593	Total 593	O 593	0	0

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

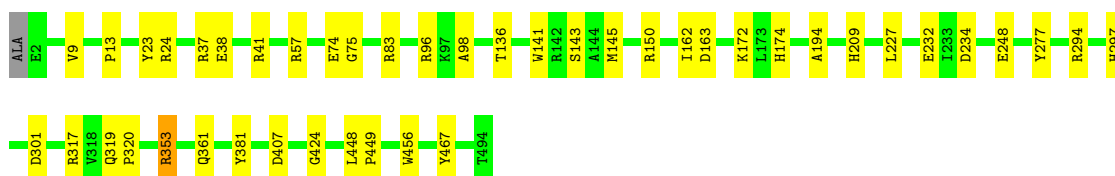
- Molecule 1: BETA-N-ACETYLHEXOSAMINIDASE

Chain A: 



- Molecule 1: BETA-N-ACETYLHEXOSAMINIDASE

Chain B: 





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.79Å 129.24Å 149.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	97.89 – 2.00 48.42 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.2 (97.89-2.00) 99.3 (48.42-2.00)	Depositor EDS
$R_{merge}$	0.18	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.75 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0033	Depositor
R, $R_{free}$	0.156 , 0.197 0.157 , 0.199	Depositor DCC
$R_{free}$ test set	4806 reflections (5.30%)	DCC
Wilson B-factor (Å <sup>2</sup> )	12.0	Xtriage
Anisotropy	0.454	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 35.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	1 of 95537 reflections (0.001%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9165	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	12.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.31% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: GC2, EDO

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	0.98	0/4041	0.94	12/5515 (0.2%)
1	B	1.00	4/4076 (0.1%)	0.90	5/5563 (0.1%)
All	All	0.99	4/8117 (0.0%)	0.92	17/11078 (0.2%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	248	GLU	CD-OE1	7.43	1.33	1.25
1	B	381	TYR	CE1-CZ	5.96	1.46	1.38
1	B	23	TYR	CE1-CZ	5.41	1.45	1.38
1	B	467	TYR	CE1-CZ	5.04	1.45	1.38

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	8	ARG	NE-CZ-NH2	-8.06	116.27	120.30
1	A	7	ASP	CB-CG-OD1	7.55	125.09	118.30
1	A	367	ARG	NE-CZ-NH2	-7.38	116.61	120.30
1	A	163	ASP	CB-CG-OD1	6.97	124.57	118.30
1	A	163	ASP	CB-CG-OD2	-6.80	112.17	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit,

and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3901	0	3805	43	0
1	B	3946	0	3813	46	0
2	A	72	0	108	13	0
2	B	64	0	96	8	0
3	A	16	0	18	0	0
3	B	16	0	18	0	0
4	A	557	0	0	23	0
4	B	593	0	0	5	0
All	All	9165	0	7858	96	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 6.

The worst 5 of 96 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:317[B]:ARG:HH12	2:B:1496[B]:EDO:C1	1.45	1.29
1:A:77[A]:TYR:CE2	4:A:2122:HOH:O	1.88	1.25
1:B:317[B]:ARG:NH1	2:B:1496[B]:EDO:H12	1.55	1.17
2:A:1507:EDO:O1	4:A:2209:HOH:O	1.71	1.06
1:B:96[B]:ARG:HH11	1:B:96[B]:ARG:CG	1.69	1.04

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	502/494 (102%)	488 (97%)	14 (3%)	0	100	100
1	B	506/494 (102%)	498 (98%)	8 (2%)	0	100	100
All	All	1008/988 (102%)	986 (98%)	22 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	398/388 (103%)	397 (100%)	1 (0%)	96	97
1	B	402/388 (104%)	398 (99%)	4 (1%)	85	88
All	All	800/776 (103%)	795 (99%)	5 (1%)	95	94

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	401	GLU
1	B	150	ARG
1	B	353[A]	ARG
1	B	353[B]	ARG
1	B	353[C]	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.5 Carbohydrates ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.6 Ligand geometry ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.7 Other polymers ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

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	494/494 (100%)	-0.80	0 100 100	5, 10, 20, 43	11 (2%)
1	B	493/494 (99%)	-0.84	0 100 100	4, 9, 17, 41	14 (2%)
All	All	987/988 (99%)	-0.82	0 100 100	4, 9, 19, 43	25 (2%)

There are no RSRZ outliers to report.

### 6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates

There are no carbohydrates in this entry.

### 6.4 Ligands

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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	EDO	A	1506	4/4	0.21	24.92	10,11,11,13	4
2	EDO	B	1507	4/4	0.23	24.43	13,13,15,17	4
2	EDO	B	1505	4/4	0.28	18.91	34,36,39,39	0
2	EDO	B	1504	4/4	0.14	15.85	21,24,25,25	1
2	EDO	A	1502	4/4	0.21	15.80	21,22,22,23	4

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	EDO	A	1501	4/4	0.25	14.74	31,36,39,40	0
2	EDO	B	1506	4/4	0.25	14.30	20,21,22,23	4
2	EDO	A	1500	4/4	0.15	13.89	24,26,27,28	0
2	EDO	A	1507	4/4	0.17	7.29	10,10,11,13	4
2	EDO	B	1498	4/4	0.14	7.22	28,30,32,37	0
2	EDO	A	1510	4/4	0.13	6.15	16,18,20,24	0
2	EDO	A	1504[A]	4/4	0.16	5.38	20,22,23,23	4
2	EDO	A	1504[B]	4/4	0.16	5.38	26,27,28,29	4
2	EDO	A	1503	4/4	0.16	4.98	29,32,34,36	0
2	EDO	A	1505	4/4	0.16	4.62	16,17,17,18	4
2	EDO	B	1499	4/4	0.10	4.59	18,19,20,21	0
2	EDO	B	1501	4/4	0.11	4.02	38,39,39,41	0
2	EDO	B	1497	4/4	0.20	3.56	29,30,30,31	0
2	EDO	B	1500	4/4	0.14	3.49	23,30,34,35	0
2	EDO	A	1498	4/4	0.13	2.64	19,21,26,35	0
2	EDO	A	1512	4/4	0.10	2.19	26,29,30,35	0
2	EDO	B	1496[B]	4/4	0.16	1.58	24,25,28,29	4
2	EDO	B	1496[A]	4/4	0.16	1.58	17,17,19,21	4
2	EDO	A	1497	4/4	0.12	1.55	29,31,31,36	0
2	EDO	B	1502	4/4	0.10	1.35	16,19,24,28	0
2	EDO	A	1496	4/4	0.18	1.10	32,38,38,42	0
3	GC2	B	1510	16/16	0.08	0.94	5,7,8,8	0
2	EDO	A	1508	4/4	0.07	0.72	20,20,21,22	0
2	EDO	B	1503	4/4	0.15	0.64	27,30,36,44	0
2	EDO	B	1509	4/4	0.15	0.60	40,40,42,43	0
2	EDO	A	1509	4/4	0.11	0.59	12,12,13,13	4
2	EDO	A	1499	4/4	0.07	0.09	18,19,19,20	0
2	EDO	B	1495	4/4	0.07	-0.25	17,17,18,18	0
3	GC2	A	1511	16/16	0.06	-0.35	7,9,11,11	0
2	EDO	B	1508	4/4	0.05	-0.84	8,8,9,10	0
2	EDO	A	1495	4/4	0.06	-0.94	9,10,10,11	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.