wwPDB X-ray Structure Validation Summary Report (i)

Apr 16, 2025 – 01:13 pm BST

PDB ID	:	9EUZ / pdb_00009euz
Title	:	Glycoside hydrolase family 191 enzyme from Thermotoga maritima
Authors	:	Roth, C.
Deposited on	:	2024-03-28
Resolution	:	2.60 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 (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 (i)) were used in the production of this report:

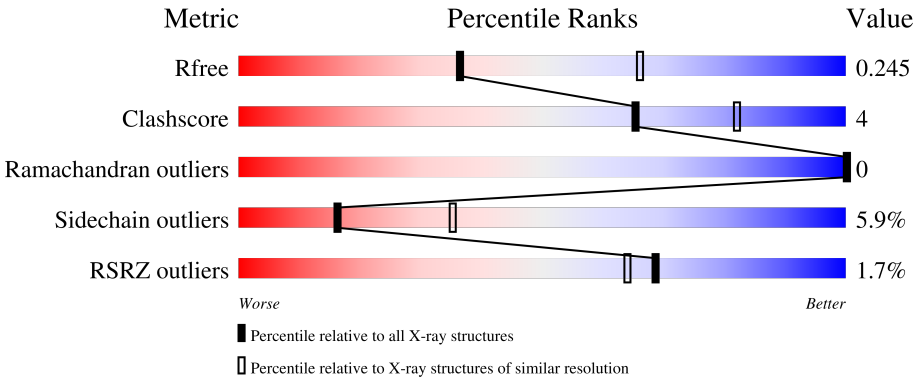
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.42

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 2.60 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	3775 (2.60-2.60)
Clashscore	180529	4181 (2.60-2.60)
Ramachandran outliers	177936	4129 (2.60-2.60)
Sidechain outliers	177891	4129 (2.60-2.60)
RSRZ outliers	164620	3775 (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 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	297	2% 77% 15%
1	B	297	78% 14% 5%
1	C	297	2% 79% 13% 5%
1	D	297	2% 79% 14% 5%
1	E	297	77% 13% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	297	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 27953 atoms, of which 13303 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 Uncharacterized protein TM_1410.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	284	Total	C	H	N	O	S	46	2	0
			4586	1519	2228	374	459	6			
1	B	283	Total	C	H	N	O	S	46	0	0
			4558	1510	2217	372	453	6			
1	C	281	Total	C	H	N	O	S	46	0	0
			4530	1500	2203	370	451	6			
1	D	283	Total	C	H	N	O	S	46	0	0
			4558	1510	2217	372	453	6			
1	E	282	Total	C	H	N	O	S	46	0	0
			4551	1508	2214	371	452	6			
1	F	283	Total	C	H	N	O	S	46	2	0
			4574	1515	2224	373	456	6			

- Molecule 2 is water.

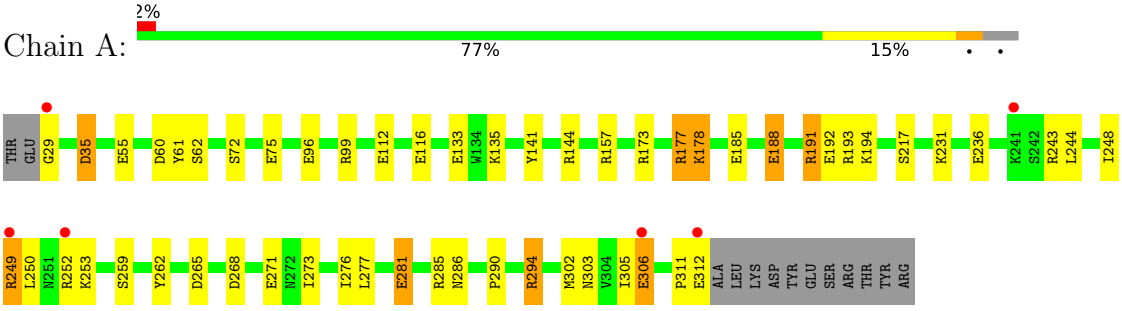
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	110	Total	O	0	0
			110	110		
2	B	107	Total	O	0	0
			107	107		
2	C	94	Total	O	0	0
			94	94		
2	D	106	Total	O	0	0
			106	106		
2	E	88	Total	O	0	0
			88	88		
2	F	91	Total	O	0	0
			91	91		

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.

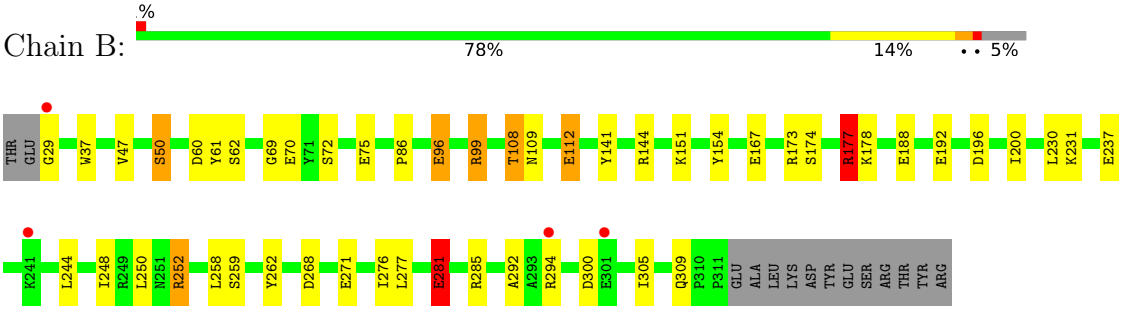
• Molecule 1: Uncharacterized protein TM_1410

Chain A:



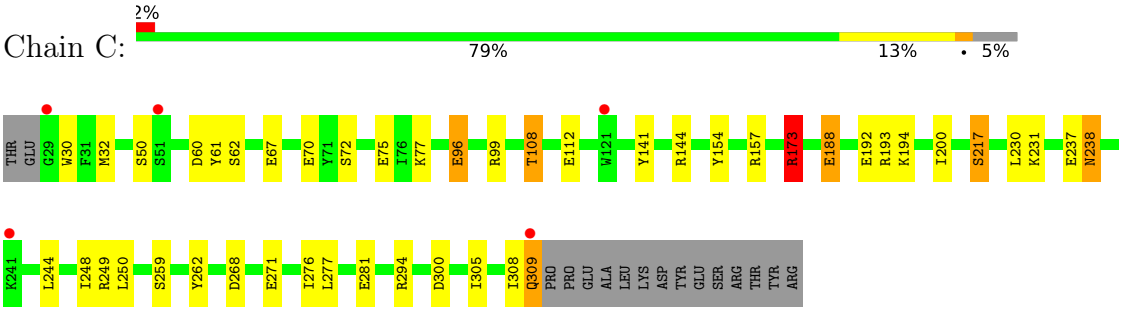
• Molecule 1: Uncharacterized protein TM_1410

Chain B:



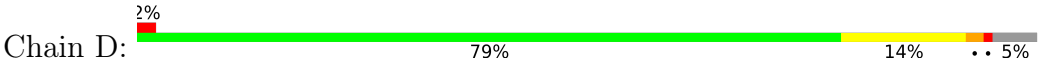
• Molecule 1: Uncharacterized protein TM_1410

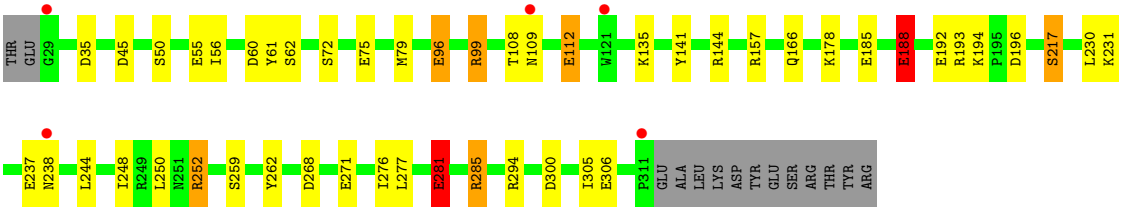
Chain C:



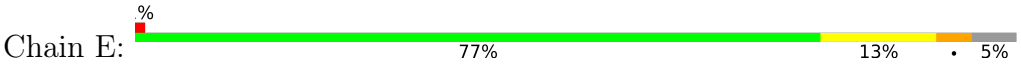
• Molecule 1: Uncharacterized protein TM_1410

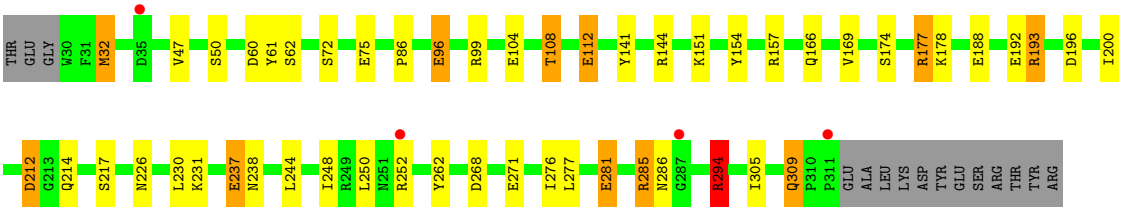
Chain D:



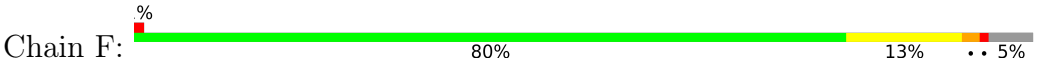


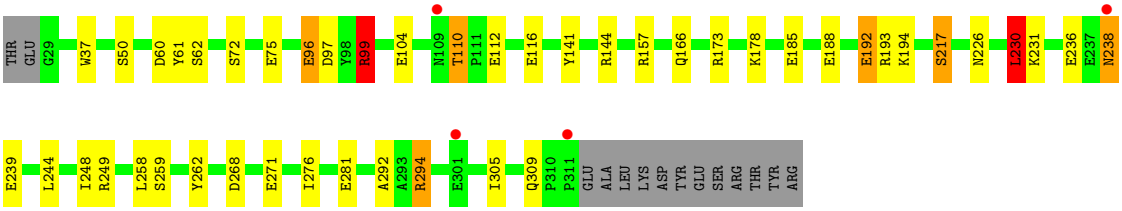
• Molecule 1: Uncharacterized protein TM_1410





• Molecule 1: Uncharacterized protein TM_1410





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	196.06Å 84.50Å 196.63Å 90.00° 119.83° 90.00°	Depositor
Resolution (Å)	49.26 – 2.60 49.26 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.3 (49.26-2.60) 98.6 (49.26-2.60)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.32 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.8.0425, REFMAC 5.8.0425	Depositor
R, R_{free}	0.216 , 0.244 0.218 , 0.245	Depositor DCC
R_{free} test set	8591 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	37.2	Xtriage
Anisotropy	0.291	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 38.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.110 for -h-l,k,h 0.110 for l,k,-h-l 0.055 for h,-k,-h-l 0.054 for -h-l,-k,l 0.055 for l,-k,h	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	27953	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.87% 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

5.1 Standard geometry

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	0.73	1/2431 (0.0%)	1.26	26/3301 (0.8%)
1	B	0.73	1/2406 (0.0%)	1.21	16/3268 (0.5%)
1	C	0.77	2/2390 (0.1%)	1.24	19/3244 (0.6%)
1	D	0.77	4/2406 (0.2%)	1.27	28/3268 (0.9%)
1	E	0.73	2/2402 (0.1%)	1.26	22/3263 (0.7%)
1	F	0.72	2/2431 (0.1%)	1.21	21/3303 (0.6%)
All	All	0.74	12/14466 (0.1%)	1.24	132/19647 (0.7%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	5
1	B	0	2
1	C	0	1
1	D	0	2
1	E	0	4
1	F	0	1
All	All	0	15

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	188	GLU	CD-OE2	13.08	1.40	1.25
1	E	192	GLU	CD-OE1	9.69	1.36	1.25
1	F	192	GLU	CD-OE1	8.36	1.34	1.25
1	D	192	GLU	CD-OE1	7.88	1.34	1.25
1	C	192	GLU	CD-OE1	7.79	1.34	1.25

The worst 5 of 132 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	188	GLU	CG-CD-OE1	-11.41	95.47	118.30
1	E	309	GLN	N-CA-CB	11.34	131.01	110.60
1	D	188	GLU	CG-CD-OE2	9.98	138.26	118.30
1	E	157	ARG	NE-CZ-NH2	-9.89	115.36	120.30
1	C	108	THR	CA-CB-OG1	-9.06	89.97	109.00

There are no chirality outliers.

5 of 15 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	173	ARG	Sidechain
1	A	191	ARG	Sidechain
1	A	249	ARG	Sidechain
1	A	294	ARG	Sidechain
1	A	302	MET	Mainchain

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	2358	2228	2217	15	0
1	B	2341	2217	2216	28	0
1	C	2327	2203	2202	12	0
1	D	2341	2217	2216	13	0
1	E	2337	2214	2213	17	0
1	F	2350	2224	2212	17	0
2	A	110	0	0	5	0
2	B	107	0	0	8	0
2	C	94	0	0	2	0
2	D	106	0	0	4	0
2	E	88	0	0	4	0
2	F	91	0	0	5	0
All	All	14650	13303	13276	101	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.

The worst 5 of 101 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:188:GLU:HG2	2:D:479:HOH:O	1.61	0.97
1:B:177:ARG:HH11	1:B:177:ARG:HG3	1.38	0.88
1:F:116:GLU:HB3	2:F:459:HOH:O	1.74	0.85
1:C:300:ASP:OD1	2:C:401:HOH:O	1.91	0.85
1:B:300:ASP:OD1	2:B:401:HOH:O	1.95	0.85

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	284/297 (96%)	274 (96%)	10 (4%)	0	100	100
1	B	281/297 (95%)	271 (96%)	10 (4%)	0	100	100
1	C	279/297 (94%)	270 (97%)	9 (3%)	0	100	100
1	D	281/297 (95%)	272 (97%)	9 (3%)	0	100	100
1	E	280/297 (94%)	270 (96%)	10 (4%)	0	100	100
1	F	283/297 (95%)	275 (97%)	8 (3%)	0	100	100
All	All	1688/1782 (95%)	1632 (97%)	56 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	253/263 (96%)	238 (94%)	15 (6%)	16	35
1	B	250/263 (95%)	237 (95%)	13 (5%)	19	41
1	C	248/263 (94%)	232 (94%)	16 (6%)	14	31
1	D	250/263 (95%)	236 (94%)	14 (6%)	17	38
1	E	250/263 (95%)	233 (93%)	17 (7%)	13	28
1	F	252/263 (96%)	238 (94%)	14 (6%)	17	38
All	All	1503/1578 (95%)	1414 (94%)	89 (6%)	16	35

5 of 89 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	281	GLU
1	E	271	GLU
1	E	61	TYR
1	E	188	GLU
1	F	50	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	309	GLN
1	C	238	ASN
1	E	181	ASN
1	E	226	ASN
1	F	309	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 oligosaccharides 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

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^{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(Å^2)	Q<0.9
1	A	284/297 (95%)	-0.70	6 (2%)	63	58	15, 34, 64, 108	1 (0%)
1	B	283/297 (95%)	-0.60	4 (1%)	73	68	18, 36, 66, 96	0
1	C	281/297 (94%)	-0.46	5 (1%)	67	62	16, 38, 69, 96	0
1	D	283/297 (95%)	-0.41	5 (1%)	67	62	20, 40, 68, 93	0
1	E	282/297 (94%)	-0.50	4 (1%)	73	68	19, 39, 69, 94	0
1	F	283/297 (95%)	-0.50	4 (1%)	73	68	12, 40, 68, 96	1 (0%)
All	All	1696/1782 (95%)	-0.53	28 (1%)	69	64	12, 38, 69, 108	2 (0%)

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	29	GLY	7.1
1	A	312	GLU	5.6
1	B	29	GLY	5.3
1	C	29	GLY	5.0
1	F	311	PRO	4.2

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 monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.