W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 01:18 pm BST

PDB ID	:	5F7C
Title	:	Crystal structure of Family 31 alpha-glucosidase (BT_0339) from Bacteroides thetaiotaomicron
Authors	:	Chaudet, M.M.; Rose, D.R.
Deposited on	:	2015-12-07
Resolution	:	2.60 Å(reported)

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

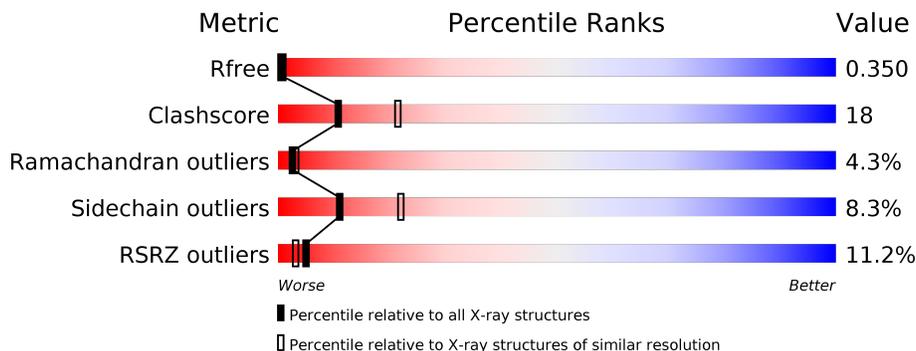|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1　Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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	Percentile Ranks	Value
Rfree		0.350
Clashscore		18
Ramachandran outliers		4.3%
Sidechain outliers		8.3%
RSRZ outliers		11.2%

Worse　　　　　　　　　　　　　　　　　*Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 >=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	761	5%　50%　26%　6% • 18%
1	B	761	16%　53%　22%　5%　20%
1	C	761	6%　57%　23% • 17%
1	D	761	9%　51%　26% • 18%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 34019 atoms, of which 14026 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 Alpha-glucosidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	625	Total	C	H	N	O	S	0	0	0
			8422	3229	3426	837	901	29			
1	B	611	Total	C	H	N	O	S	0	0	0
			8403	3150	3512	832	881	28			
1	C	632	Total	C	H	N	O	S	0	0	0
			8637	3278	3548	865	917	29			
1	D	623	Total	C	H	N	O	S	0	0	0
			8490	3185	3540	840	895	30			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	LEU	PHE	conflict	UNP Q8AAX3
A	13	PHE	LEU	conflict	UNP Q8AAX3
A	750	HIS	-	expression tag	UNP Q8AAX3
A	751	HIS	-	expression tag	UNP Q8AAX3
A	752	HIS	-	expression tag	UNP Q8AAX3
A	753	HIS	-	expression tag	UNP Q8AAX3
A	754	HIS	-	expression tag	UNP Q8AAX3
A	755	HIS	-	expression tag	UNP Q8AAX3
A	756	LEU	-	expression tag	UNP Q8AAX3
A	757	ARG	-	expression tag	UNP Q8AAX3
A	758	VAL	-	expression tag	UNP Q8AAX3
A	759	PRO	-	expression tag	UNP Q8AAX3
A	760	ARG	-	expression tag	UNP Q8AAX3
A	761	GLY	-	expression tag	UNP Q8AAX3
A	762	SER	-	expression tag	UNP Q8AAX3
B	12	LEU	PHE	conflict	UNP Q8AAX3
B	13	PHE	LEU	conflict	UNP Q8AAX3
B	750	HIS	-	expression tag	UNP Q8AAX3
B	751	HIS	-	expression tag	UNP Q8AAX3
B	752	HIS	-	expression tag	UNP Q8AAX3
B	753	HIS	-	expression tag	UNP Q8AAX3

Continued on next page...
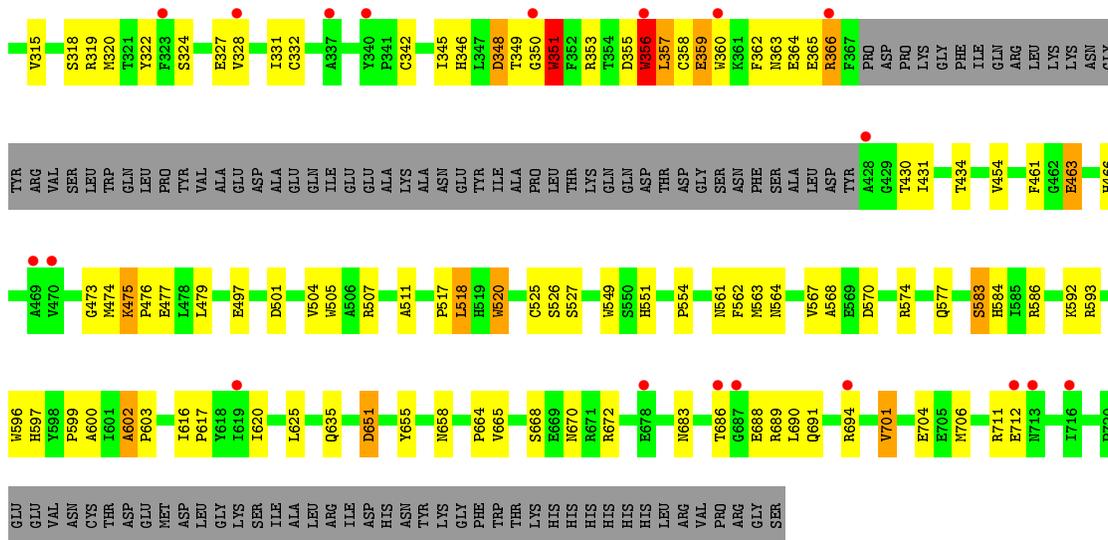
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	754	HIS	-	expression tag	UNP Q8AAX3
B	755	HIS	-	expression tag	UNP Q8AAX3
B	756	LEU	-	expression tag	UNP Q8AAX3
B	757	ARG	-	expression tag	UNP Q8AAX3
B	758	VAL	-	expression tag	UNP Q8AAX3
B	759	PRO	-	expression tag	UNP Q8AAX3
B	760	ARG	-	expression tag	UNP Q8AAX3
B	761	GLY	-	expression tag	UNP Q8AAX3
B	762	SER	-	expression tag	UNP Q8AAX3
C	12	LEU	PHE	conflict	UNP Q8AAX3
C	13	PHE	LEU	conflict	UNP Q8AAX3
C	750	HIS	-	expression tag	UNP Q8AAX3
C	751	HIS	-	expression tag	UNP Q8AAX3
C	752	HIS	-	expression tag	UNP Q8AAX3
C	753	HIS	-	expression tag	UNP Q8AAX3
C	754	HIS	-	expression tag	UNP Q8AAX3
C	755	HIS	-	expression tag	UNP Q8AAX3
C	756	LEU	-	expression tag	UNP Q8AAX3
C	757	ARG	-	expression tag	UNP Q8AAX3
C	758	VAL	-	expression tag	UNP Q8AAX3
C	759	PRO	-	expression tag	UNP Q8AAX3
C	760	ARG	-	expression tag	UNP Q8AAX3
C	761	GLY	-	expression tag	UNP Q8AAX3
C	762	SER	-	expression tag	UNP Q8AAX3
D	12	LEU	PHE	conflict	UNP Q8AAX3
D	13	PHE	LEU	conflict	UNP Q8AAX3
D	750	HIS	-	expression tag	UNP Q8AAX3
D	751	HIS	-	expression tag	UNP Q8AAX3
D	752	HIS	-	expression tag	UNP Q8AAX3
D	753	HIS	-	expression tag	UNP Q8AAX3
D	754	HIS	-	expression tag	UNP Q8AAX3
D	755	HIS	-	expression tag	UNP Q8AAX3
D	756	LEU	-	expression tag	UNP Q8AAX3
D	757	ARG	-	expression tag	UNP Q8AAX3
D	758	VAL	-	expression tag	UNP Q8AAX3
D	759	PRO	-	expression tag	UNP Q8AAX3
D	760	ARG	-	expression tag	UNP Q8AAX3
D	761	GLY	-	expression tag	UNP Q8AAX3
D	762	SER	-	expression tag	UNP Q8AAX3

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	23	Total 23	O 23	0	0
2	B	9	Total 9	O 9	0	0
2	C	18	Total 18	O 18	0	0
2	D	17	Total 17	O 17	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 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: Alpha-glucosidase

- Molecule 1: Alpha-glucosidase

- Molecule 1: Alpha-glucosidase

Chain C:

6%

57%   23%   17%

• Molecule 1: Alpha-glucosidase

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	82.34Å 75.00Å 232.86Å 90.00° 95.14° 90.00°	Depositor
Resolution (Å)	48.81 – 2.60 48.81 – 2.60	Depositor EDS
% Data completeness (in resolution range)	100.0 (48.81-2.60) 100.0 (48.81-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.57 (at 2.61Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.302 , 0.349 0.306 , 0.350	Depositor DCC
R_{free} test set	2000 reflections (2.29%)	wwPDB-VP
Wilson B-factor (Å^2)	46.0	Xtriage
Anisotropy	0.049	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 59.5	EDS
L-test for twinning[2]	$<|L|> = 0.45$, $<L^2> = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.84	EDS
Total number of atoms	34019	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.85% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical 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.

5 Model quality (i)

5.1 Standard geometry (i)

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	1/5147 (0.0%)	0.58	1/7000 (0.0%)
1	B	0.71	2/5035 (0.0%)	0.48	1/6839 (0.0%)
1	C	0.74	2/5240 (0.0%)	0.55	1/7115 (0.0%)
1	D	0.74	2/5092 (0.0%)	0.54	1/6917 (0.0%)
All	All	0.71	7/20514 (0.0%)	0.54	4/27871 (0.0%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	324	SER	CA-CB	-37.04	0.97	1.52
1	B	324	SER	CA-CB	-37.03	0.97	1.52
1	D	324	SER	CA-CB	-36.99	0.97	1.52
1	C	324	SER	CA-CB	-36.97	0.97	1.52
1	B	117	TRP	CA-CB	-25.68	0.97	1.53

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	476	PRO	N-CA-CB	5.63	110.06	103.30
1	D	476	PRO	N-CA-CB	5.57	109.99	103.30
1	A	476	PRO	N-CA-CB	5.31	109.67	103.30
1	C	476	PRO	N-CA-CB	5.24	109.58	103.30

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	4996	3426	4731	213	0
1	B	4891	3512	4668	166	0
1	C	5089	3548	4863	161	0
1	D	4950	3540	4704	182	0
2	A	23	0	0	0	0
2	B	9	0	0	0	0
2	C	18	0	0	0	0
2	D	17	0	0	2	0
All	All	19993	14026	18966	711	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 18.

The worst 5 of 711 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:D:49:LYS:HB2	1:D:64:ARG:HG2	1.26	1.12
1:A:14:ASP:HA	1:A:15:PHE:HB2	1.17	1.12
1:A:49:LYS:HB2	1:A:64:ARG:HD2	1.39	1.03
1:A:14:ASP:CA	1:A:15:PHE:HB2	1.94	0.98
1:A:37:LYS:HB2	1:A:39:GLY:H	1.26	0.96

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	619/761 (81%)	518 (84%)	71 (12%)	30 (5%)	2	2
1	B	605/761 (80%)	523 (86%)	59 (10%)	23 (4%)	3	4
1	C	626/761 (82%)	537 (86%)	66 (10%)	23 (4%)	3	4
1	D	617/761 (81%)	532 (86%)	55 (9%)	30 (5%)	2	2

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	2467/3044 (81%)	2110 (86%)	251 (10%)	106 (4%)	2 3

5 of 106 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	87	GLU
1	A	133	GLU
1	A	137	LEU
1	A	142	GLU
1	A	162	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	512/663 (77%)	460 (90%)	52 (10%)	7 14
1	B	504/663 (76%)	460 (91%)	44 (9%)	10 20
1	C	527/663 (80%)	495 (94%)	32 (6%)	18 38
1	D	507/663 (76%)	464 (92%)	43 (8%)	10 21
All	All	2050/2652 (77%)	1879 (92%)	171 (8%)	11 22

5 of 171 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	430	THR
1	B	713	ASN
1	D	563	MET
1	B	465	ILE
1	B	592	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	591	ASN
1	D	259	HIS
1	B	649	HIS
1	B	255	HIS
1	D	113	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)

There are no ligands in this entry.

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, 95th 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(Å²)	Q<0.9
1	A	625/761 (82%)	0.47	38 (6%)	21	16	22, 45, 78, 95	0
1	B	611/761 (80%)	1.21	125 (20%)	1	0	47, 64, 88, 96	0
1	C	632/761 (83%)	0.62	45 (7%)	16	11	25, 49, 83, 95	0
1	D	623/761 (81%)	0.82	71 (11%)	5	3	32, 53, 87, 101	0
All	All	2491/3044 (81%)	0.78	279 (11%)	5	3	22, 56, 85, 101	0

The worst 5 of 279 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	350	GLY	7.7
1	D	113	GLN	6.7
1	B	59	ASP	6.5
1	B	468	ASP	6.5
1	B	52	LEU	6.3

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.