

Full wwPDB X-ray Structure Validation Report (i)

Jan 24, 2023 – 09:20 PM EST

PDB ID : 4HET

Title : Crystal structure of a putative glycoside hydrolase (BT3745) from Bacteroides

thetaiotaomicron VPI-5482 at 2.10 A resolution

Authors : Joint Center for Structural Genomics (JCSG)

Deposited on : 2012-10-04

Resolution : 2.10 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*A user guide is available at

https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

 $Mol Probity \quad : \quad 4.02b\text{--}467$

Mogul: 1.8.5 (274361), CSD as541be (2020)

Xtriage (Phenix) : 1.13 EDS : 2.31.2

Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)

 $Refmac \quad : \quad 5.8.0158$

CCP4 : 7.0.044 (Gargrove)

Ideal geometry (proteins) : Engh & Huber (2001) Ideal geometry (DNA, RNA) : Parkinson et al. (1996)

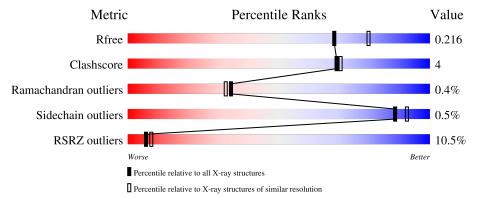
Validation Pipeline (wwPDB-VP) : 2.31.2

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 2.10 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 $(\# \mathrm{Entries})$	$\begin{array}{c} {\rm Similar\ resolution} \\ (\#{\rm Entries},{\rm resolution\ range}(\mathring{\rm A})) \end{array}$
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)
RSRZ outliers	127900	5083 (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 of the lower bar indicate the fraction of residues that contain outliers for >=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 <=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	360	90%	6% •
1	В	360	12% 89%	5% 6%



2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 5759 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.

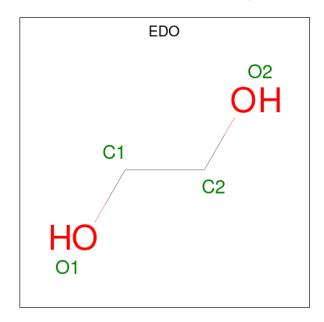
• Molecule 1 is a protein called HYPOTHETICAL GLYCOSIDE HYDROLASE.

Mol	Chain	Residues		Atoms			ZeroOcc	AltConf	Trace		
1	A	348	Total 2728	C 1739	N 437		S 3	Se 6	0	8	0
1	В	340	Total 2660	C 1697		O 529	S 3	Se 6	0	7	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled			Reference
A	0	GLY	-	expression tag	UNP Q8A1C1
В	0	GLY	-	expression tag	UNP Q8A1C1

• Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



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

Continued on next page...



Continued from previous page...

Mol	Chain	Residues	Atoms	5	ZeroOcc	AltConf
2	В	1	Total C 4 2	O 2	0	0

• Molecule 3 is water.

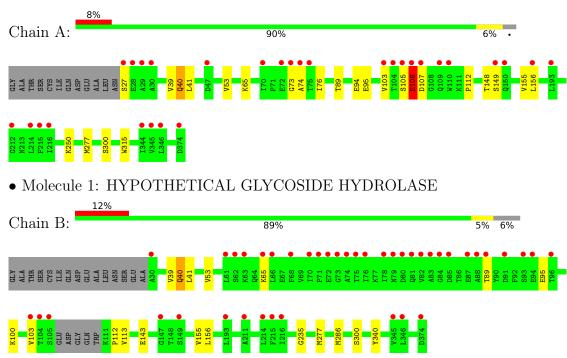
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	236	Total O 237 237	0	1
3	В	126	Total O 126 126	0	0



3 Residue-property plots (i)

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 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: HYPOTHETICAL GLYCOSIDE HYDROLASE





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 61	Depositor
Cell constants	180.38Å 180.38Å 63.75Å	Domositon
a, b, c, α , β , γ	90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.52 - 2.10	Depositor
Resolution (A)	29.52 - 2.10	EDS
% Data completeness	99.9 (29.52-2.10)	Depositor
(in resolution range)	100.0 (29.52-2.10)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) > 1$	2.35 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
D.D.	0.178 , 0.207	Depositor
R, R_{free}	0.188 , 0.216	DCC
R_{free} test set	3508 reflections $(5.06%)$	wwPDB-VP
Wilson B-factor (Å ²)	39.5	Xtriage
Anisotropy	0.250	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.31 , 53.8	EDS
L-test for twinning ²	$< L >=0.49, < L^2>=0.32$	Xtriage
Estimated twinning fraction	0.030 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	5759	wwPDB-VP
Average B, all atoms (Å ²)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.01% of the height of the origin peak. No significant pseudotranslation is detected.

²Theoretical values of <|L|>, $<L^2>$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



¹Intensities estimated from amplitudes.

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: 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
MIOI	Chain	RMSZ	# Z > 5	RMSZ	# Z > 5
1	A	0.59	0/2802	0.78	0/3784
1	В	0.54	0/2728	0.75	0/3682
All	All	0.56	0/5530	0.76	0/7466

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	2728	0	2655	21	0
1	В	2660	0	2597	18	0
2	A	4	0	6	0	0
2	В	4	0	6	0	0
3	A	237	0	0	1	0
3	В	126	0	0	1	0
All	All	5759	0	5264	39	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 4.

All (39) 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:40[B]:GLN:HE22	1:B:156:LEU:HA	1.45	0.78
1:A:148:THR:HG22	1:A:149:SER:N	2.11	0.66
1:B:65:LYS:HG2	1:B:89:THR:HG22	1.79	0.63
1:A:94:GLU:HG2	3:A:624:HOH:O	1.99	0.63
1:B:41:LEU:CD2	1:B:155:VAL:HG13	2.29	0.62
1:A:41:LEU:HD23	1:A:155:VAL:CG1	2.31	0.60
1:A:41:LEU:HD23	1:A:155:VAL:HG13	1.84	0.59
1:B:40[B]:GLN:NE2	1:B:156:LEU:HA	2.17	0.59
1:B:41:LEU:HD23	1:B:155:VAL:CG1	2.34	0.57
1:A:65:LYS:HG2	1:A:89:THR:HG22	1.87	0.57
1:B:41:LEU:HD23	1:B:155:VAL:HG13	1.86	0.57
1:A:40[B]:GLN:HE22	1:A:156:LEU:HD12	1.71	0.56
1:B:41:LEU:CD2	1:B:155:VAL:CG1	2.84	0.55
1:A:41:LEU:CD2	1:A:155:VAL:HG13	2.36	0.55
1:B:39:VAL:HG22	1:B:53:VAL:CG1	2.37	0.55
1:B:100:LYS:HE2	1:B:113:VAL:HG11	1.90	0.54
1:A:41:LEU:CD2	1:A:155:VAL:CG1	2.86	0.53
1:B:277:MSE:HE3	1:B:340:TYR:HB3	1.90	0.53
1:A:27:SER:N	1:A:107:ASP:HB2	2.26	0.50
1:B:277:MSE:HE3	1:B:340:TYR:CB	2.42	0.49
1:A:148:THR:HG22	1:A:149:SER:H	1.78	0.49
1:A:74:ALA:HB2	1:A:105:SER:O	2.14	0.48
1:B:41:LEU:HD22	1:B:155:VAL:HG13	1.96	0.48
1:B:65:LYS:CG	1:B:89:THR:HG22	2.42	0.48
1:A:76:ILE:HD13	1:A:103:VAL:HG22	1.97	0.47
1:B:277:MSE:HB3	1:B:300:SER:HB3	1.99	0.44
1:A:105:SER:O	1:A:106:GLU:CB	2.66	0.44
1:A:250:LYS:HE3	1:A:315:TRP:CE2	2.53	0.43
1:A:148:THR:CG2	1:A:149:SER:N	2.77	0.43
1:B:103:VAL:O	1:B:112:PRO:HD2	2.19	0.43
1:A:39:VAL:HG22	1:A:53:VAL:CG1	2.49	0.43
1:A:277:MSE:HB3	1:A:300:SER:HB3	2.01	0.42
1:A:65:LYS:CG	1:A:89:THR:HG22	2.49	0.42
1:A:73:GLY:C	1:A:106:GLU:HB3	2.40	0.41
1:B:39:VAL:HG22	1:B:53:VAL:HG12	2.01	0.41
1:A:103:VAL:O	1:A:112:PRO:HD2	2.20	0.41
1:B:143:GLU:HG2	3:B:501:HOH:O	2.21	0.41
1:B:235:GLY:O	1:B:286:MSE:HE3	2.21	0.41
1:A:155:VAL:HG12	1:A:156:LEU:N	2.37	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	Perce	ntiles
1	A	354/360~(98%)	347 (98%)	5 (1%)	2 (1%)	25	21
1	В	343/360 (95%)	337 (98%)	5 (2%)	1 (0%)	41	41
All	All	697/720 (97%)	684 (98%)	10 (1%)	3 (0%)	34	32

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	106	GLU
1	A	95	GLU
1	В	95	GLU

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	296/300 (99%)	293 (99%)	3 (1%)	76 82		
1	В	290/300 (97%)	288 (99%)	2 (1%)	84 88		
All	All	586/600 (98%)	581 (99%)	5 (1%)	88 84		

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

Mol	Chain	Res	Type
1	A	40[A]	GLN
1	A	40[B]	GLN
1	A	106	GLU

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type
1	В	40[A]	GLN
1	В	40[B]	GLN

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

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)

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 Lin		Link	B	Bond lengths			Bond angles			
WIOI	Moi Type Chain R	nes	Res Link	Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2	
2	EDO	A	401	-	3,3,3	0.43	0	2,2,2	0.26	0
2	EDO	В	400	-	3,3,3	0.44	0	2,2,2	0.24	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
2	EDO	A	401	-	-	0/1/1/1	-
2	EDO	В	400	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	В	400	EDO	O1-C1-C2-O2

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 (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^{th} 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		$OWAB(A^2)$	Q<0.9
1	A	343/360 (95%)	0.01	29 (8%) 10 13	38, 56, 105, 150	0
1	В	335/360~(93%)	0.48	42 (12%) 3 5	41, 66, 130, 168	0
All	All	678/720 (94%)	0.24	71 (10%) 6 8	38, 60, 121, 168	0

All (71) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	В	82	VAL	10.1
1	В	73	GLY	9.6
1	A	110	TRP	7.1
1	В	70	ILE	6.7
1	В	105	SER	6.7
1	В	83	ALA	6.4
1	В	84	GLY	6.1
1	В	76	ILE	6.0
1	A	29	ALA	5.6
1	В	71	PRO	5.3
1	В	75	THR	4.8
1	В	74	ALA	4.8
1	В	103	VAL	4.7
1	В	80	ASP	4.7
1	A	27	SER	4.4
1	В	104	THR	4.4
1	В	81	GLN	4.2
1	В	30	ALA	4.1
1	A	109	GLN	3.9
1	В	63	LYS	3.6
1	В	72	GLU	3.5
1	В	96	THR	3.5
1	В	62	SER	3.5
1	A	73	GLY	3.4

Continued on next page...



Continued from previous page...

Mol	nued fron Chain	Res	Type	RSRZ
1	В	214	LEU	3.3
1	В	88	ALA	3.2
1	В	374	ASP	3.2
1	В	85	ASP	3.1
1	A	346	LEU	3.0
1	В	215	PHE	2.9
1	В	87	GLU	2.9
1	A	70	ILE	2.8
1	A	74	ALA	2.8
1	В	216	ILE	2.8
1	A	193	LEU	2.8
1	A	75	THR	2.8
1	В	65	LYS	2.8
1	В	66	LEU	2.8
1	A	107	ASP	2.7
1	A	30	ALA	2.7
1	A	106	GLU	2.7
1	A	214	LEU	2.7
1	A	374	ASP	2.7
1	A	216	ILE	2.7
1	В	147	GLY	2.7
1	A	105	SER	2.7
1	A	345	VAL	2.6
1	В	149	SER	2.6
1	В	93	SER	2.6
1	В	68	PHE	2.6
1	A	104	THR	2.5
1	A	47	ASP	2.4
1	В	61	LEU	2.4
1	В	346	LEU	2.3
1	В	345	VAL	2.3
1	В	193	LEU	2.3
1	В	89	THR	2.3
1	В	91	ASP	2.3
1	A	344	ILE	2.2
1	A	150	GLN	2.2
1	В	211	ALA	2.2
1	В	78	ILE	2.2
1	A	215	PHE	2.2
1	A	28	GLU	2.1
1	В	79	ASN	2.1
1	A	156	LEU	2.1



Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	72	GLU	2.1
1	В	94	GLU	2.1
1	A	103	VAL	2.1
1	A	212	GLY	2.0
1	A	149	SER	2.0

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 monosaccharides 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^{th} 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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
2	EDO	В	400	4/4	0.84	0.21	69,86,88,94	0
2	EDO	A	401	4/4	0.90	0.19	73,78,87,105	0

6.5 Other polymers (i)

There are no such residues in this entry.

