



wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 06:40 pm BST

PDB ID : 2J7H
Title : Beta-glucosidase from *Thermotoga maritima* in complex with azafagomine
Authors : Gloster, T.M.; Meloncelli, P.; Stick, R.V.; Zechel, D.; Davies, G.J.
Deposited on : 2006-10-08
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

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

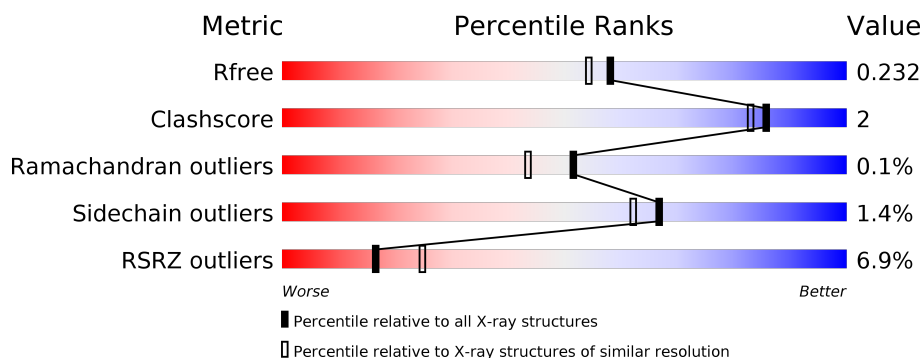
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 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))
R_{free}	130704	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (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 ≥ 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 $\leq 5\%$. 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	468	<div> <div>2%</div> <div> <div></div> <div>88%</div> <div>5%</div> <div>6%</div> </div> </div>
1	B	468	<div> <div>11%</div> <div> <div></div> <div>88%</div> <div>6%</div> <div>6%</div> </div> </div>

2 Entry composition [i](#)

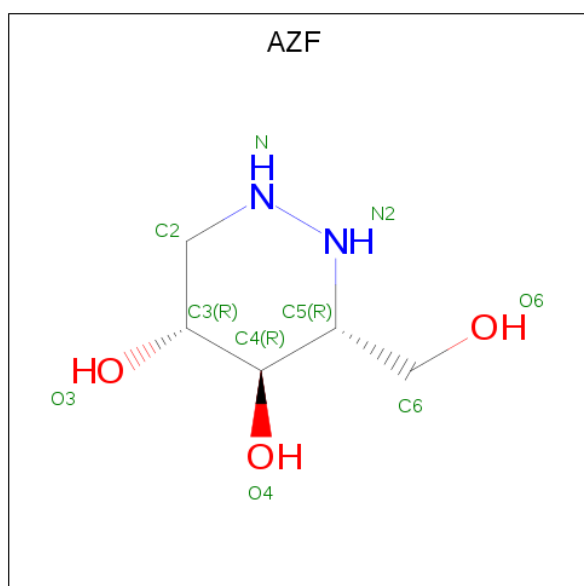
There are 5 unique types of molecules in this entry. The entry contains 7890 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 BETA-GLUCOSIDASE A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	438	Total	C	N	O	S	0	11	0
			3633	2368	614	644	7			
1	B	442	Total	C	N	O	S	0	9	0
			3639	2368	610	654	7			

- Molecule 2 is AZAFAGOMINE (three-letter code: AZF) (formula: $C_5H_{12}N_2O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			10	5	2	3		
2	B	1	Total	C	N	O	0	0
			10	5	2	3		

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

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

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	362	Total O 362 362	0	0
5	B	221	Total O 221 221	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	94.23Å 94.48Å 113.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.83 – 1.95 29.81 – 1.95	Depositor EDS
% Data completeness (in resolution range)	98.2 (29.83-1.95) 90.9 (29.81-1.95)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.63 (at 1.95Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.188 , 0.238 0.184 , 0.232	Depositor DCC
R_{free} test set	3399 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	32.0	Xtriage
Anisotropy	0.072	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 47.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.013 for k,h,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	7890	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.94% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: CA, AZF, ACT

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	0.74	0/3779	0.70	0/5128
1	B	0.64	0/3783	0.64	0/5146
All	All	0.70	0/7562	0.67	0/10274

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	3633	0	3526	17	0
1	B	3639	0	3480	14	0
2	A	10	0	12	1	0
2	B	10	0	12	0	0
3	A	1	0	0	0	0
3	B	2	0	0	0	0
4	A	8	0	6	0	0
4	B	4	0	3	0	0
5	A	362	0	0	7	0
5	B	221	0	0	2	0
All	All	7890	0	7039	31	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 2.

The worst 5 of 31 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:A:28:ASP:HB3	1:A:96:LYS:HD3	1.75	0.69
1:A:240[A]:ARG:HD3	5:A:2228:HOH:O	1.94	0.67
1:B:339[A]:LYS:HD2	1:B:391:PRO:HD3	1.81	0.63
1:A:188:TYR:O	1:A:192[B]:ARG:HG2	2.00	0.61
1:A:240[A]:ARG:CD	5:A:2228:HOH:O	2.47	0.61

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	444/468 (95%)	433 (98%)	11 (2%)	0	100	100
1	B	450/468 (96%)	436 (97%)	13 (3%)	1 (0%)	47	38
All	All	894/936 (96%)	869 (97%)	24 (3%)	1 (0%)	51	43

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	213	LYS

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	377/399 (94%)	372 (99%)	5 (1%)	69	65
1	B	375/399 (94%)	368 (98%)	7 (2%)	57	50
All	All	752/798 (94%)	740 (98%)	12 (2%)	67	58

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	246[A]	ASN
1	B	246[B]	ASN
1	B	359	VAL
1	A	246[B]	ASN
1	B	313	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	381	GLN
1	B	381	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](#)

Of 8 ligands modelled in this entry, 3 are monoatomic - leaving 5 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	ACT	B	1448	3	1,3,3	1.66	0	0,3,3	0.00	-
2	AZF	B	1445	-	7,10,10	0.75	0	8,13,13	1.46	2 (25%)
2	AZF	A	1446	-	7,10,10	1.13	1 (14%)	8,13,13	1.51	1 (12%)
4	ACT	A	1449	-	1,3,3	1.78	0	0,3,3	0.00	-
4	ACT	A	1448	3	1,3,3	1.25	0	0,3,3	0.00	-

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	AZF	A	1446	-	-	0/2/16/16	0/1/1/1
2	AZF	B	1445	-	-	0/2/16/16	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1446	AZF	C3-C4	2.29	1.55	1.52

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1446	AZF	C2-C3-C4	3.73	114.71	110.33
2	B	1445	AZF	O3-C3-C2	2.40	114.30	109.61
2	B	1445	AZF	C2-C3-C4	2.11	112.80	110.33

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1446	AZF	1	0

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(Å ²)	Q<0.9
1	A	438/468 (93%)	0.06	11 (2%) 57 66	18, 29, 48, 66	0
1	B	442/468 (94%)	0.60	50 (11%) 5 8	19, 39, 61, 77	0
All	All	880/936 (94%)	0.33	61 (6%) 16 25	18, 33, 59, 77	0

The worst 5 of 61 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	308	ALA	4.9
1	A	3	VAL	4.7
1	B	307	PRO	4.6
1	B	359	VAL	4.2
1	B	361	SER	4.1

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. The B-factors column lists the minimum, median, 95th 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	B-factors(\AA^2)	Q<0.9
4	ACT	A	1449	4/4	0.83	0.19	39,40,40,41	0
4	ACT	A	1448	4/4	0.88	0.17	46,46,47,47	0
2	AZF	A	1446	10/10	0.94	0.13	23,28,30,33	0
4	ACT	B	1448	4/4	0.95	0.17	47,48,48,48	0
3	CA	B	1447	1/1	0.95	0.06	60,60,60,60	0
2	AZF	B	1445	10/10	0.96	0.17	29,30,31,33	0
3	CA	B	1446	1/1	0.97	0.05	36,36,36,36	0
3	CA	A	1447	1/1	1.00	0.03	27,27,27,27	0

6.5 Other polymers [i](#)

There are no such residues in this entry.