

DOI: <http://doi.org/10.32792/utq.jceps.10.01.01>

Exploring Sequence Conservation and Functional Diversity in Beta-Galactosidase Enzymes: A Comparative Analysis of Mesophilic and Thermophilic Bacterial Species

Anaheed H. Kareem¹, Mutaman H. Abdullah^{2*}, Ahmed T. Abduladheem^{3*},

anaheedhussein@gmail.com

Sefatullah Zakary^{4*}, Abdul Q. Basser^{5*} and Mohd S. Shamsir^{6*}.

¹ Department of Prosthodontics, College of Health and Medical Technology, Al-Ayen University, Thi-Qar, Iraq.

²Advanced Medical & Dental Institute, University Science Malaysia, Bandar Putra Bertam, Kepala Batas, Pulau Pinang, Malaysia.

³Department of Community Health College of Health and Medical Technologies, National University for Science and Technology, Thi-Qar-64001,

IRAQ.

⁴Department of Botany, Faculty of Biology, Kabul, Afghanistan.

⁵Department of Botany, Faculty of Biology, Kabul, Afghanistan.

⁶Department of Bioscience, Faculty of Science, University Technology Malaysia, Johor Bahru, Skudai.

Received 30/10/2023, Accepted 23/11/2023, Published 1/12/2023



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Abstract:

The structural comparative analysis of β -Galactosidase in *Sulfolobus acidocaldarius* and *Escherichia coli* highlights its thermostable nature and potential industrial applications under extreme conditions. This study investigates the 3D structure of β -Galactosidase in the thermophilic archaeon *Sulfolobus acidocaldarius* and compares it with the mesophilic bacterium *Escherichia coli*. The aim is to highlight the distinct structural features of the thermostable β -Galactosidase in *S. acidocaldarius*, emphasizing its potential for industrial applications under extreme conditions. The absence of a known 3D structure for this enzyme in *S. acidocaldarius* prompted modeling efforts. The findings reveal significant structural differences, particularly in thermal stability, making *S. acidocaldarius* β -Galactosidase promising for applications in the dairy industry, pharmaceuticals, and biotechnology. This study underscores the importance of understanding extremophile enzymes' adaptability to extreme environments and their potential for biotechnological advancements. The comparative analysis lays the foundation for future research aimed at harnessing thermostable β -Galactosidase enzymes' unique properties, offering innovative possibilities across various industries.

Keywords: bioinformatics, thermostability, β -Galactosidase, *S. acidocaldarius*.

1-Introduction

β -Galactosidase (lactase; EC 3.2.1.23) is a multifunctional enzyme known for its critical role in *lactose metabolism* [1]. It possesses three distinct enzymatic activities: the hydrolysis of lactose into glucose and galactose, the catalysis of trans-galactosylation to form allolactose, and the subsequent cleavage of allolactose into monosaccharides[2]. This enzyme, which is essentially a tetramer comprising four identical polypeptide chains consisting of 409 amino acids [3], is sourced from various microorganisms, including bacteria, fungi, and yeasts, as well as certain plant species like almonds, peaches, apples, and apricots. Notably, β -Galactosidase enzymes derived from bacterial sources are of particular interest due to their high activity, ease of production, and stability, making them valuable assets in industrial processes, especially those involving lactose-containing fluids [4].

β -Galactosidase plays a crucial role in alleviating lactose intolerance, a common issue affecting individuals worldwide, by facilitating the digestion of dairy products such as milk [5]. Additionally, the production of sweet syrups and their utilization in the cheese industry, soft drinks, ice cream, and confectionery benefit from the enzymatic capabilities of β -Galactosidase. Furthermore, this enzyme is integral to pharmaceutical drug development, offering unique opportunities for drug synthesis [6].

Sulfolobus acidocaldarius has been discovered as a thermoacidophilic archaeon capable of thriving in extreme conditions, including high temperatures (75-80°C) and low pH (optimal range: pH 2-3) [7]. This microorganism, belonging to the phylum Crenarchaeota, exhibits facultative autotrophy for sustenance. Initially isolated from the geothermal environments of Yellowstone National Park, USA, *S. acidocaldarius* inhabits hot springs with pH levels below 3 and temperatures ranging from 65-90°C [8]. This unique extremophile serves as a natural source of thermostable enzymes, including β -Galactosidase, with promising industrial potential [9].

Despite the significance of β -Galactosidase in *S. acidocaldarius*, there is a paucity of structural data for this enzyme in publicly accessible databases such as UniProt and the Protein Data Bank (PDB). This study endeavours to bridge this gap by employing computational modelling techniques to elucidate the 3D structure of thermostable β -Galactosidase in *S. acidocaldarius*. Subsequently, we aim to compare this structure with its mesophilic counterpart from *Escherichia coli*. Such a comparative analysis will unveil structural similarities and differences, paving the way for a deeper understanding of these enzymes' adaptability to extreme environments and their potential in various industrial applications.

This investigation holds promise for harnessing the unique properties of thermostable β -Galactosidase enzymes from extremophiles like *S. acidocaldarius*, offering prospects for innovative solutions in biotechnology, pharmaceuticals, and other industrial sectors.

2-Methodology of proposed methods

The FASTA format and other information regarding the PDB structure (no 3D structure) are received from UniProt. Then, we performed the UniProt blast to obtain similar sequences for Beta-galactosidase of thermophilic and mesophilic bacterial species for comparison sequence and structure.

The alignment was to make a meticulous comparative analysis of Beta-galactosidase enzymes sourced from a diverse array of both mesophilic and thermophilic bacterial species—the selection of these enzymes aimed to unravel the evolutionary and functional attributes inherent in their sequences.

For the thermophilic group, the target sequence of Beta-galactosidase was aligned with counterparts from ten distinct bacterial species known for their thermophilic adaptations. These species encompass *Sulfolobus acidocaldarius*, *Saccharolobus solfataricus*, *Acidilobus saccharovorans*, *Thermococcus celer*, *Pyrococcus furiosus*, *Thermococcus sibiricus*, *Pseudothermotoga hypogea*, *Pyrococcus woesei*, *Thermococcus sibiricus*, and *Ferroplasma acidarmanus*. This alignment revealed critical residues that have been evolutionarily preserved across these thermophilic enzymes, guiding our understanding of their structural and functional significance.

Concurrently, for the mesophilic ensemble, the target sequence of Beta-galactosidase from *Escherichia coli* was aligned with counterparts from ten distinct mesophilic bacterial species. These species encompass *Escherichia coli*, *Shigella sonnei*, *Klebsiella pneumoniae*, *Pluralibacter gergoviae*, *Shigella dysenteriae*, *Citrobacter rodentium*, *Superficieibacter electus*, *Salmonella arizonae*, *Enterobacteriaceae bacterium*, and *Izhakiella australiensis*.

This mesophilic alignment provides a valuable perspective on the shared and unique sequence attributes within the mesophilic bacterial realm.

Throughout the alignment process, residues that exhibited evolutionary conservation were denoted with asterisks (*), while those showcasing conservation among groups with similar properties were marked with colons (:). Conservation between groups with weakly similar properties was indicated by periods (.) within the alignment, while variable regions were aptly designated by gaps.

This comprehensive alignment initiative aims to unravel the evolutionary and functional intricacies of Beta-galactosidase enzymes across a spectrum of temperature-adapted bacterial species, contributing to our broader understanding of these enzymes' adaptations and potential applications in diverse fields.

Table 1: A curated list of 10 thermophilic enzymes, derived from extremophilic organisms adapted to high-temperature environments, has been thoughtfully selected for alignment. These enzymes, which include DNA polymerase, amylase, lipase, protease, cellulase, catalase, ribonuclease, dehydrogenase, phosphatase, and esterase, offer a fascinating opportunity to explore sequence conservation and the unique structural features that enable them to function effectively in extreme thermal conditions. This alignment provides valuable insights into the molecular adaptations of thermophilic enzymes, shedding light on their potential applications in biotechnology, industrial processes, and environmental remediation.

Table 1. List of 10 thermophilic enzymes

Uniprot ID	Gene Name	Organism
14288P	1849bgaS Sac_	<i>Sulfolobus acidocaldarius</i>
22498P	3019lazS SSO	<i>Saccharolobus solfataricus</i>
50388P	bglY	<i>Saccharolobus shibatae</i>
08PZ9D	1390ASAC_	<i>Acidilobus saccharovorans</i>
180P218A0A	03345_03L3A	<i>Thermococcus celer</i>
14KZ1Q	1453PTO	<i>torridus Picrophilus</i>
VKSH1X0A0A	08155_81AJ	<i>Pseudothermotoga hypogea</i>
52629O		<i>Pyrococcus woesei</i>
190A6C	0324TSIB_	<i>Thermococcus sibiricus</i>
3ATA0S	1310G0001FACI_IFERC	<i>Ferroplasma acidarmanus</i> <i>1fert</i>

Table 2: A list of 10 mesophilic enzymes has been thoughtfully curated for alignment. These enzymes encompass a diverse range of functions, including alpha-amylase, lipase, protease, cellulase, catalase, ribonuclease, dehydrogenase, phosphatase, esterase, and lysozyme. This selection allows for a comprehensive examination of the conserved elements and functional attributes within mesophilic proteins through sequence alignment and analysis.

Table 2. List of 10 mesophilic enzymes has been thoughtfully curated for alignment

Uniprot ID	Name Gene	Organism
2.GU^Q	1SGR	<i>Rauvolfia serpentine</i>
1.W^V^L.A.A	109600.g.1Vigan_ε^LR	<i>Phaseolus angularis</i>
1.TKG.103A.A	1.6706360LOC	<i>Vigna radiata var. radiata</i>
1.CTP^V	g.76700.G.1PHAVU_	<i>Phaseolus vulgaris</i>
1.LJS^I	1.07767.8 129700.G^1GLYMA_ 129800.G^1GLYMA_	<i>max Glycine</i>
1.RNP^B.A.A	1.29791_60Y.D 10000.glysoja_	<i>Glycine soja</i>
1.HLεε0A.A	1.33447_60Y.D	<i>Glycine soja</i>
ε1RG101A.A	1.37127_1KK	<i>Gajanus cajan</i>
1.IHR^G	1.9802.g2MTR_11430241	<i>Medicago truncatula</i>
1.GSL.61A.A	1.4408TCM_	<i>Theobroma cacao</i>

3-Results

3.1 Sstructure predictio econdaryn

The determination of the secondary structure of the enzyme *Beta-galactosidase* in *Sulfolobus acidocaldarius* was carried out with utmost precision using the CFSSP (Chou-Fasman Secondary Structure Prediction) database, accessible at <http://www.biogem.org/tool/chou-fasman/>. This robust computational tool was employed to predict the specific regions within the amino acid sequence of Beta-galactosidase that correspond to various secondary structure elements, notably including alpha helices, beta sheets, and turns. This analytical approach provides essential insights into the protein's structural composition, thereby enhancing our comprehension of its functional attributes and contributing to a more comprehensive understanding of its biochemical properties.

3.2 Primary sequence

The amino acid sequences of beta-galactosidase sourced from *Sulfolobus acidocaldarius* were retrieved meticulously from UniProt, specifically under the UniProt Identifier (ID): P14288. This precise data acquisition from a reputable and well-curated protein database ensures the accuracy and reliability of the sequence information, forming a solid foundation for subsequent analyses and investigations in our research endeavours.

```
>sp|P14288|BGAL_SULAC Beta-galactosidase OS=Sulfolobus acidocaldarius (strain  
ATCC 33909 / DSM 639 / JCM 8929 / NBRC 15157 / NCIMB 11770) OX=330779  
GN=bgaS PE=1 SV=2  
MLSFPKGFKFGWSQSGFQSEMGTGSEDPNSDWHVWVHDRENIVSQVVSDDL  
LNAVRINVEWSRIFPRPLPKPEMQTGTDKPENGPGYWGNYKRFHDEAEKIG  
NSPVISVDL NESKLREMDNYANHEALSHYRQILEDLRNRGFHIVLNMYHWTL  
IWLHDPPIRVRRGDFTGPTGWLNSRTVYEFARFSAYVAWKLDDLASEYATMNE  
PNVWVGAGYAFPRAGFPPNYLSFRLSEIAKWNIIQAHARAYDAIKSVSKKSVGI  
IYANTSYYPLRPQDNEAVEIAERLNRWSFFDSIIKGEITSEGQNVREDLRNRLD  
KAESGYLTLPYGDR CERN SLSLANLPTSDFGWEFFPEGLWIGVNY YTRTVVT  
YDVLLKYWNRYGLPLYVMENGIADDADYQRPYYLVSHIYQVHRALNEGVDV  
RGYLHWSLADNYEWSSGFSMRFGLLKVDYLTKRLYWRPSALVYREITRSNGI  
PEELEHLNRVPPIKPLRH
```

The primary sequence analysis result by using protparam from beta-glucosidase in Sulfolobus acidocaldarius ID (P14288):, Number of amino acids: 491, Molecular weight: 57143.24

Table 2.Amino acid composition.

Ala (A) 27	5.5%
Arg (R) 37	7.5%
Asn (N) 31	6.3%
Asp (D) 27	5.5%
Cys (C) 1	0.2%
Gln (Q) 10	2.0%
Glu (E) 34	6.9%
Gly (G) 34	6.9%
His (H) 14	2.9%
Ile (I) 25	5.1%
Leu (L) 43	8.8%
Lys (K) 18	3.7%
Met (M) 8	1.6%
Phe (F) 20	4.1%
Pro (P) 29	5.9%
Ser (S) 36	7.3%
Thr (T) 17	3.5%
Trp (W) 18	3.7%
Tyr (Y) 31	6.3%
Val (V) 31	6.3%
Pyl (O) 0	0.0%

Sec (U) 0	0.0%
-----------	------

Total number of negatively charged residues (Asp + Glu): 61
Total number of positively charged residues (Arg + Lys): 55

Table 4. Atomic composition.

Carbon	C	2599
Hydrogen	H	3881
Nitrogen	N	707
Oxygen	O	739
Sulphur	S	9

Formula: C₂₅₉₉H₃₈₈₁N₇₀₇O₇₃₉S₉

Total number of atoms: 7935

3.3 Sequence alignment analysis

Sequence alignment analysis is a vital technique in molecular biology that involves comparing biological sequences, like DNA, RNA, or proteins, to identify regions of similarity. It helps reveal evolutionary relationships, conserved elements, and structural motifs. By introducing gaps, it accommodates insertions or deletions in sequences. This analysis is fundamental in various fields, including phylogenetics, protein structure prediction, and functional annotation, providing insights into biomolecule structure and function.

3.3.1 Thermophilic

These microorganisms are known for their ability to thrive and reproduce in high-temperature environments, often exceeding 45 degrees Celsius (113 degrees Fahrenheit). These heat-loving bacteria have adapted to extreme conditions, such as hot springs, volcanic vents, and geothermal areas. Their unique enzymes and cellular structures are adapted to function optimally at elevated temperatures, making them of particular interest in biotechnology and industrial applications, where heat-resistant enzymes play a crucial role. Thermophilic bacteria offer valuable

insights into extremophiles' biology and the potential for harnessing their specialized traits for various scientific and practical purposes.



Figure .1: Illustrating about conserve domain.

Figure1 show conserve domain group that was founded in packages one , two , four , five and six . This similarity means there are important segments which have functional roles in these proteins and without these conserved sequences the proteins are denatured and suspended. Moreover, there are active sites common between these groups except for two of them. Besides that, there is no binding site in thermophilic groups except one that has a binding site in ‘R’ and ‘K’.

3.3.2 Mesophilic

These microorganisms that thrive in moderate temperature conditions are typically found on Earth. They prefer temperatures ranging from about 20 to 45 degrees Celsius (68 to 113 degrees Fahrenheit). These adaptable bacteria can be found in various environments, including soil, water, and the human body. Many mesophilic bacteria are essential for processes like decomposition,

fermentation, and nutrient cycling, making them crucial for maintaining ecological balance. They are also commonly used in biotechnology and food industries for processes like fermentation and the production of various products. Mesophilic bacteria play a significant role in both natural ecosystems and human activities, making them a widely studied and applied group of microorganisms.

However, the comprehensive analysis across various packages has revealed a plethora of striking similarities within the mesophilic groups of proteins. These commonalities extend beyond mere sequence conservation and traverse the realms of conserved domains, roles, and functions. One of the most noteworthy findings is the presence of conserved domains that transcend different protein packages, signifying a shared evolutionary heritage and functional significance among these mesophilic proteins. This structural and functional conservation points to a fundamental role these domains play in the biological processes of these organisms.

These findings illuminate the adaptive strategies of mesophilic organisms and highlight their ability to thrive in moderate-temperature environments by harnessing shared molecular mechanisms and functional motifs. The recognition of these conserved features paves the way for a deeper understanding of the intricate relationships between sequence, structure, and function in these mesophilic proteins, ultimately advancing our knowledge of their biology and potential biotechnological applications.

```
P00722 BGAL_ECOLI 241 EAEVQMCSELRDYLRTVTSIQGEIQVAGSTAPFGSEIIDERGGYADRVTLRLNVENFK 300
Q32583 BGAL_SHISS 241 EAEVQMYSELRDELRTVTSIQGEIQVAGSTAPFGSEIIDERGGYADRVTLRLNVENFA 300
A6TI29 BGAL2_KLEP7 241 EADVQMYSELRDELRTVTSIQGEIQVAGSTAPFGSEIIDERGGYADRVTLRLNVENFK 300
Q32JB6 BGAL_SHIDS 241 EAEVQMYSELRDELRTVTSIQGEIQVAGSTAPFGSEIIDERGGYADRVTLRLNVENFK 300
D2TK51 D2TK51_CITRI 241 EAEVRLRSELCDELRTVTSIQGEIQVAGSTAPFGSEIIDERGGYADRVTLRLNVENFK 300
A9MQ82 BGAL_SALAR 241 EAEVRLRSELCDELRTVTSIQGEIQVAGSTAPFGSEIIDERGGYADRVTLRLNVENFA 300
L0MA64 L0MA64_ENTBF 241 EADVRIAAGNVQHDVQVLELHKNKQSLIGQVSARPSAPVDERGNYALRALICLPEVEHFA 300
A0A0J5KFW5 A0A0J5KFW5_PLUGE 241 EADVQMYSELRDELRTVTSIQGEIQVAGSTAPFGSEIIDERGGYADRVTLRLNVENFK 300
A0A2P5GMD4 A0A2P5GMD4_9ENTR 241 EAEVRIAAGLQDDLOVELELHKNKQSLIGQVTSISLGESEIIDERGGYADRVTLRLNVENFA 300
A0A4P8YKQ3 A0A4P8YKQ3_9ENTR 241 EALMALDSDVTDQDCQIAIQGLDGLSLIGEQQAVGSDIIVDERGAWRDRTRTLRAVKKOFR 300
```

```
P00722 BGAL_ECOLI 301 LWSAIIINLYRAVVEIHTADGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 360
Q32583 BGAL_SHISS 301 LWSAIIINLYRAVVEIHTDDGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 360
A6TI29 BGAL2_KLEP7 301 LWSAIIINLYRAVVEIHTADGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 360
Q32JB6 BGAL_SHIDS 301 LWSAIIINLYRAVVEIHTADGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 360
D2TK51 D2TK51_CITRI 301 LWSAIIINLYRAVVEIHTADGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 360
A9MQ82 BGAL_SALAR 301 LWSAIIINLYRAVVEIHTADGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 360
L0MA64 L0MA64_ENTBF 301 LWSAIIINLYRAVVEIHTADGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 360
A0A0J5KFW5 A0A0J5KFW5_PLUGE 301 LWSAIIINLYRAVVEIHTADGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 360
A0A2P5GMD4 A0A2P5GMD4_9ENTR 301 LWSAIIINLYRAVVEIHTADGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 360
A0A4P8YKQ3 A0A4P8YKQ3_9ENTR 301 LWSAIIINLYRAVVEIHTADGTLIEAEACDVGFREVRLENGLLLNCKPFLIRGVNRHEH 359
```

```
P00722 BGAL_ECOLI 361 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 420
Q32583 BGAL_SHISS 361 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 420
A6TI29 BGAL2_KLEP7 361 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 420
Q32JB6 BGAL_SHIDS 361 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 420
D2TK51 D2TK51_CITRI 361 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 420
A9MQ82 BGAL_SALAR 361 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 420
L0MA64 L0MA64_ENTBF 361 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 420
A0A0J5KFW5 A0A0J5KFW5_PLUGE 361 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 420
A0A2P5GMD4 A0A2P5GMD4_9ENTR 361 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 420
A0A4P8YKQ3 A0A4P8YKQ3_9ENTR 360 HPLHGVMDQEQVMVDIILMKQNNFNNAVRCSHYPNHFLVITLCPDRYGLYVVDEANIEIHC 419
```

```
P00722 BGAL_ECOLI 421 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 480
Q32583 BGAL_SHISS 421 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 480
A6TI29 BGAL2_KLEP7 421 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 480
Q32JB6 BGAL_SHIDS 421 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 480
D2TK51 D2TK51_CITRI 421 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 480
A9MQ82 BGAL_SALAR 421 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 480
L0MA64 L0MA64_ENTBF 421 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 480
A0A0J5KFW5 A0A0J5KFW5_PLUGE 421 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 480
A0A2P5GMD4 A0A2P5GMD4_9ENTR 421 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 480
A0A4P8YKQ3 A0A4P8YKQ3_9ENTR 420 MFMNRLTDDPRFLPAMSRVTRMVRQDRNHPSVLIWSLGN SGHGANNHDALYRWIKASV 479
```

```
P00722 BGAL_ECOLI 1 MTMITDSLAVVLQRRDWNPCVTLNRLAAHPPEASWRNSEEARTDRPSQQLRSLNGEWR 60
Q32583 BGAL_SHISS 1 MIMITDSLAVVLQRRDWNPCVTLNRLAAHPPEASWRNSEEARTDRPSQQLRSLNGEWR 60
A6TI29 BGAL2_KLEP7 1 MTMITDSLAVVLQRRDWNPCVTLNRLAAHPPEASWRNSEEARTDRPSQESRSLNGEWR 60
Q32JB6 BGAL_SHIDS 1 MTMITDSLAVVLQRRDWNPCVTLNRLAAHPPEASWRNSEEARTNRPSQQLRSLNGEWO 60
D2TK51 D2TK51_CITRI 1 MNLNTDSLAAVLAARRDWNPCVTLNRLAAHPPEASWRNSEEARTDRPSQQLRSLNGEWR 60
A9MQ82 BGAL_SALAR 1 MTPERDSLAAVLAARRDWNPCVTLNRLAAHPPEASWRNSEEARTDRPSQESRSLNGEWR 60
L0MA64 L0MA64_ENTBF 1 MTSEQDSLAAVLAARRDWNPCVTLNRLAAHPPEASWRNSEEARTDRPSQQLRSLNGEWO 60
A0A0J5KFW5 A0A0J5KFW5_PLUGE 1 MTMITDSLAVVLQRRDWNPCVTLNRLAAHPPEASWRNSEEARTDRPSQESRSLNGEWR 60
A0A2P5GMD4 A0A2P5GMD4_9ENTR 1 MSTGDSLAAVLAARRDWNPCVTLNRLAAHPPEASWRNSEEARTDRPSQESRSLNGEWO 60
A0A4P8YKQ3 A0A4P8YKQ3_9ENTR 1 MTVKALSLAAVLAARRDWNPCVTLNRLAAHPPEASWRNSEEARTDRPSQQLRSLNGAAR 60
```

```
P00722 BGAL_ECOLI 61 FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
Q32583 BGAL_SHISS 61 FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
A6TI29 BGAL2_KLEP7 61 FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
Q32JB6 BGAL_SHIDS 61 FVWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
D2TK51 D2TK51_CITRI 61 FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
A9MQ82 BGAL_SALAR 61 FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
L0MA64 L0MA64_ENTBF 61 FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
A0A0J5KFW5 A0A0J5KFW5_PLUGE 61 FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
A0A2P5GMD4 A0A2P5GMD4_9ENTR 61 FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
A0A4P8YKQ3 A0A4P8YKQ3_9ENTR 61 FAWFPAPEAVPESWLECDLPEADTVVPSNWQMHGYDAPITYVTYPIVNPFFVPEENP 120
```

```
P00722 BGAL_ECOLI 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
Q32583 BGAL_SHISS 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
A6TI29 BGAL2_KLEP7 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
Q32JB6 BGAL_SHIDS 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
D2TK51 D2TK51_CITRI 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
A9MQ82 BGAL_SALAR 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
L0MA64 L0MA64_ENTBF 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
A0A0J5KFW5 A0A0J5KFW5_PLUGE 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
A0A2P5GMD4 A0A2P5GMD4_9ENTR 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
A0A4P8YKQ3 A0A4P8YKQ3_9ENTR 121 TGCYSLTFENVDESILQEGQTRIIIDGVNSAFHLWCNRRVVGYGODSRLPSEFDLSAFIRA 180
```

```
P00722 BGAL_ECOLI 181 GENRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
Q32583 BGAL_SHISS 181 GKNRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
A6TI29 BGAL2_KLEP7 181 GENRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
Q32JB6 BGAL_SHIDS 181 GENRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
D2TK51 D2TK51_CITRI 181 GENRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
A9MQ82 BGAL_SALAR 181 GENRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
L0MA64 L0MA64_ENTBF 181 GENRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
A0A0J5KFW5 A0A0J5KFW5_PLUGE 181 GENRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
A0A2P5GMD4 A0A2P5GMD4_9ENTR 181 GENRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
A0A4P8YKQ3 A0A4P8YKQ3_9ENTR 181 GANRLAVVLRWSDGSYLEDDMMWRMSGIFRDVSLHKEPTQISDPFHVATHFNDDFSRAV 240
```

P00722	BGAL_ECOLI	718	WQWRLEAENSVTLPAAASHAI	PHLTTSEMDFCEELGNKRWOENRQSEF	LSQMWIGDKKQ	777
Q3Z583	BGAL_SHISS	718	WQWRLEAENSVTLPAAASHAI	PHLTTSEMDFCEELGNKRWOENRQSEF	LSQMWIGDKKQ	777
A6TI29	BGAL2_KLEP7	718	WQWRLEAENSVTLPAAASHAI	PHLTTSEMDFCEELGNKRWOENRQSEF	LSQMWIGDKKQ	777
Q3ZJB6	BGAL_SHIDS	718	WQWRLEAENSVTLPAAASHAI	PHLTTSEMDFCEELGNKRWOENRQSEF	LSQMWIGDKKQ	777
D2TK51	D2TK51_CITRI	721	WQWRPLAEKLSVTLPRAAAA	POLKVENAAFEVVVQQRWQFCRQRST	LSQYWLADAQ	780
A9MQ82	BGAL_SALAR	721	WQWRPLAEKLSVTLPRAAAA	POLKVENAAFEVVVQQRWQFCRQRST	LSQYWLADAQ	780
LOMA64	LOMA64_ENTBF	721	WQWRLEAENSVTLPAAASHAI	PHLTTSEMDFCEELGNKRWOENRQSEF	LSQMWIGDKKQ	780
AOA0J5KFW5	AOA0J5KFW5_PLUGE	718	WQWRLEAENSVTLPAAASHAI	PHLTTSEMDFCEELGNKRWOENRQSEF	LSQMWIGDKKQ	777
AOA2P5GMD4	AOA2P5GMD4_9ENTR	721	WQWRLEAENSVTLPAAASHAI	PHLTTSEMDFCEELGNKRWOENRQSEF	LSQMWIGDKKQ	780
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	717	WQWRLEAENSVTLPAAASHAI	PHLTTSEMDFCEELGNKRWOENRQSEF	LSQMWIGDKKQ	776
*** **						
P00722	BGAL_ECOLI	778	LFPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	837
Q3Z583	BGAL_SHISS	778	LFPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	837
A6TI29	BGAL2_KLEP7	778	LFPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	837
Q3ZJB6	BGAL_SHIDS	778	LFPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	837
D2TK51	D2TK51_CITRI	781	LFPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	840
A9MQ82	BGAL_SALAR	781	LFPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	840
LOMA64	LOMA64_ENTBF	781	LFPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	840
AOA0J5KFW5	AOA0J5KFW5_PLUGE	778	LFPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	837
AOA2P5GMD4	AOA2P5GMD4_9ENTR	781	LFPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	840
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	777	ATPLRDOETRAPLNDNDIGVSE	ATRIDPNANVERWKAAGHYQAEAAAL	LOCTADTLADAVLI	836
*** **						
P00722	BGAL_ECOLI	838	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	897
Q3Z583	BGAL_SHISS	838	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	897
A6TI29	BGAL2_KLEP7	838	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	897
Q3ZJB6	BGAL_SHIDS	838	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	897
D2TK51	D2TK51_CITRI	841	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	900
A9MQ82	BGAL_SALAR	841	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	900
LOMA64	LOMA64_ENTBF	841	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	900
AOA0J5KFW5	AOA0J5KFW5_PLUGE	838	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	897
AOA2P5GMD4	AOA2P5GMD4_9ENTR	841	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	900
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	837	TTAHAWCHOGKFLFISRKTYR	IDGSSGOMALTVDVEVASDTEHPARI	GLINCOLAQVAERVN	896
*** **						
P00722	BGAL_ECOLI	898	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	957
Q3Z583	BGAL_SHISS	898	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	957
A6TI29	BGAL2_KLEP7	898	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	957
Q3ZJB6	BGAL_SHIDS	898	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	957
D2TK51	D2TK51_CITRI	901	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	960
A9MQ82	BGAL_SALAR	901	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	960
LOMA64	LOMA64_ENTBF	901	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	960
AOA0J5KFW5	AOA0J5KFW5_PLUGE	898	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	957
AOA2P5GMD4	AOA2P5GMD4_9ENTR	901	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	960
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	897	WLGGLPPOENYPDRLLTAAC	FDROWLPLSDMYTPYVFPSENGLR	CETREINYPHOWRGDFQ	956
*** **						
P00722	BGAL_ECOLI	481	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	540
Q3Z583	BGAL_SHISS	481	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	540
A6TI29	BGAL2_KLEP7	481	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	540
Q3ZJB6	BGAL_SHIDS	481	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	540
D2TK51	D2TK51_CITRI	481	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	540
A9MQ82	BGAL_SALAR	481	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	540
LOMA64	LOMA64_ENTBF	481	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	540
AOA0J5KFW5	AOA0J5KFW5_PLUGE	481	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	540
AOA2P5GMD4	AOA2P5GMD4_9ENTR	481	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	540
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	480	PSRPVQYEGGGADTTATDII	ICPMYARVDEDDQPPAVPKWSIKK	WLSLPGCETRPLILCYA	539
*** **						
P00722	BGAL_ECOLI	541	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	600
Q3Z583	BGAL_SHISS	541	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	600
A6TI29	BGAL2_KLEP7	541	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	600
Q3ZJB6	BGAL_SHIDS	541	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	600
D2TK51	D2TK51_CITRI	541	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	600
A9MQ82	BGAL_SALAR	541	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	600
LOMA64	LOMA64_ENTBF	541	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	600
AOA0J5KFW5	AOA0J5KFW5_PLUGE	541	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	600
AOA2P5GMD4	AOA2P5GMD4_9ENTR	541	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	600
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	540	HAMGNSLGGFAKYWAQFRQY	PRLQGGFVVDWDVQDQLIKYDENG	PNWSAYGGDFGDTPNDR	599
*** **						
P00722	BGAL_ECOLI	601	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	657
Q3Z583	BGAL_SHISS	601	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	657
A6TI29	BGAL2_KLEP7	601	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	657
Q3ZJB6	BGAL_SHIDS	601	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	657
D2TK51	D2TK51_CITRI	601	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	660
A9MQ82	BGAL_SALAR	601	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	660
LOMA64	LOMA64_ENTBF	601	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	660
AOA0J5KFW5	AOA0J5KFW5_PLUGE	601	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	657
AOA2P5GMD4	AOA2P5GMD4_9ENTR	601	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	660
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	600	QFCMGLVFADRTPHPALTEAKH	QOOFFQFRLLSGQTTEVTSEYLFRHS	DNELLHWV	657
*** **						
P00722	BGAL_ECOLI	658	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	717
Q3Z583	BGAL_SHISS	658	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	717
A6TI29	BGAL2_KLEP7	658	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	717
Q3ZJB6	BGAL_SHIDS	658	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	717
D2TK51	D2TK51_CITRI	661	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	720
A9MQ82	BGAL_SALAR	661	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	720
LOMA64	LOMA64_ENTBF	661	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	720
AOA0J5KFW5	AOA0J5KFW5_PLUGE	658	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	717
AOA2P5GMD4	AOA2P5GMD4_9ENTR	661	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	720
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	658	ALDGKPLASGEVPLDVAPOG	QKQIIELELPELPQPSAGQLWLT	VHVQPNATWSAAGHISA	716
*** **						

P00722	BGAL_ECOLI	958	FNISRYSQQLMETSRRHLLHAEECTNLNIDGFHMGIGGDDS	SPSVSAEFQLSAGRHYH	1017	
Q3Z583	BGAL_SHISS	958	FNISRYSQQLMETSRRHLLHAEECTNLNIDGFHMGIGGDDS	SPSVSAEFQLSAGSYHY	1017	
A6TI29	BGAL2_KLEP7	958	FNISRYSQQLMETSRRHLLHAEECTNLNIDGFHMGIGGDDS	SPSVSAEFQLSAGSYHY	1017	
Q32JB6	BGAL_SHIDS	958	FNISRYSQQLMETSRRHLLHAEECTNLNIDGFHMGIGGDDS	SPSVSAEFQLSAGRHYH	1017	
D2TK51	D2TK51_CITRI	961	FNISRYSQQLMETSRRHLLRENGTNLNIDGYHMGVGGDDS	SPSVSPEYQLSAGRHYH	1020	
A9MQ82	BGAL_SALAR	961	FNISRYSQQLMETSRRHLLQAAGVNLNIDGYHMGVGGDDS	SPSVSPEFQLSARHYHY	1020	
L0MA64	L0MA64_ENTBF	961	FNISRYSQQLMETSRRHLLQAAGVNLNIDGYHMGVGGDDS	SPSVSPEFQLSARHYHY	1020	
A0A0J5KFW5	A0A0J5KFW5_PLUGE	958	FNISRYSQQLMETSRRHLLHAEECTNLNIDGFHMGIGGDDS	SPSVSAEFQLSAGSYHY	1017	
A0A2P5GMD4	A0A2P5GMD4_9ENTR	961	FNISRYSQQLMETSRRHLLQAAGVNLNIDGYHMGVGGDDS	SPSVSPEFQLSARHYNY	1020	
A0A4P8YKQ3	A0A4P8YKQ3_9ENTR	957	FNLSRYSQQLMETSRRHLLKEEFGS	NLNIDGFHMGVGGDDS	SPSVAPPELLFQRHYHY	1016
			:**:*****: * * *****:*****:*****: * * *			
P00722	BGAL_ECOLI	1018	QLVWCQK		1024	
Q3Z583	BGAL_SHISS	1018	QLVWCQK		1024	
A6TI29	BGAL2_KLEP7	1018	QLVWCQK		1024	
Q32JB6	BGAL_SHIDS	1018	QLVWCQK		1024	
D2TK51	D2TK51_CITRI	1021	QLVWGQK		1027	
A9MQ82	BGAL_SALAR	1021	QLTARK--		1025	
L0MA64	L0MA64_ENTBF	1021	QLINQ--		1025	
A0A0J5KFW5	A0A0J5KFW5_PLUGE	1018	QLVWCQK		1024	
A0A2P5GMD4	A0A2P5GMD4_9ENTR	1021	QLINE--		1025	
A0A4P8YKQ3	A0A4P8YKQ3_9ENTR	1017	AVSARR		1023	
			:			

Figure .2: Similarities between mesophilic groups.

Figure 2 Provides information about more similarities between mesophilic groups that are found in all packages. This means there are many conserved domains common between these types of proteins that give similar roles and functions. Besides that, mesophilic groups show similarity in binding sites and active sites which are found in different packages.

There is a common conserved domain between meso and thermo groups that begin from 367 to 966 terminals. That means there is a sharing segment and an important position is playing a functional role in this enzyme. Moreover, among twenty of them, there is no common binding site and the active site is a tool used to study closely related genes or proteins to find the evolutionary relationships between genes and to identify shared patterns among functionally or structurally related genes. Besides that, many programs provide both progressive global and local alignments. Moreover, we can identify active sites, similarities, and binding sites as well as make phylogenetic

P14288	BGAL_SULAC	1	-----	0
P22498	BGAL_SACS2	1	-----	0
P50388	BGAL_SACSH	1	-----	0
D9P208	D9P208 ACIS3	1	-----	0
AOA218P180	AOA218P180 THECE	1	-----	0
Q6KZ14	Q6KZ14 PICTO	1	-----	0
AOA0X1KSH7	AOA0X1KSH7_9THEM	1	-----	0
O52629	BGAL_PYRWO	1	-----	0
C6A195	C6A195 THESM	1	-----	0
SOATA3	SOATA3_FERAC	1	-----	0
P00722	BGAL_ECOLI	121	TGCYSLTFNVDES WLQEGQTRI IPDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLRA	180
Q3Z583	BGAL_SHISS	121	TGCYSLTFNIDESWLQEGQTRI IPDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLRA	180
A6TI29	BGAL2_KLEP7	121	TGCYSLTFNIDESWLQEGQTRI IPDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLHA	180
Q3ZJB6	BGAL_SHIDS	121	TGYYSLTFNVDES WLQEGQTRI IPDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLRA	180
D2TK51	D2TK51 CITRI	121	TGCYSLTFTVDDAWLREGQTRI IPDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSAFLTS	180
A9MQ82	BGAL_SALAR	121	TGCYSLTFCMDDDLWTEGQTRI IPDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSEYLOV	180
L0MA64	L0MA64 ENTBF	121	TGCYSLTFCVDNSWLAEQQTRI IPDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLTYLHA	180
AOA0J5KFW5	AOA0J5KFW5_PLUGE	121	TGCYSLTFNIDESWLQEGQTRI IPDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSVFLHA	180
AOA2P5GMD4	AOA2P5GMD4_9ENTR	121	TGCYSLTFNVDDSWLEDGQTRV IPDGVNSAFHLWCNGRWVGYGQDSRLPSEFDLSTYLOA	180
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	121	TGCYSLTFPSVDTDLWLSGQTRV IPDGVNSAFHLWCNGEWIGYAQDSRLPSEFDLSQALKP	180

P14288	BGAL_SULAC	1	-----	0
P22498	BGAL_SACS2	1	-----	0
P50388	BGAL_SACSH	1	-----	0
D9P208	D9P208 ACIS3	1	-----	0
AOA218P180	AOA218P180 THECE	1	-----	0
Q6KZ14	Q6KZ14 PICTO	1	-----	0
AOA0X1KSH7	AOA0X1KSH7_9THEM	1	-----	0
O52629	BGAL_PYRWO	1	-----	0
C6A195	C6A195 THESM	1	-----	0
SOATA3	SOATA3_FERAC	1	-----	0
P00722	BGAL_ECOLI	181	GENRLAVMVLWRS DGSYLEDDMWRMSGIFRDV3LLHKPTTQISDFHVATR FNDDFSRAV	240
Q3Z583	BGAL_SHISS	181	GKNRLAVMVLWRS DGSYLEDDMWRMSGIFRDV3LLHKPTTQISDFHVATR FNDDFSRAV	240
A6TI29	BGAL2_KLEP7	181	GENRLAVMVLWRS DGSYLEDDMWRMSGIFRDV3LLHKPTTQISDFHVATR FNDDFSRAV	240
Q3ZJB6	BGAL_SHIDS	181	GENRLAVMVLWRS DGSYLEDDMWRMSGIFRDV3LLHKPTTQIRDFHVATR FNDDFSRAV	240
D2TK51	D2TK51 CITRI	181	GENRLAVMVLWRC DGTYLEDDMWRMSGIFRDV3LLHKPTTQISDLRIATHFNDDFSRAE	240
A9MQ82	BGAL_SALAR	181	GENRLAVLVLWRS DGSYLEDDMWRMSGIFRDV3LLHKPTTQITDFHLHATHFNDDFTQAV	240
L0MA64	L0MA64 ENTBF	181	GENRLAVLVLWRS DGTYLEDDMWRMSGIFRDV3LLHKPSVQIADDFHINTHLSNENRAR	240
AOA0J5KFW5	AOA0J5KFW5_PLUGE	181	GENRLAVMVLWRS DGSYLEDDMWRMSGIFRDV3LLHKPTTQISDFHVATR FNDDFSRAV	240
AOA2P5GMD4	AOA2P5GMD4_9ENTR	181	GENRLAVLVLWRS DGTYLEDDMWRMSGIFRDV3LLHKPTTQIADDFHINTHLSNENRAR	240
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	181	GANRLAVLVLWRS DGTYLEDDMWRMSGIFRDV3LLHKPTTRLSDMRLT PHLNEDFTHGE	240

P14288	BGAL_SULAC	1	-----	0
P22498	BGAL_SACS2	1	-----	0
P50388	BGAL_SACSH	1	-----	0
D9P208	D9P208 ACIS3	1	-----	0
AOA218P180	AOA218P180 THECE	1	-----	0
Q6KZ14	Q6KZ14 PICTO	1	-----	0
AOA0X1KSH7	AOA0X1KSH7_9THEM	1	-----	0
O52629	BGAL_PYRWO	1	-----	0
C6A195	C6A195 THESM	1	-----	0
SOATA3	SOATA3_FERAC	1	-----	0
P00722	BGAL_ECOLI	1	MTMITDLSLAVVLQRRD WENPGVTQLNRLAAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60
Q3Z583	BGAL_SHISS	1	MIMITDLSLAVVLQRRD WENPGVTQLNRLAAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60
A6TI29	BGAL2_KLEP7	1	MTMITDLSLAVVLQRRD WENPGVTQLNRLAAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60
Q3ZJB6	BGAL_SHIDS	1	MTMITDLSLAVVLQRRD WENPGVTQLNRLAAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60
D2TK51	D2TK51 CITRI	1	MNLNTDLSLAAV LARRD WENPGVTQLNRLAAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60
A9MQ82	BGAL_SALAR	1	MTPERDLSLAAV LARRD WENPAVTFQNR LTAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60
L0MA64	L0MA64 ENTBF	1	MTSEQDLSLAAV LARRD WEEPGVTQRNRMAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60
AOA0J5KFW5	AOA0J5KFW5_PLUGE	1	MTMITDLSLAVVLQRRD WENPGVTQLNRLAAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60
AOA2P5GMD4	AOA2P5GMD4_9ENTR	1	MSTGHDSLAAV LARRD WENPGVTQRNRMAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	1	MTVKALSLAAV LARRD WENPAVTFQLHQLDAHPPFASWRNSEEARTDRFSQQLRSLNGEWR	60

P14288	BGAL_SULAC	1	-----	0
P22498	BGAL_SACS2	1	-----	0
P50388	BGAL_SACSH	1	-----	0
D9P208	D9P208 ACIS3	1	-----	0
AOA218P180	AOA218P180 THECE	1	-----	0
Q6KZ14	Q6KZ14 PICTO	1	-----	0
AOA0X1KSH7	AOA0X1KSH7_9THEM	1	-----	0
O52629	BGAL_PYRWO	1	-----	0
C6A195	C6A195 THESM	1	-----	0
SOATA3	SOATA3_FERAC	1	-----	0
P00722	BGAL_ECOLI	61	FAWFPAPEAVPESWLECDLPEADTVVPSNWMHGYDAP IYTVTYPIAVNPPVPTENP	120
Q3Z583	BGAL_SHISS	61	FAWFPAPEAVPESWLECDLPEADTVVPSNWMHGYDAP IYTVTYPIAVNPPVPTENP	120
A6TI29	BGAL2_KLEP7	61	FAWFPAPEAVPESWLECDLPEADTVVPSNWMHGYDAP IYTVTYPIAVNPPVPTENP	120
Q3ZJB6	BGAL_SHIDS	61	FVWFPAPEAVPESWLECDLPEADTVVPSNWMHGYDAP IYTVTYPIAVNPPVPTENP	120
D2TK51	D2TK51 CITRI	61	FAWFPSPEAVPESWLTDDLQADSVQLPSNWMHGYDAP IYTVTYPIAVNPPVPTENP	120
A9MQ82	BGAL_SALAR	61	FAWFSSPQAVPENWRLDLEAGTINVP SNWMHGYDAP IYTVTYPIAVNPPVPTENP	120
L0MA64	L0MA64 ENTBF	61	FRWYPSPEAVPESWLECDLPEADTVVPSNWMHGYDAP IYTVTYPIAVNPPVPTENP	120
AOA0J5KFW5	AOA0J5KFW5_PLUGE	61	FAWFPAPEAVPESWLECDLPEADTVVPSNWMHGYDAP IYTVTYPIAVNPPVPTENP	120
AOA2P5GMD4	AOA2P5GMD4_9ENTR	61	FAWFPSPEAVPESWLTDDLQADTVVPSNWMHGYDAP IYTVTYPIAVNPPVPTENP	120
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	61	FSYFPCPEAVPERWIDEDLTDADM IQVPSNWMHGYDAP IYTVTYPIAVNPPVPTENP	120

P14288	BGAL SULAC	1	-----		0
P22498	BGAL SACS2	1	-----		0
P50388	BGAL SACS	1	-----		0
D9P208	D9P208 ACIS3	1	-----		0
A0A218P180	A0A218P180 THECE	1	-----		0
Q6K214	Q6K214 PICTO	1	-----		0
A0A0X1KSH7	A0A0X1KSH7_9THEM	1	-----		0
O52629	BGAL PYRWO	1	-----		0
C6A195	C6A195 THESM	1	-----		0
SOATA3	SOATA3 FERAC	1	-----		0
P00722	BGAL ECOLI	361	HPLHGQVMDEQTMVQDILLMKQNNFNNAVRC	SHYPNHPLWYTLCDRYGLYVVDEANIETHG	420
Q32583	BGAL SHISS	361	HPLHGQVMDEQTMVQDILLMKQNNFNNAVRC	SHYPNHPLWYTLCDRYGLYVVDEANIETHG	420
A6T129	BGALZ KLEP7	361	HPLHGQVMDEQTMVQDILLMKQNNFNNAVRC	SHYPNHPLWYTLCDHYGLYVVDEANIETHG	420
Q32JB6	BGAL SHIDS	361	HPLHGQVMDEQTMVQDILLMKQNNFNNAVRC	SHYPNHPLWYTLCDRYGLYVVDEANIETHG	420
D2TK51	D2TK51 CITRI	361	HPQNGQVMDEATMVQDILLMKQNNFNNAVRC	SHYPNHPLWYTLCDRYGLYVVDEANIETHG	420
A9MQ82	BGAL SALAR	361	HPERGQVMVDYDTMVQDILLMKQNNFNNAVRC	SHYPNHPLWYTLCDRYGLYVVDEANIETHG	420
L0MA64	L0MA64 ENTBF	361	HPLNGQVMDEATMIQDILLMKQNNFNNAVRC	SHYPNHPLWYTLCDQYGLYVVDEANIETHG	420
A0A0J5KFW5	A0A0J5KFW5_PLUGE	361	HPLHGQVMDEQTMVQDILLMKQNNFNNAVRC	SHYPNHPLWYTLCDHYGLYVVDEANIETHG	420
A0A2P5GMD4	A0A2P5GMD4_9ENTR	361	HPERGQVMDEATMIRIDVLMKQNNFNNAVRC	SHYPNHPLWYTLCDQYGLYVVDEANIETHG	420
A0A4P8YKQ3	A0A4P8YKQ3_9ENTR	360	HPENGQVMDEVTMRRDILLMKQNNFNNAVRC	SHYPNHPLWYRLLCDYGLYVVDEANIETHG	419
P14288	BGAL SULAC	1	-----		0
P22498	BGAL SACS2	1	-----		0
P50388	BGAL SACS	1	-----		0
D9P208	D9P208 ACIS3	1	-----		0
A0A218P180	A0A218P180 THECE	1	-----		0
Q6K214	Q6K214 PICTO	1	-----		0
A0A0X1KSH7	A0A0X1KSH7_9THEM	1	-----		0
O52629	BGAL PYRWO	1	-----		0
C6A195	C6A195 THESM	1	-----		0
SOATA3	SOATA3 FERAC	1	-----		0
P00722	BGAL ECOLI	421	MVPMNRLTDDPRWLPAMSERVTRMVQRDRNHPSVI	IWSLGN SGHGANNHDALYRWIKSV	480
Q32583	BGAL SHISS	421	MVPMNRLTDDPRWLPAMSERVTRMVQRDRNHPSVI	IWSLGN SGHGANNHDALYRWIKSV	480
A6T129	BGALZ KLEP7	421	MVPMNRLTDDPRWLPAMSERVTRMVQRDRNHPSVI	IWSLGN SGHGANNHDALYRWIKSV	480
Q32JB6	BGAL SHIDS	421	MVPMNRLTDDPRWLPAMSERVTRMVQRDRNHPSVI	IWSLGN SGHGANNHDALYRWIKSV	480
D2TK51	D2TK51 CITRI	421	MTFMNRLSDDPDWLPAMSERVTRMVQRDRNHPSII	IWSLGN SGHGANNHDALYRWIKSV	480
A9MQ82	BGAL SALAR	421	MTFMNRLSDDPDWLPAMSERVTRMVQRDRNHPSII	IWSLGN SGHGANNHDALYRWIKSV	480
L0MA64	L0MA64 ENTBF	421	MVPMNRLTDDPDWLPAMSERVTRMVQRDRNHPSII	IWSLGN SGHGANNHDALYRWIKSV	480
A0A0J5KFW5	A0A0J5KFW5_PLUGE	421	MVPMNRLTDDPRWLPAMSERVTRMVQRDRNHPSVI	IWSLGN SGHGANNHDALYRWIKSV	480
A0A2P5GMD4	A0A2P5GMD4_9ENTR	421	MVPMNRLTDDPDWLPAMSERVTRMVQRDRNHPSII	IWSLGN SGHGANNHDALYRWIKSV	480
A0A4P8YKQ3	A0A4P8YKQ3_9ENTR	420	MVPMNRLSDDPRWLPAMSERVTRMVQRDRNHPSII	IWSLGN SGYGGNHDALYHWIKASD	479
P14288	BGAL SULAC	1	-----	M-----LSPFKGKFG-----	WSQSGFQ 18
P22498	BGAL SACS2	1	-----	M-----YSFPNSFRFG-----	WSQAGFQ 18
P50388	BGAL SACS	1	-----	M-----YSFPKFRFG-----	WSQAGFQ 18
D9P208	D9P208 ACIS3	1	-----	MA-----VTFPKDFLFG-----	WSQAGFQ 19
A0A218P180	A0A218P180 THECE	1	-----	M-----YKFPDFVFG-----	YSWSGFQ 18
Q6K214	Q6K214 PICTO	1	-----	M-----MLPKNFLG-----	FSLAGFQ 16
A0A0X1KSH7	A0A0X1KSH7_9THEM	1	-----	M-----MFPKDFLFG-----	ASMSGFQ 16
O52629	BGAL PYRWO	1	-----	M-----MPEKFLWG-----	VAQSGFQ 16
C6A195	C6A195 THESM	1	-----	M-----MEKEFLWG-----	VAQSGFQ 15
SOATA3	SOATA3 FERAC	1	-----	M-----MFG-----	TATSPFQ 10
P00722	BGAL ECOLI	481	PSRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SLPGETRPLILCEYA 540
Q32583	BGAL SHISS	481	PSRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SLPGETRPLILCEYA 540
A6T129	BGALZ KLEP7	481	PSRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SLPGETRPLILCEYA 540
Q32JB6	BGAL SHIDS	481	PSRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SLPGETRPLILCEYA 540
D2TK51	D2TK51 CITRI	481	PSRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SMPGEQRPLILCEYA 540
A9MQ82	BGAL SALAR	481	PSRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SLPGQRPLILCEYA 540
L0MA64	L0MA64 ENTBF	481	PTRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SLPGQRPLILCEYA 540
A0A0J5KFW5	A0A0J5KFW5_PLUGE	481	PSRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SLPGETRPLILCEYA 540
A0A2P5GMD4	A0A2P5GMD4_9ENTR	481	PSRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SLPGQRPLILCEYA 540
A0A4P8YKQ3	A0A4P8YKQ3_9ENTR	480	PSRPVQYEGGGADTTATDI	ICPMYARVDEDQPPFPAVPKWSIKKWL	SLPDESRLILCEYA 539
P14288	BGAL SULAC	19	SEMGT--PGSEDPNSDWHVWHDRENIVSQVVS	GD-----LPENGPYWGNYKRFH--	67
P22498	BGAL SACS2	19	SEMGT--PGSEDPNTDHYKWHVHDPENMAAGL	VSGD-----LPENGPYWGNYKTFH--	67
P50388	BGAL SACS	19	SEMGT--PGSEDPNTDHYKWHVHDPENISAGL	VSGD-----LPENGPYWGNYKTFH--	67
D9P208	D9P208 ACIS3	20	SEMGT--PGSEDPNSDHYAWVHDRENIAAGL	VSGD-----FPENGPYWGNYKRFH--	68
A0A218P180	A0A218P180 THECE	19	FEMGL--KGSVEPNSDWVWVHDMENIMTGLV	SGD-----LPENGPYWHLYSKDH--	67
Q6K214	Q6K214 PICTO	17	SEMGI--SD-PDSNSDWWLWVHDVFNITGLV	SGD-----LPENGIYWDLYKKNY--	64
A0A0X1KSH7	A0A0X1KSH7_9THEM	17	VEMGYA-KGDLDPNTDHFVWVREPENLINSV	VSGH-----LPEYGVYWNFFPIH--	66
O52629	BGAL PYRWO	17	FEMGDKLRRNIDTNTDWWHWVDRKTNIEKGL	VSGD-----LPEGINNYELYEKDH--	67
C6A195	C6A195 THESM	16	FEMGDRFNRLHDIRSDWVWVHDVFNITGLV	SGD-----LPEGINNYELYPIDH--	66
SOATA3	SOATA3 FERAC	11	VEMGRS-DNSTSSESDWYKWSHDSNIIQKTY	VSGD-----FPDDGPDFWNNYKRFI--	60
P00722	BGAL ECOLI	541	HAMGNSLGG---FAKYWQAFRQYPRLQGGFV	WDVVDQSLIKYDENG-NPWSAYGGDFD	595
Q32583	BGAL SHISS	541	HAMGNSLGG---FAKYWQAFRQYPRLQGGFV	WDVVDQSLIKYDENG-NPWSAYGGDFD	595
A6T129	BGALZ KLEP7	541	HAMGNSLGG---FAKYWQAFRQYPRLQGGFV	WDVVDQSLIKYDENG-NPWSAYGGDFD	595
Q32JB6	BGAL SHIDS	541	HAMGNSLGG---FAKYWQAFRQYPRLQGGFV	WDVVDQSLIKYDENG-NPWSAYGGDFD	595
D2TK51	D2TK51 CITRI	541	HAMGNSLGG---FAKYWQAFRQYPRLQGGFV	WDVVDQSLIKYDENG-NPWSAYGGDFD	595
A9MQ82	BGAL SALAR	541	HAMGNSLGG---FAKYWQAFRQYPRLQGGFV	WDVVDQSLIKYDADG-KPWSAYGGDFD	595
L0MA64	L0MA64 ENTBF	541	HAMGNSLGG---FAKYWQAFRQYPRLQGGFV	WDVVDQSLIKYDEEG-KPWSAYGGDFD	595
A0A0J5KFW5	A0A0J5KFW5_PLUGE	541	HAMGNSLGG---FAKYWQAFRQYPRLQGGFV	WDVVDQSLIKYDENG-NPWSAYGGDFD	595
A0A2P5GMD4	A0A2P5GMD4_9ENTR	541	HAMGNSLGG---FAKYWQAFRQYPRLQGGFV	WDVVDQSLIKYDENG-NPWSAYGGDFD	595
A0A4P8YKQ3	A0A4P8YKQ3_9ENTR	540	HAMGNSLGG---FRKYWDAFRQHPRLQGGFV	WDVVDQSLTRYDENG-QPWSAYGGDFD	594

P14288	BGAL SULAC	68	-DEAEKIGLNVRIN-----VWSRIFFRPLPKPEMGTGTDKEN--SPV	108
P22498	BGAL SACS2	68	-DNAQKMGLKIARLN-----VWSRIFFPNPLFRPNFD---ESK--QDV	105
P50388	BGAL SACSH	68	-DNAQKMGLKIARLN-----VWSRIFFPNPLFRPNFD---DSK--QDV	105
D9PZ08	D9PZ08 ACIS3	69	-DAAQAMGLTAARIG-----VWSRIFFRPTFDVKV--DAEVKG--DDV	107
AOA218P180	AOA218P180 THECE	68	-DMAEKLGMDAIRGG-----IEWARIFPEPTFDVVRVTERDE-E--GRI	107
Q6KZ14	Q6KZ14 PICTO	65	-GLAVQTMNAARIG-----VWSRIFFPKSTEEVKVME--DYKD--DDL	103
AOA0X1KSH7	AOA0X1KSH7 9THEM	67	-KLASDFGMVNLRTN-----IEWSRIFFRPTFDVVKV--EQTE--SDI	105
O52629	BGAL PYRWO	68	-EIARKLGLNAYRIG-----IEWSRIFFPWPTFDVVDYSYNSY--NLI	108
C6A195	C6A195 THESM	67	-LLAKKLGANAYSLN-----LEWSRIFFPCATYGLDVIDYELD--SN--GLI	106
SOATA3	SOATA3_FERAC	61	-DASIDMGNSTRIG-----IDWARIFFKTSSTESVDVAVASKN-EK--GDV	100
P00722	BGAL ECOLI	596	TFNDRQFCMGLVFADRTPHPALTEAKHQQQFFQFRL---SGRTIEVTSEYLFPHSDNEL	652
Q3Z583	BGAL SHISS	596	TFNDRQFCMGLVFADRTPHPALTEAKHQQQFFQFRL---SGRTIEVTSEYLFPHSDNEL	652
A6T129	BGALZ KLEP7	596	TFNDRQFCMGLVFADRTPHPALTEAKHQQQFFQFRL---SGRTIEVTSEYLFPHSDNEL	652
Q3ZJB6	BGAL SHIDS	596	TFNDRQFCMGLVFADRTPHPALTEAKHQQQFFQFRL---SGRTIEVTSEYLFPHSDNEL	652
D2TK51	D2TK51 CITRI	596	TFNDRQFCMGLVFADRTPHPSLYEAKHAQQFFQFRL---SGRTIEVTSEYLFPHSDNEI	655
A9MQ82	BGAL SÄLAR	596	TFNDRQFCMGLVFADRTPHPALYEAHVQFQFQFRL---SGRTIEVTSEYLFPHSDNEI	655
L0MA64	L0MA64_ENTBF	596	TFNDRQFCMGLVFADRTPHPALYEAHVQFQFQFRL---SGRTIEVTSEYLFPHSDNEI	655
AOA0J5KFW5	AOA0J5KFW5_PLUGE	596	TFNDRQFCMGLVFADRTPHPALTEAKHQQQFFQFRL---SGRTIEVTSEYLFPHSDNEL	652
AOA2P5GMD4	AOA2P5GMD4_9ENTR	596	TFNDRQFCMGLVFADRTPHPSLYEAKHEQQFFQFRLLAGKEYSIEVLSEYLFPHSDNEI	655
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	595	TFNDRQFCMGLVFADRTAHPPALYEAHVQFQFQFRL---RLEVTSEYLFRTTDRNR	652

P14288	BGAL SULAC	109	ISVDLINESKLREMDNYANHEALSHYRQIILEDLNRNGFHIVLNMHWHT----LPIWLHDP	163
P22498	BGAL SACS2	106	TEVEINENELKRLDEYANKDALNHYREIFKDLKSRGLYPTLNMYHWP-----LPLWLHDP	160
P50388	BGAL SACSH	106	TEVEINENELKRLDEYANKDALNHYREIFKDLKSRGLYPTLNMYHWP-----LPLWLHDP	160
D9PZ08	D9PZ08 ACIS3	108	LSVYVSEGALQLDLMANRDANHYREMFDLRSRGITPILNLYHWP-----LPLWLHDP	162
AOA218P180	AOA218P180 THECE	108	TSVDVPEASAIIELEKRLALEAHEHYRKYLSDWREKGVFLNLYHWP-----LPLWLHDP	162
Q6KZ14	Q6KZ14 PICTO	104	ISVDVNEGSLEKLDLRLANQAINRYMEIFMNIKENNMTLIVNLYHWP-----LPIYLHDP	158
AOA0X1KSH7	AOA0X1KSH7 9THEM	106	TSVQIDERALRELDLADKEAVEHYREIFSDMRKGLKVFVNLVHFT-----LPIWLHDP	160
O52629	BGAL PYRWO	109	EDVKITKDTLEELDEIANREVAAYRSVINSLSRSGFKVIVLNHFT-----LPYWLHDP	163
C6A195	C6A195 THESM	107	KEVKITKEVLEELNANIIEVEHYMSVLSNLKKGKVFITIVHYT-----HPLWLHDP	161
SOATA3	SOATA3_FERAC	101	YAMSFDPNVIQRMSIADNDIAVKHYVIMEYIKAKNLKLLTAYHWP-----LPLWLHDP	155
P00722	BGAL ECOLI	653	LHWMVA-----LDGKPLASGEVPLDVAPOGKQLIE-LPELPOPEASAGQLWLVTR	700
Q3Z583	BGAL SHISS	653	LHWSVA-----LDGKPLASGEMPLDVAPOGKQLIE-LPELPOPEASAGQLWLVTR	700
A6T129	BGALZ KLEP7	653	LHWMVA-----LDGKPLASGEVPLDVAPOGKQLIE-LPELPOPEASAGQLWLVTR	700
Q3ZJB6	BGAL SHIDS	653	LHWTVA-----LDGKPLASSEVPMNVAFPOGKQVIE-LPELPRIESTGQLWLVTR	700
D2TK51	D2TK51 CITRI	656	LHWSIT-----LDGNVPAAGEAALDIAPGQRQLIA-LPDIAAPDAGQLWLVTR	703
A9MQ82	BGAL SÄLAR	656	LRWMLA-----QENGLASGEVVLIDAPGQRQIIL-LPAFPPQFETAGQLWLVTR	703
L0MA64	L0MA64_ENTBF	656	LHWTLM-----RDGNQLAYGEMVLIDAPGQRQIIT-LPEVSTPQAFGQLWLVTR	703
AOA0J5KFW5	AOA0J5KFW5_PLUGE	653	LHWMVA-----LDGKPLASGEVPLDVAPOGKQLIE-LPELPOPEASAGQLWLVTR	700
AOA2P5GMD4	AOA2P5GMD4_9ENTR	656	LHWTFS-----LDGTSLASGEVALDIAPGQRQVIT-LPDIPAPQTAGQWLVTR	703
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	653	LHWTLK-----LDGESLACGEALLNAPQGGQIIT-L-DLDPVDRSQGLWLVNVE	699

P14288	BGAL SULAC	164	IRVRR-----GDFTE	172
P22498	BGAL SACS2	161	IRVRR-----GDFTE	169
P50388	BGAL SACSH	161	IRVRR-----GDLSE	169
D9PZ08	D9PZ08 ACIS3	163	IAIRR-----GNLS	171
AOA218P180	AOA218P180 THECE	163	IKVRR-----FGPDR	172
Q6KZ14	Q6KZ14 PICTO	159	IEARN-----SGLSN	168
AOA0X1KSH7	AOA0X1KSH7 9THEM	161	VAVHK-----RQET	169
O52629	BGAL PYRWO	164	IEARE-----RALTN	173
C6A195	C6A195 THESM	162	IESRE-----TNLKN	171
SOATA3	SOATA3_FERAC	156	VKCNQ-----DFANC	165
P00722	BGAL ECOLI	701	VVQP NATAWSEAGHISAQQWR LAENLSVTLPAAASHAIPHLTTSSEMDFCIELGNKRWQFN	760
Q3Z583	BGAL SHISS	701	VVQP NATAWSEAGHISAQQWR LAENLSVTLPAAASHAIPHLTTSSEMDFCIELGNKRWQFN	760
A6T129	BGALZ KLEP7	701	VVQP NATATSAAGHISAQQWR LAENLSVTLPASAPHAIPQLTTSETDFCIELDNKRWQFN	760
Q3ZJB6	BGAL SHIDS	701	VVQP NATAWSEAGHISAQQWR LAENLSVTLPASAPHAIPQLTTSETDFCIELDNKRWQFN	760
D2TK51	D2TK51 CITRI	704	VEQPQATAWSPAGHISAQQWPLAEKLSVTIPFRAAAAPQLKVENAAFEVVNVQRWQFC	763
A9MQ82	BGAL SÄLAR	704	VEQPLATWSAEGHISAQQWPLEEKLCVSKPTRASVAPVLTMRDGFPCVQGNLRWQFC	763
L0MA64	L0MA64_ENTBF	704	VEQP NATAWSEAGHISAQQWALEETLVANPPP LADEVPTLSANQREFMVTAQDKRWQFC	763
AOA0J5KFW5	AOA0J5KFW5_PLUGE	701	VVQP NATATWSAAGHISAQQWR LAENLSVTLPASAPHAIPQLTTSETDFCIELDNKRWQFN	760
AOA2P5GMD4	AOA2P5GMD4_9ENTR	704	VEQP PRASAWSAGHISAQQWALEETLSVQQA PRASDAPALATDDNTFCVTLGDKRWQFC	763
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	700	VKHIEATPWCDAGHLCAWDQWR LATRLSEPDADFVVGAAPELETNDAFIEVRHGAQHWRPFD	759

P14288	BGAL SULAC	173	GPTGWLNSRTVVEFARFSAIYVAWKLD-----DLASEYATMNPVNVVWGAGYAF--PRAGF	225
P22498	BGAL SACS2	170	GPSGLWLSRTVVEFARFSAIYVAWKLD-----DLVDEYSTMNPVNVVGLGYVG--VKSGF	222
P50388	BGAL SACSH	170	GPTGWLSTRVVEFARFSAIYVAWKLD-----DLVDEYSTMNPVNVVGLGYVG--VKSGF	222
D9PZ08	D9PZ08 ACIS3	172	APSGWLDRVTVEFAKFSAYVAWKLD-----DLVYMYSTFNPVNVVGLGYAA--VKSGF	224
AOA218P180	AOA218P180 THECE	173	APSGWLDRSVEFAKFAAFVAYHLN-----DFVDSWSTFNPVNVVYNGYGR--PNSGF	225
Q6KZ14	Q6KZ14 PICTO	169	KRNGWLNHKTVEFAYKYLAWKFS-----DVADMF SIMNEPNVFNBYNGYFN--VKSGF	221
AOA0X1KSH7	AOA0X1KSH7 9THEM	170	DKLGWASKRIVFAKFAAYVWVKFD-----DLVDMYSTFNEPNVVSQMGYV--SVSGF	222
O52629	BGAL PYRWO	174	KRNGWVNPRTVEFAKFAAYIAYKFG-----DIVDMWSTFNPVNVVGLGYLA--PYSGF	226
C6A195	C6A195 THESM	172	ERNGWVNRQSIIEFTKFAAYLAYKFG-----HLVDMWSTFNEPNVIVGLGYLA--PYSGF	224
SOATA3	SOATA3_FERAC	166	REKGWGDKATVEFGKYYIYIKNFH-----RYVDIWNLTNEPNIIAINGNYYVGNLEGF	219
P00722	BGAL ECOLI	761	RQSGFLSQMWIGDKKQLLPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYQAE	820
Q3Z583	BGAL SHISS	761	RQSGFLSQMWIGGAEKQLLPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYQAE	820
A6T129	BGALZ KLEP7	761	RQSGFLSQMWIGDKKQLLPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYQAE	820
Q3ZJB6	BGAL SHIDS	761	RQSGFLSQMWIGDEKQLLPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYQAE	820
D2TK51	D2TK51 CITRI	764	RQSGTLSQYWIADAQLLPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYRAE	823
A9MQ82	BGAL SÄLAR	764	RQGGWLTFQWRDDEAQLLPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYCAE	823
L0MA64	L0MA64_ENTBF	764	RQGGWLTFQWLGDQQLLPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYCAE	823
AOA0J5KFW5	AOA0J5KFW5_PLUGE	761	RQSGFLSQMWIGDKKQLLPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYQAE	820
AOA2P5GMD4	AOA2P5GMD4_9ENTR	764	RQGGWLTFQWRNDEAQLLPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYRME	823
AOA4P8YKQ3	AOA4P8YKQ3_9ENTR	760	RESGLLCQWFSHGQPTLATPLRQDTPRAPLNDIGVSEATRIPDPAWVERWKAAGHYEMS	819

P14288	BGAL_SULAC	402	YYLVSHIYQVHRAINNEG-VDVRGYLHWSLADNYEWSSGFSMRFGLLKVDYLTKRLYWR-P	459
P22498	BGAL_SACSS2	400	YYLVSHVYQVHRAINSRG-ADVRGYLHWSLADNYEWASGFSMRFGLLKVDYNTKRLYWR-P	457
P50388	BGAL_SACSH	400	YYLVSHVYQVHRAINSRG-ADVRGYLHWSLADNYEWASGFSMRFGLLKVDYGTKRLYWR-P	457
D9P208	D9P208_ACIS3	398	YYLVSHVYQVHRAIQDG-VNVIYGLHWSLADNYEWASGFSKRFGLLMVDYSTKRLHWR-P	455
A0A218F180	A0A218F180_THECE	401	YYLVSHLRAIHSALTEAG-ADIRGYLHWSLTDNYEWAKGFQMKFGLLKVDWESKRRYIR-P	458
Q6KZ14	Q6KZ14_PICTO	410	RYLVSHIKSVEKALSMG-MDIRGYLHWSLTDNYEWASGFSMKFGLYGIDLNNKKIQHR-P	467
A0A0X1KSH7	A0A0X1KSH7_9THEM	404	YHLVAHMYAVEFAVEEG-LNVKGYLHWSLVDNYEWAKGYHMRFGLAETNYQTKSYNPR-P	461
O52629	BGAL_PYRWO	427	YYIASHIKMTEKAFEDG-YEVKGYFHWALTDNFEWALGFRMRFGLYEVNLTKERIPREK	485
C6A195	C6A195_THESM	415	YFIAASHIDYIEKALIEEG-FDVRGYFHWALTDNYEWAMGFRMRFGLYVVDMITKERIPRKE	473
SOATA3	SOATA3_FERAC	396	EFIEKHLIELHKAIKEDYIPVKGYFHWSLVDNYEWARGYKDKFGLYKINNGEF---VKTE	452
P00722	BGAL_ECOLI	968	LMETSHRH-LLHAEEGTWNLDGF-HMGIGGDDS*SPSVSAEFQLSAGRYHYQLVWCQ-K	1024
Q3Z583	BGAL_SHISS	968	LMETSHRH-LLHAEEGTWNLDGF-HMGIGGDDS*SPSVSAEFQLSAGSYHYQLVWCQ-K	1024
A6TI29	BGAL2_KLEP7	968	LMETSHRH-LLHAEEGTWNLDGF-HMGIGGDDS*SPSVSAEFQLSAGSYHYQLLWCQ-K	1024
Q3ZJB6	BGAL_SHIDS	968	LMETSHRH-LLHAEEGTWNLDGF-HMGIGGDDS*SPSVSAEFQLSAGRYHYQLVWCQ-K	1024
D2TK51	D2TK51_CITRI	971	LMETSHRH-LLRENGTWNLDGF-HMGVGGDDS*SPSVSPEYQLSAGRYHYQLVWCQ-K	1027
A9MQ82	BGAL_SALAR	971	LMETSHRH-LLQAEAGTWNLDGF-HMGVGGDDS*SPSVSPEFQLSARHYHYQLIWK---	1025
L0MA64	L0MA64_ENTBF	971	LMETSHRH-LLQAEAGTWNLDGF-HMGVGGDDS*SPSVSPEFQLSARHYHYQLIWK---	1025
A0A0J5KFW5	A0A0J5KFW5_PLUGE	968	LMETSHRH-LLHAEEGTWNLDGF-HMGIGGDDS*SPSVSAEFQLSAGSYHYQLLWCQ-K	1024
A0A2P5GMD4	A0A2P5GMD4_9ENTR	971	LMETSHRH-LLQAEAGTWNLDGF-HMGVGGDDS*SPSVSPEFQLSARHYNYQLIWE---	1025
A0A4P8YKQ3	A0A4P8YKQ3_9ENTR	967	LMETSHRH-LLKEEPPGSLNLDGF-HMGVGGDDS*SPSVSAEFQLSARHYHYAVSWAR-R	1023
* : * : * : . : . : . : . : . : *				
P14288	BGAL_SULAC	460	SALVYREITRSNGIPEELEHLNRVPPKPLRH---	491
P22498	BGAL_SACSS2	458	SALVYREIATNGAITDEIEHLNSVPPVKPLRH---	489
P50388	BGAL_SACSH	458	SALVYREIATNGGITDEIEHLNTVPPKPLRH---	489
D9P208	D9P208_ACIS3	456	SALFYREIATNGSRAITDEIEHLNSVPPKPLRGLSPGHR	490
A0A218F180	A0A218F180_THECE	459	SALVFKEIATQKALPEELSHLSDLRPLLQD-----	488
Q6KZ14	Q6KZ14_PICTO	468	SALVFKEIANANGVPEEFVMAQQHNS-----	495
A0A0X1KSH7	A0A0X1KSH7_9THEM	462	SMYIFREIVKELSTKFRSYLSSPYQIWRQK-G--	493
O52629	BGAL_PYRWO	486	SVSIFREIVANNGVTKKIEEELLRG-----	510
C6A195	C6A195_THESM	474	SVGVYREIENDGITDRIRKEYLIRGDMI-----	501
SOATA3	SOATA3_FERAC	453	ASEFYSKICHDRGVEDEDFKTY-----	474
P00722	BGAL_ECOLI	1025	-----	1024
Q3Z583	BGAL_SHISS	1025	-----	1024
A6TI29	BGAL2_KLEP7	1025	-----	1024
Q3ZJB6	BGAL_SHIDS	1025	-----	1024
D2TK51	D2TK51_CITRI	1028	-----	1027
A9MQ82	BGAL_SALAR	1026	-----	1025
L0MA64	L0MA64_ENTBF	1026	-----	1025
A0A0J5KFW5	A0A0J5KFW5_PLUGE	1025	-----	1024
A0A2P5GMD4	A0A2P5GMD4_9ENTR	1026	-----	1025
A0A4P8YKQ3	A0A4P8YKQ3_9ENTR	1024	-----	1023

Figure .3: Explain the alignment between the different groups mesophilic and thermophilic.

3.4 Secondary Structure Prediction

Analysis of secondary structure prediction from CFSSP on beta-galactosidase had shown that there are 119 amino acid residues involved in the formation of helix, 109 amino acids for extended strands (beta sheet) formation and 263 amino acid residues in the formation of Random Coil, which consists 24.24%, 22.20% and 53.56% amino acid residues respectively).

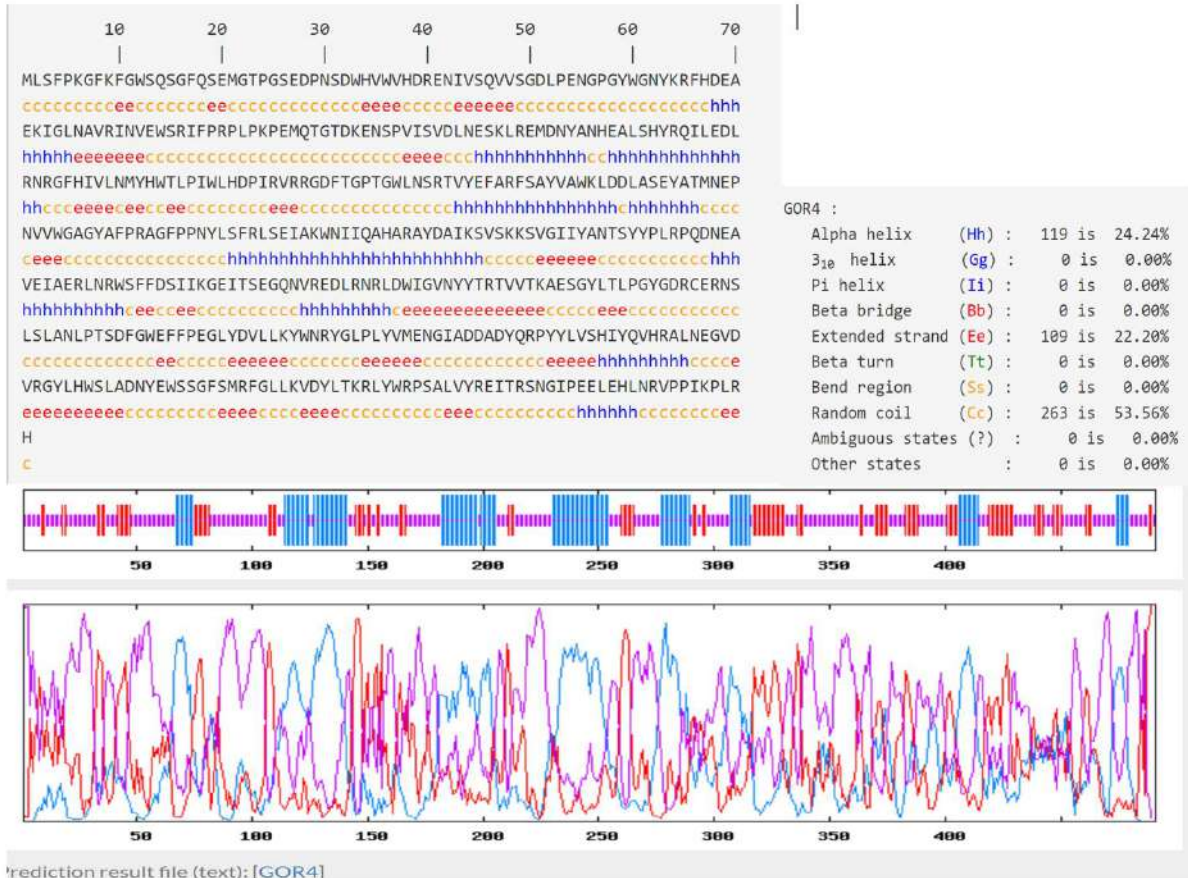


Figure .4: Secondary structure prediction of Beta-galactosidase

3.4 Sequence analysis

The alignment results of thermophilic types, meticulously analyzed within MEGAX, have unveiled a region of exceptional conservation. This region, thoughtfully identified through Weblogo analysis (<https://weblogo.berkeley.edu/logo.cgi>), serves as a prominent testament to the shared functional attributes among these thermophilic enzyme sequences. It signifies a segment of the enzyme sequences that play specific roles critical to their biological function, a commonality that underscores their adaptability to high-temperature environments and highlights the evolutionary convergence of these thermophilic variants towards similar enzymatic functionalities. This observation opens intriguing avenues for further exploration and understanding of the unique traits and capabilities of these thermophilic enzymes.

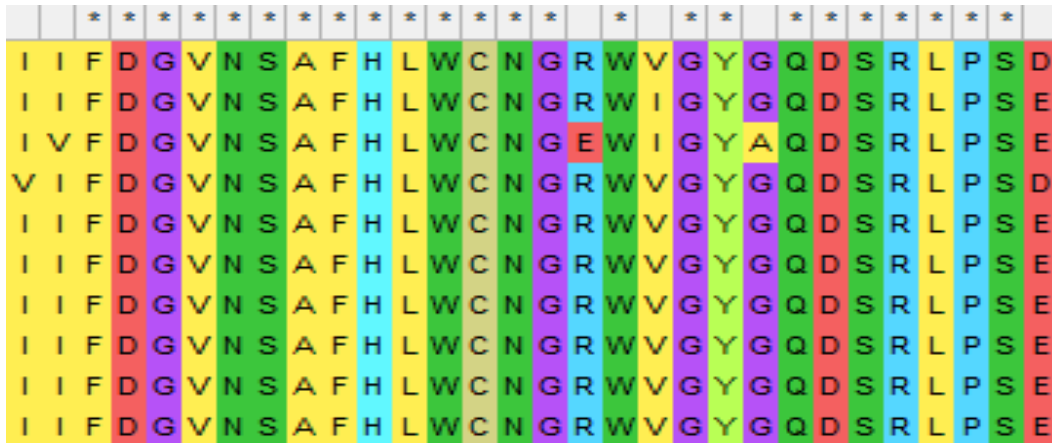


Figure .5: Shows the sequence alignment of thermophilic types.

meticulously conducted through the robust software MEGAX, which offers a comprehensive depiction of the evolutionary relationships and shared genetic characteristics among these thermophilic organisms. This alignment serves as a foundational framework for elucidating the conserved and divergent elements within their genetic sequences, shedding light on the molecular adaptations that enable these organisms to thrive in high-temperature environments.

Through this alignment, we gain valuable insights into the conserved regions and motifs, which are indicative of critical functional roles common to these thermophilic species. Such conserved elements are essential for their survival and growth under extreme thermal conditions, underscoring the remarkable evolutionary strategies employed by thermophiles to maintain their biological functions at elevated temperatures.



Figure .6: The conserved domain sequence alignment logo.

The construction of a phylogenetic tree using MEGAX serves as a powerful tool to elucidate the evolutionary relationships among different species, specifically focusing on the thermophilic and mesophilic types. The phylogenetic analysis spans from thermophilic organisms, commencing with thermo types, and extending through to *Thermococcus sibiricus*. In parallel, the analysis includes mesophilic organisms, with the starting point at *Izhakiella* sp. and concluding at *Pluralibacter gergoviae*. This comprehensive phylogenetic tree provides valuable insights into the evolutionary divergence and relatedness among these diverse microbial species, contributing to our understanding of their evolutionary history and ecological niches.

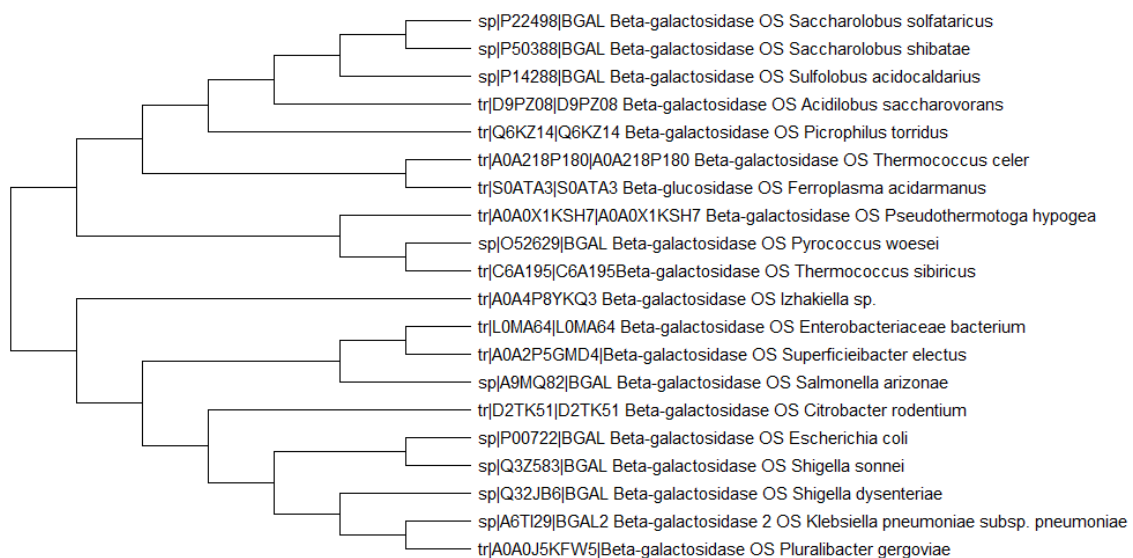


Figure .7: The phylogenetic tree, meticulously constructed using the robust software MEGAX.

The active site of beta-galactosidase from *Sulfolobus acidocaldarius*, as identified and documented by UniProt, is characterized by specific amino acid residues located at positions 209 and 389 within the enzyme's primary sequence. These particular amino acid positions play a critical role in substrate binding, catalysis, and overall enzymatic activity. The precise arrangement and functional significance of these residues underscore their pivotal contribution to the enzyme's active site, facilitating its biological function with precision and efficiency.

```
      110      120      290      300
TDKENSPIVIS VDLNESKLRE VEIAERLNRW SFFDSIIKGE
      160      170      340      350
MYHWTLPDWL HDPIRVRGGD KAESGYLTLP GYGDRCERNL
      210      220      390      400
ASEYATMNEP NVVWGAGYAF YGLPLYVMEN GIADDADYQR
      260      270      440      450
DAIKSVSKKS VGIIYANTSY DNYEWSSGFS MRFGLLKVDY
      310      320
---
```

Figure .8: Illustrates the active sites of *beta-galactosidase* from *Sulfolobus acidocaldarius*.

3.5 Three-dimensional (3D) structure.

The Swiss-model modelling tool played a pivotal role in the prediction of the three-dimensional (3D) structure for the given protein sequence. Within the tool's repertoire, numerous potential template structures closely matching the query protein sequence were available for consideration. From this array of choices, meticulous selection led us to opt for the most akin 3D structures, namely, those of β -Glycosidase from *Sulfolobus solfataricus* and β -Glycosidase from *Acidilobus saccharovorans*. This judicious choice of templates ensured that our modelling efforts were anchored in the most relevant and structurally analogous frameworks, enhancing the accuracy and reliability of our 3D structure prediction.

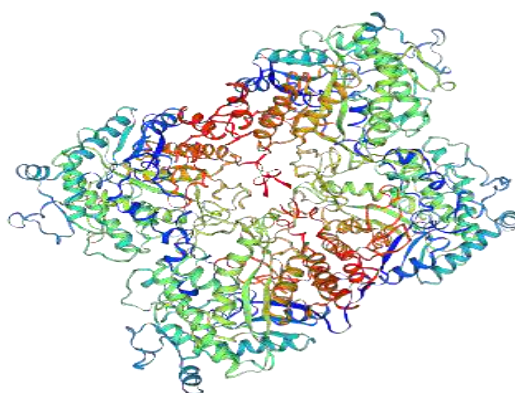


Figure .9: Shows modeled 3D structure of *Beta-galactosidase* in *Sulfolobus acidocaldarius*.

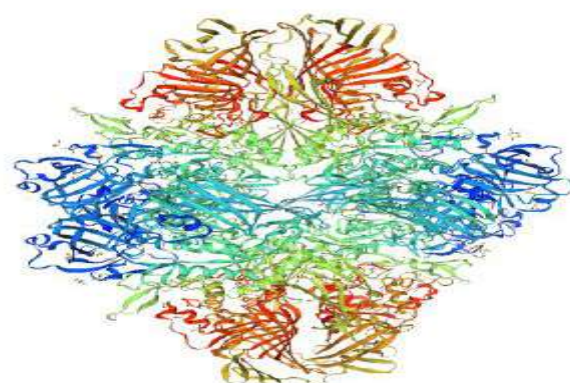


Figure .10: Shows modelled 3D structure of beta-galactosidase in *Escherichia coli* (strain k12)

Moreover, the computational prediction of local structural similarity in the building mode has revealed a noteworthy pattern of conserved residues. Specifically, this analysis has illuminated a region characterized by an escalating and continuous increase in conserved residues, spanning from position 120 to 300 within the protein sequence. Furthermore, this conserved motif seamlessly extends, persisting across . positions 350 through 480. This observed pattern of residue conservation underscores the functional significance of this specific protein segment and suggests its potential involvement in critical molecular interactions or structural stability within the protein's tertiary structure.

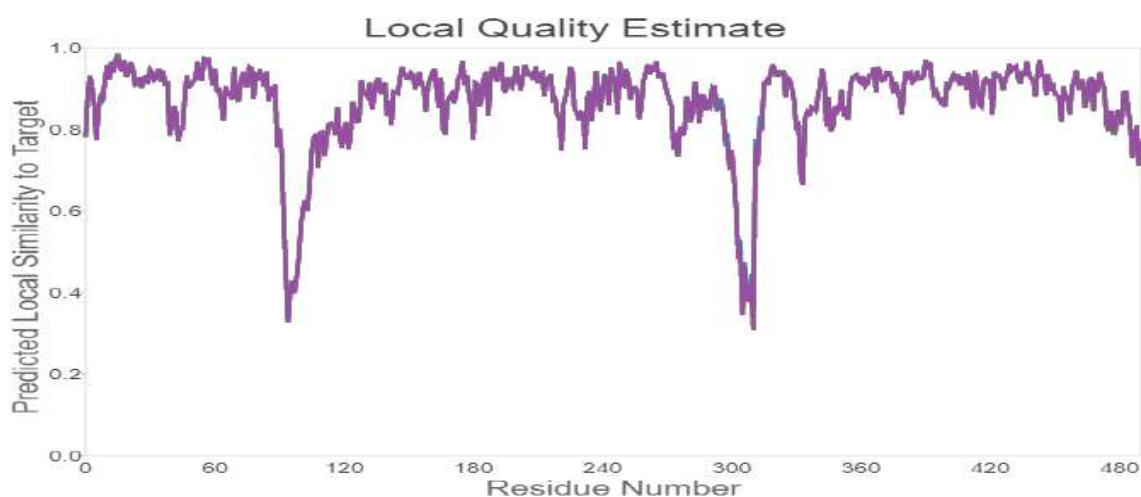


Figure .11: Shows the predicted local similarity of residue number

The predicted local similarity of residue number made by the Swiss model through the build model of *Sulfolobus acidocaldarius* 3D structure modelling. The chains for the build model showed similarity in different residues with almost more similarities between them with tiny differences.

4. Analysis model 3D

The Swiss Model structure analysis has been instrumental in identifying critical amino acid residues essential for our alignment, particularly those constituting the active site. Specifically, our focus has honed in on amino acids positioned at residues 209 and 389 within the enzyme's structure. Furthermore, the three-dimensional structural representation has revealed a noteworthy insight: the active sites exhibit close spatial proximity to each other. This spatial arrangement hints at potential cooperative interactions between these active sites, which may have significant implications for the enzyme's catalytic function and substrate binding.

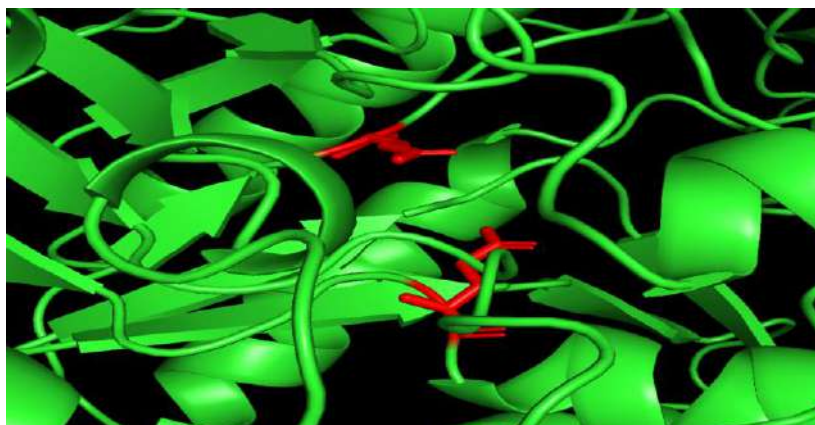


Figure .13: Shows the critical active site of the enzyme

The critical active site of the enzyme is defined by the specific amino acid residue located at position 209 within its primary sequence. This particular amino acid residue holds pivotal significance in the enzyme's catalytic function, substrate binding, and interaction with its molecular targets. The precision and strategic placement of this residue at position 209 highlight its essential role in the enzyme's overall activity, underscoring its contribution to the biological and biochemical functions of the protein.

6. Discussion

The comparative analysis of β -Galactosidase enzymes from *Sulfolobus acidocaldarius* and *Escherichia coli* yields compelling insights that underscore the validity and significance of the findings. The research discovered that *Sulfolobus* isolates possess thermostable enzymes that can effectively break down various glycosidic substances [10]. The B-galactosidase from *Sulfolobus acidocaldarius* is being investigated for its thermostability and organic solvent tolerance to hydrolyze lactose in dairy products, along with other *Sulfolobus* enzymes. The study delves into the implications of our results, emphasizing the rationale behind our methodology and the broader implications for biotechnology and industrial

applications. The modelling of the 3D structure of β -Galactosidase from *Sulfolobus acidocaldarius* provides crucial insights into the enzyme's structural adaptation to extreme conditions. Notably, *S. acidocaldarius* thrives in high-temperature, low-pH environments, and its β -Galactosidase enzyme demonstrates exceptional thermostability [11]. By comparing this thermostable structure with the mesophilic counterpart from *Escherichia coli*, we shed light on the structural basis for the extraordinary stability observed in *S. acidocaldarius*. The results reveal distinct structural features in *S. acidocaldarius*' β -Galactosidase that contribute to its thermostability. These may include enhanced structural rigidity, increased hydrophobic interactions, and unique amino acid compositions. Our findings align with the well-established principle that extremophiles often possess specialised structural adaptations that allow them to thrive in harsh environments. This not only validates the biological significance of our structural model but also opens avenues for the engineering of thermostable enzymes for various industrial processes. The industrial utility of β -Galactosidase, particularly in lactose-containing fluid processing, is well-recognized [12]. The results emphasise the potential advantages of harnessing the thermostable β -Galactosidase from *S. acidocaldarius* for such applications. The structural insights gained through our modelling provide a rational basis for the development of more robust and efficient enzymatic processes. The ability to operate at elevated temperatures can lead to improved reaction kinetics and reduced contamination risks, enhancing the overall efficiency of lactose hydrolysis in industrial settings. Furthermore, the thermostable β -Galactosidase enzyme may find applications in diverse sectors, including the dairy industry, pharmaceuticals, and the production of sweet syrups for confectionery and soft drinks. The enhanced stability and activity observed in this enzyme make it an attractive candidate for these applications, potentially revolutionizing processes and products in these fields. The results also have broader implications for environmental sustainability. By offering a thermostable enzyme alternative for lactose breakdown, we contribute to the reduction of lactose-containing waste in the food industry. This has positive environmental consequences by mitigating the environmental footprint associated with lactose disposal [13].

6. Conclusion:

Finally, our study opens the door to further investigations. Future research could involve experimental validation of the predicted structural features and functional properties of the *S. acidocaldarius* β -Galactosidase. Additionally, directed evolution or protein engineering techniques can be employed to enhance its performance even further for specific industrial applications.

In conclusion, our comprehensive analysis of β -Galactosidase enzymes from *Sulfolobus acidocaldarius* and *Escherichia coli* yields compelling evidence that supports the robustness of our results. The structural insights gained provide a solid foundation for the rational design and utilization of thermostable enzymes in biotechnology, industrial processes, and environmental sustainability efforts. These findings not only advance our understanding of extremophile enzymes but also hold the potential to drive innovation in various sectors, ultimately benefiting society at large.

REFERENCES

- [1] -βS. Ozarslan and T. Taş, “Treatment of lactose by fermentation: Production process on vol. 1. *Int. Food Res. J.* galactosidase using *Kluyveromyces marxianus* isolated from kefir grains,” 2022, no. 19.
- [2] galactosidase and its potential β A. Shafi and Q. Husain, “Structural and functional insights of 2347–2353pp. 2023 Elsevier, *Glycoside Hydrolases applications*,” in
- [3] in as a therapeutic target J. Gao, Y. Hao, X. Piao, and X. Gu, “Aldehyde dehydrogenase *Int. J. translational modifications deserve more attention*,” -related diseases: post-oxidative stress 2022, 2682p. no. 23 vol. 1. *Mol. Sci*
- [4] value products by -M. Bilal and H. M. N. Iqbal, “Sustainable bioconversion of food waste into high *Food Res. A review*,” –economy challenges and opportunities-zed enzymes to meet bioimmobili 2019, 240–246pp. 123 vol. 1. *Int*
- [5] T. Capcanari, A. Chirsanova, E. Covaliov, and R. Siminiuc, “Development of Lactose Free Yogurt 2021, 1130–1136pp. 11 no. 12 vol. 1. *Food Nutr. Sci* Technology for Personalized Nutrition,”
- [6] galactosidase in lactose -βA. Nivetha and V. Mohanasrinivasan, “Mini review on role of 2017 IOP Publishing, *IOP Conference Series: Materials Science and Engineering* intolerance,” in 22046p.
- [7] *The Physicochemical Characterization of Microvesicles Secreted by Sulfolobus* A. P. Bonanno 2019 Temple University, *acidocaldarius*
- [8] J. K. Lee, C. Park, and O. J. Kwon, “Experimental study of moving throat plug in a shock tunnel,” 2007, 10–11pp. 10, 1007doi: 2010, 442–431pp. 4 no. 20 vol. 1. *Shock Waves*
- [9] D. A. Prangishvili, R. P. Vashakidze, M. G. Chelidze, and I. Y. Gabriadze, “A restriction *FEBS* endonuclease Sual from the thermoacidophilic archaeobacterium *Sulfolobus acidocaldarius*,” 1980, 60–67pp. 1 no. 192 ol. v 1. *Lett*
- [10] G. Chong, “Certain, but not all, tetraether lipids from the thermoacidophilic -A. Bonanno and P. L. archaeon *Sulfolobus acidocaldarius* can form black lipid membranes with remarkable stability and 22 vol. 1. *Int. J. Mol. Sci* sensitivity,” + activity with unusually high Ca exhibiting mthk channel 2021, 12941p. 23 no.
- [11] galactosidase from -β elbatsomreht a gnidocne eneg a fo noisserpxe suogoloreteH“ 1. *et al*/T. Yuan 2008, 348–353pp. 30 ol. v 1. *Biotechnol. Lett* Alicyclobacillus acidocaldarius,”
- [12] P. Katrolia, M. Zhang, Q. Yan, Z. Jiang, C. Song, and L. Li, “Characterisation of a thermostable galactosidase (BgalC) family from *Thermotoga maritima* showing efficient lactose -β 4 family

.۲۰۱۱، ۶۲۱-۶۱۱ pp. ،۲no. ،۱۲۰vol. ،*Food Chemhydrolysis*،”

[۱۳] P. Stanchev, V. Vasilaki, D. Egas, J. Colon, S. Ponsá, and E. Katsou, “Multilevel environmental assessment of the anaerobic treatment of dairy processing effluents in the context of circular

.۲۰۲۰، ۱۲۱۱۳۹p. ،۲۶۱vol. ،*J. Clean. Prodeconomy*،”