



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 19, 2020 – 02:38 PM EST

PDB ID : 6VE1  
Title : Crystal structure of endo-beta-N-acetylglucosaminidase H at high pH  
Authors : Stachowski, T.R.; Snell, M.E.; Snell, E.S.  
Deposited on : 2019-12-28  
Resolution : 2.10 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	:	2.14.6
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)	:	2.14.6

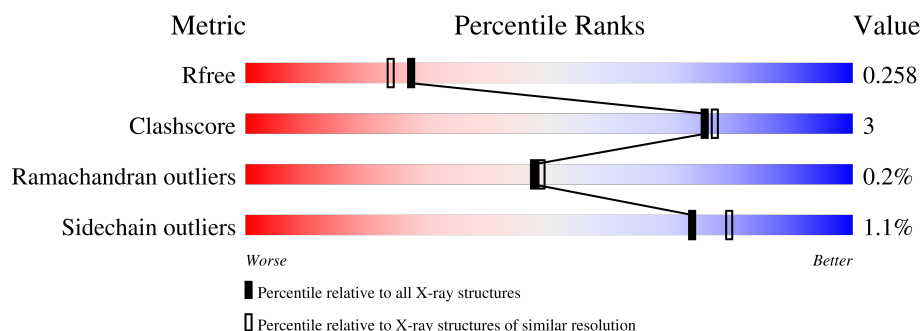
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*





The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)

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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	281	 88% 6% 5%
1	B	281	 90% 5% 5%
1	C	281	 88% 7% 5%
1	D	281	 86% 8% 5%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 17020 atoms, of which 7767 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 Endo-beta-N-acetylglucosaminidase H.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	267	Total	C	H	N	O	S	0	0	0
			3969	1274	1942	349	402	2			
1	B	267	Total	C	H	N	O	S	0	0	0
			3969	1274	1942	349	402	2			
1	C	268	Total	C	H	N	O	S	0	0	0
			3971	1276	1940	349	403	3			
1	D	267	Total	C	H	N	O	S	0	0	0
			3969	1274	1943	348	401	3			

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	expression tag	UNP P04067
A	2	LEU	-	expression tag	UNP P04067
A	3	SER	-	expression tag	UNP P04067
A	4	THR	-	expression tag	UNP P04067
A	5	GLY	-	expression tag	UNP P04067
A	6	CYS	-	expression tag	UNP P04067
A	7	TYR	-	expression tag	UNP P04067
A	8	MET	-	expression tag	UNP P04067
A	276	HIS	-	expression tag	UNP P04067
A	277	HIS	-	expression tag	UNP P04067
A	278	HIS	-	expression tag	UNP P04067
A	279	HIS	-	expression tag	UNP P04067
A	280	HIS	-	expression tag	UNP P04067
A	281	HIS	-	expression tag	UNP P04067
B	1	SER	-	expression tag	UNP P04067
B	2	LEU	-	expression tag	UNP P04067
B	3	SER	-	expression tag	UNP P04067
B	4	THR	-	expression tag	UNP P04067
B	5	GLY	-	expression tag	UNP P04067
B	6	CYS	-	expression tag	UNP P04067
B	7	TYR	-	expression tag	UNP P04067

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	8	MET	-	expression tag	UNP P04067
B	276	HIS	-	expression tag	UNP P04067
B	277	HIS	-	expression tag	UNP P04067
B	278	HIS	-	expression tag	UNP P04067
B	279	HIS	-	expression tag	UNP P04067
B	280	HIS	-	expression tag	UNP P04067
B	281	HIS	-	expression tag	UNP P04067
C	1	SER	-	expression tag	UNP P04067
C	2	LEU	-	expression tag	UNP P04067
C	3	SER	-	expression tag	UNP P04067
C	4	THR	-	expression tag	UNP P04067
C	5	GLY	-	expression tag	UNP P04067
C	6	CYS	-	expression tag	UNP P04067
C	7	TYR	-	expression tag	UNP P04067
C	8	MET	-	expression tag	UNP P04067
C	276	HIS	-	expression tag	UNP P04067
C	277	HIS	-	expression tag	UNP P04067
C	278	HIS	-	expression tag	UNP P04067
C	279	HIS	-	expression tag	UNP P04067
C	280	HIS	-	expression tag	UNP P04067
C	281	HIS	-	expression tag	UNP P04067
D	1	SER	-	expression tag	UNP P04067
D	2	LEU	-	expression tag	UNP P04067
D	3	SER	-	expression tag	UNP P04067
D	4	THR	-	expression tag	UNP P04067
D	5	GLY	-	expression tag	UNP P04067
D	6	CYS	-	expression tag	UNP P04067
D	7	TYR	-	expression tag	UNP P04067
D	8	MET	-	expression tag	UNP P04067
D	276	HIS	-	expression tag	UNP P04067
D	277	HIS	-	expression tag	UNP P04067
D	278	HIS	-	expression tag	UNP P04067
D	279	HIS	-	expression tag	UNP P04067
D	280	HIS	-	expression tag	UNP P04067
D	281	HIS	-	expression tag	UNP P04067

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Mg 1 1	0	0
2	A	2	Total Mg 2 2	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	1	Total 1	Mg 1	0	0
2	C	2	Total 2	Mg 2	0	0


- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	288	Total 288	O 288	0	0
3	B	300	Total 300	O 300	0	0
3	C	285	Total 285	O 285	0	0
3	D	263	Total 263	O 263	0	0

### 3 Residue-property plots

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. 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. 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.

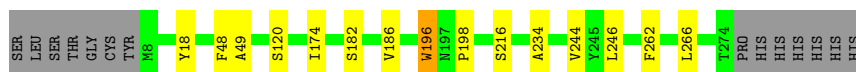
- Molecule 1: Endo-beta-N-acetylglucosaminidase H

Chain A: 




- Molecule 1: Endo-beta-N-acetylglucosaminidase H

Chain B: 



- Molecule 1: Endo-beta-N-acetylglucosaminidase H

Chain C: 



- Molecule 1: Endo-beta-N-acetylglucosaminidase H

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.48Å 99.48Å 135.12Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.03 – 2.10 76.88 – 1.82	Depositor EDS
% Data completeness (in resolution range)	99.7 (41.03-2.10) 99.8 (76.88-1.82)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.57 (at 1.82Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, $R_{free}$	0.221 , 0.258 0.220 , 0.258	Depositor DCC
$R_{free}$ test set	5666 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	9.4	Xtriage
Anisotropy	1.160	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 42.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.82	EDS
Total number of atoms	17020	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 53.52 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 4.1666e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: MG

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	0.44	0/2069	0.54	0/2820
1	B	0.44	0/2069	0.54	0/2820
1	C	0.43	0/2073	0.52	0/2826
1	D	0.45	0/2068	0.55	0/2818
All	All	0.44	0/8279	0.54	0/11284

There are no bond length outliers.

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	2027	1942	1941	12	0
1	B	2027	1942	1942	11	0
1	C	2031	1940	1940	12	0
1	D	2026	1943	1943	14	0
2	A	2	0	0	0	0
2	B	1	0	0	0	0
2	C	2	0	0	0	0
2	D	1	0	0	0	0
3	A	288	0	0	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	300	0	0	0	0
3	C	285	0	0	2	0
3	D	263	0	0	2	0
All	All	9253	7767	7766	49	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 3.

The worst 5 of 49 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:234:ALA:HB1	1:B:266:LEU:CD1	2.15	0.75
1:B:234:ALA:HB1	1:B:266:LEU:HD13	1.69	0.73
1:D:236:ARG:O	1:D:239:ASP:O	2.19	0.60
1:B:262:PHE:HD1	1:B:266:LEU:HD22	1.66	0.59
1:C:8:MET:HG2	1:C:37:GLY:HA3	1.87	0.57

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	265/281 (94%)	253 (96%)	11 (4%)	1 (0%)	34	32
1	B	265/281 (94%)	258 (97%)	7 (3%)	0	100	100
1	C	266/281 (95%)	256 (96%)	10 (4%)	0	100	100
1	D	265/281 (94%)	252 (95%)	12 (4%)	1 (0%)	34	32
All	All	1061/1124 (94%)	1019 (96%)	40 (4%)	2 (0%)	47	49

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	55	ASP
1	A	12	GLY

### 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	209/223 (94%)	208 (100%)	1 (0%)	88	92
1	B	209/223 (94%)	207 (99%)	2 (1%)	76	82
1	C	209/223 (94%)	206 (99%)	3 (1%)	67	73
1	D	209/223 (94%)	206 (99%)	3 (1%)	67	73
All	All	836/892 (94%)	827 (99%)	9 (1%)	73	79

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	164	MET
1	D	196	TRP
1	D	8	MET
1	B	196	TRP
1	C	196	TRP

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

Mol	Chain	Res	Type
1	C	11	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](#)

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.