TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

STRUCTURAL STUDIES OF ENZYMES FROM THERMOPHILIC ORGANISMS AND IDENTIFICATION OF THEIR THERMOSTABILITY FACTORS

Mohsin Imran

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 738 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2025

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

STRUCTURAL STUDIES OF ENZYMES FROM THERMOPHILIC ORGANISMS AND IDENTIFICATION OF THEIR THERMOSTABILITY FACTORS

Mohsin Imran

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AI OSA – TOM. 738 | ASTRONOMICA – CHEMICA – PHYSICA – MATHEMATICA | TURKU 2025

University of Turku

Faculty of Technology
Department of Life Technologies
Biochemistry
Doctoral Programme in Technology

Supervised by

Docent Anastassios Papageorgiou
Turku Bioscience Centre
University of Turku and
Åbo Akademi University
Turku, Finland

Reviewed by

Prof. Panos Soultanas
School of Chemistry
University of Nottingham
UK

Docent Tommi Kajander
Institute of Biotechnology
University of Helsinki
Finland

Opponent

Prof. Matthew Groves
Faculty of Science and Engineering
University of Groningen
Netherlands

ISBN 978-952-02-0183-8 (Print)
ISBN 978-952-02-0184-5 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)
Painosalama, Turku, Finland 2025

UNIVERSITY OF TURKU
Faculty of Technology
Department of Life Technologies
Biochemistry
MOHSIN IMRAN: Structural studies of enzymes from thermophilic organisms and identification of their thermostability factors
Doctoral Dissertation, 146 pp.
Doctoral Programme in Technology
June 2025

ABSTRACT

Thermophilic enzymes are characterized by their ability to resist temperature while maintaining their catalytic efficiency. They are mostly found in organisms, such as bacteria or fungi, that reside in natural habitats of high temperatures, usually between 50–80 °C. Thermophilic enzymes are of utmost importance for industrial and research applications and understanding their properties is key for their further improvement by genetic engineering approaches.

This study investigated four thermophilic enzymes, three from fungal sources and one from a bacterial source. X-ray crystallography technique was employed to extract the three-dimensional (3D) structural details of the enzymes. A brief description of the investigated thermophilic enzymes is given below.

The crystal structure of a β-glucosidase from a thermophilic fungus (*Chaetomium thermophilum*), referred here as *Ct*BGL, was determined to a resolution of 2.99 Å. *Ct*BGL structure revealed the nucleophilic (Asp287) and acid/base (Glu517) catalytic residues. The structure of *Ct*BGL showed a three-domain architecture as in other β-glucosidases but with variations in loops and linker regions. Glycosylation and charged residues were suggested as the potential thermostability contributing factors in *Ct*BGL.

The crystal structure of an Auxiliary Activity 9 Lytic Polysaccharide Mono Oxygenase produced by the thermophilic fungus *Thermoascus aurantiacus*, referred to here as native *Ta*AA9A (n*Ta*AA9A), was determined to a resolution of 1.36 Å. n*Ta*AA9A was found to be active in producing C1- and C4-oxidized products from cellulose and xylan. The n*Ta*AA9A structure was compared with the recombinant form of the enzyme expressed in *Aspergillus oryzae* (r*Ta*AA9A). The compared structures exhibited a root mean square deviation (RMSD) of 0.43 Å after structural superposition, suggesting subtle changes. Differences were observed in surface loops and glycosylation sites. n*Ta*AA9A revealed higher degree of glycosylation than r*Ta*AA9A. In n*Ta*AA9A, Asn138 residue was found glycosylated with at least two NAG molecules. Glycosylation and electrostatic interactions were suggested as possible thermostability contributing factors.

The crystal structure of a Cu,Zn superoxide dismutase from a thermophilic fungus *Chaetomium thermophilum* (*Ct*SOD) was determined to a resolution of 1.56 Å. *Ct*SOD was crystallized with eight molecules (A-H) in the crystallographic asymmetric unit, resulting in eight distinct interfaces. Zn^{2+} was present in all

subunits, but Cu^{2+} in 4 subunits only (C, D, E, and F). The active-site pocket region, along with the copper- and zinc-binding sites, were found highly conserved. A higher degree of oligomerization and an elevated contribution of polar residues in *Ct*SOD were suggested as thermostability contributing factors.

The crystal structure of a carbonic anhydrase from a thermophilic bacterium *Caloramator australicus* (γ-*Ca*CA) was determined to 1.11 Å resolution. This is the highest resolution thus far for a γ-family carbonic anhydrase. The enzyme was crystallized with 3 molecules in the asymmetric unit. The active site of each molecule was found at the interface of two neighbouring molecules. The γ-*Ca*CA structure was found highly conserved, but differences were noticed in loop regions compared to other CAs. Charged residues and hydrophobic clusters were suggested as possible thermostability contributing factors in γ-*Ca*CA.

Thermozymes hold a promising future for bio-economy and green chemistry. The results presented here could offer new ideas to develop sustainable and environment-friendly solutions for a better future.

TURUN YLIOPISTO
Teknillinen tiedekunta
Bioteknologian laiton
Biokemia
MOHSIN IMRAN: Structural studies of enzymes from thermophilic organisms and identification of their thermostability factors
Väitöskirja, 146 s.
Teknologian tohtoriohjelma
Kesäkuu 2025

TIIVISTELMÄ

Termofiilisille entsyymeille on tunnusomaista niiden kyky kestää korkeita lämpötiloja säilyttäen samalla katalyyttisen tehokkuutensa. Niitä esiintyy enimmäkseen luonnonvaraisissa organismeissa, kuten bakteereissa tai sienissä, jotka elävät luonnollisissa elinympäristöissä, joissa lämpötila on korkea, yleensä 50–80 °C. Termofiiliset entsyymit ovat äärimmäisen tärkeitä teollisuuden ja tutkimuksen sovelluksille, ja niiden ominaisuuksien ymmärtäminen on avainasemassa niiden edelleen parantamiseksi geenitekniikan lähestymistapojen avulla.

Tässä tutkimuksessa tutkittiin neljä termofiilistä entsyymiä, kolme sienilähteistä ja yksi bakteerilähteestä. Röntgenkristallografiatekniikkaa käytettiin entsyymien kolmiulotteisten (3D) rakenteellisten yksityiskohtien erottamiseen. Tässä esitetään lyhyt kuvaus tutkituista termofiilisistä entsyymeistä.

Termofiilisen sienen (*Chaetomium thermophilum*), jota kutsutaan tässä nimellä *Ct*BGL, β-glukosidaasin kiderakenne määritettiin 2,99 Å:n resoluutiolla. *Ct*BGL-rakenne paljasti nukleofiiliset (Asp287) ja happo/emäs (Glu517) katalyyttiset aminohapot. *Ct*BGL:n rakenne muodostui kolmesta domeenista, kuten muissa β-glukosidaaseissa, mutta rakennevaihtelua esiintyi silmukoissa ja linkkerialueilla. Glykosylaatio ja varaukselliset aminohapporyhmät voivat olla lämpöstabiilisuutta edistäviä tekijöitä *Ct*BGL:ssä.

Termofiilisestä sienestä (*Thermoascus aurantiacus*) peräisin olevan natiivin AA9 LPMO:n kiderakenne, jota kutsutaan tässä natiiviksi *Ta*AA9A:ksi (n*Ta*AA9A), määritettiin 1,36 Å:n resoluutiolla. n*Ta*AA9A:n havaittiin olevan aktiivinen C1- ja C4-hapettuneiden tuotteiden tuottamisessa selluloosasta ja ksylaanista. n*Ta*AA9A-rakennetta verrattiin Aspergillus oryzaessa (r*Ta*AA9A) ilmennetyn entsyymin rekombinantti-muotoon. Näiden kahden rakenteen laskettu neliökeskiarvon poikkeama (RMSD) oli 0,43 Å rakenteellisen superposition jälkeen, mikä viittaa hienovaraisiin muutoksiin. Proteiinin uloimmissa silmukoissa ja glykosylaatiokohdissa havaittiin eroja. n*Ta*AA9A:n rakenne paljasti korkeamman glykosylaatioasteen kuin r*Ta*AA9A:n rakenne. n*Ta*AA9A:ssa Asn138-tähde havaittiin glykosyloituneena vähintään kahdella NAG-molekyylillä. Glykosylaatio ja sähköstaattisia vuorovaikutuksia voivat olla lämpöstabiilisuutta edistäviä tekijöitä n*Ta*AA9A:ssa.

Termofiilisen sienen *Chaetomium thermophilum* (*Ct*SOD) Cu,Zn-superoksididismutaasin kiderakenne määritettiin 1,56 Å:n resoluutiolla. *Ct*SOD kiteytyi kahdeksalla molekyylillä (A-H) kristallografisessa asymmetrisessä yksikössä, mikä

johti kahdeksaan erilliseen rajapintaan. Sinkkiä oli kaikissa alayksiköissä, mutta kuparia vain 4 alayksikössä (C, D, E ja F). Entsyymin aktiivinen kohta sekä kuparia ja sinkkiä sitovat kohdat havaittiin erittäin konservoituneiksi. Korkeampi oligomerisoitumisaste ja suurempi polaariset aminohappojäännösten osuus *Ct*SOD:ssa voivat olla lämpöstabiilisuutta edistäviä tekijöitä. .

Termofiilisestä *Caloramator australicus* -bakteerista (γ-*Ca*CA) peräisin olevan hiilihappoanhydraasin kiderakenne määritettiin 1,11 A:n resoluutioon. Tämä on tähän mennessä korkein resoluutio γ-perheen hiilihappoanhydraasille. Entsyymi kiteytettiin 3 molekyylin asymmetrisessä yksikössä. Jokaisen molekyylin aktiivinen kohta löydettiin kahden vierekkäisen molekyylin rajapinnasta. γ-*Ca*CA-rakenne havaittiin erittäin konservoituneeksi, mutta silmukka-alueilla havaittiin eroja. Varauksellisia aminohappotähteitä ja hydrofobisia klustereita ehdotettiin mahdollisiksi lämpöstabiilisuutta edistäviksi tekijöiksi γ-*Ca*CA:ssa.

Termosyymeillä on lupaava tulevaisuus biotaloudelle ja vihreälle kemialle. Tässä esitetyt tulokset voivat tarjota uusia ideoita kestävien ja ympäristöystävällisten ratkaisujen kehittämiseen parempaa tulevaisuutta varten.

Table of Contents

Abbreviations

ASA	Accessible Surface Area
AA9	Auxiliary Activity 9
BGC	β-D-glucose
CAZy	Carbohydrate-Active enZymes
*Ct*BGL	*Chaetomium thermophilum* β-glucosidase
*Ct*SOD	*Chaetomium thermophilum* Superoxide Dismutase
Cu,Zn-SOD	Copper,Zinc-Superoxide Dismutase
DP	Degree of Polymerization
EMBL	European Molecular Biology Laboratory
Fe-SOD	Iron-Superoxide Dismutase
GH3	Glycoside Hydrolase Family 3
HPLC-RID	High Performance Liquid Chromatography-Refractive Index Detector
LPMO	Lytic Polysaccharide Monooxygenase
MALDI-TOF MS	Matrix-Assisted Laster Desorption Ionization-Time Of Flight Mass Spectrometry
Mn-SOD	Manganese-Superoxide Dismutase
MPD	2-methyl-2,4-pentanediol
MR	Molecular replacement
PDB	Protein Data Bank
NAG	N-acetyl-β-D-glucosamine
NCS	Non-Crystallographic Symmetry
Ni-SOD	Nickel-Superoxide Dismutase
*Ta*AA9A	*Thermoascus aurantiacus* Auxiliary Activity 9
RMSD	Root mean square deviation.
SAS	Solvent Accessible Surface
SDG	Sustainable Development Goal
SOD	Superoxide Dismutase
$T_{1/2}$	Half-life
TFA	Trifluoroacetic acid
TLC	Thin Layer Chromatography
Tm	Melting temperature

Amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamate	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Selenocysteine	Sec	U
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I. **Mohsin I**, Poudel N, Li DC, Papageorgiou AC. Crystal Structure of a GH3 β-Glucosidase from the Thermophilic Fungus *Chaetomium thermophilum*. Int J Mol Sci. 2019 Nov 27;20(23):5962. doi: 10.3390/ijms20235962. PMID: 31783503; PMCID: PMC6929035. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6929035/

II. Yu, W., **Mohsin, I**., Papageorgiou, A.C., Li DC. Purification and Structural Characterization of the Auxiliary Activity 9 Native Lytic Polysaccharide Monooxygenase from *Thermoascus aurantiacus* and Identification of Its C1- and C4-Oxidized Reaction Products. Catalysts 2022, 12, 139; https://doi.org/10.3390/catal12020139, SN: 2073-4344. https://www.mdpi.com/2073-4344/12/2/139/html

III. **Mohsin I**, Zhang LQ, Li DC, Papageorgiou AC. Crystal Structure of a Cu,Zn Superoxide Dismutase From the Thermophilic Fungus *Chaetomium thermophilum*. Protein Pept Lett. 2021;28(9):1043-1053. doi: 10.2174/0929866528666210316104919. PMID: 33726638. https://pubmed.ncbi.nlm.nih.gov/33726638/

IV. Bodourian, CS, **Mohsin I**, Georgakis, ND, Papageorgiou, AC, Labrou, NE. Structural and functional characterization of a metagenomically-derived inactive γ-type carbonic anhydrase and its engineering into a hyperthermostable esterase (in preparation).

The original publications have been reproduced with the permission of the copyright holders.

Original Publications Not Included in the Thesis

I. Papageorgiou AC, **Mohsin I**. The SARS-CoV-2 Spike Glycoprotein as a Drug and Vaccine Target: Structural Insights into Its Complexes with ACE2 and Antibodies. Cells. 2020 Oct 22;9(11):2343. doi: 10.3390/cells9112343. PMID: 33105869. PMCID: PMC7690584. https://www.mdpi.com/2073-4409/9/11/2343

II. **Mohsin, I**., Papageorgiou, A.C. (2022). Fungal Extremozymes: A Potential Bioresource for Green Chemistry. In: Sahay, S. (eds) Extremophilic Fungi. Springer, Singapore. http://doi.org/10.1007/978-981-16-4907-3_27 https://link.springer.com/chapter/10.1007/978-981-16-4907-3_27

1 Introduction

Currently, Earth's population is more than 7.8 billion and is projected to grow to 9.7 billion by 2050 and 11.1 billion by 2100 [1]. The growing population has led to rapid industrial growth to fulfil increasing demands. However, the resources required to meet human needs and living standards are steadily diminishing. In addition, the distribution of resources is not uniform, and the stockpiles of waste and pollutants produced by humans are amassing. This situation not only threatens the sustainability of our planet but also the survival of every form of life on it.

In 2015, the United Nations (UN) chalked out 17 Aims to achieve sustainability and prosperity for human beings by 2030. These Aims are termed Sustainable Development Goals (SDGs) and include good health and well-being, clean water and sanitation, affordable and clean energy, sustainable cities and communities, responsible consumption and production, climate action, etc.[2]. These goals are implemented through international watchdogs such as the World Health Organization (WHO) and the UN that are leading us to achieve a sustainable future for our planet. However, sustainability is a relative term that can only be achieved if every sphere of life contributes to it.

For a sustainable future, techniques and processes that result in maximum yields with low carbon emissions must be developed. We need to i) eliminate food shortages without polluting the environment, ii) improve health and well-being by harvesting disease-free crops and inventing non-invasive diagnostic and treatment methods, and iii) harness clean energy, invest in green chemistry, develop sustainable cities, and rediscover industrialization with no carbon footprint. For all of these to happen, we need to invest in research and development to obtain green solutions.

Enzymes (i.e., proteins that act as catalysts) are one of the tools for realizing and addressing SDGs. Enzymes have broad range of applications in healthcare, agriculture, and other industries. However, the production and use of enzymes can have significant environmental impacts, highlighting the need for sustainable and environmentally friendly processes. One of the key challenges in enzyme production and use is their stability and activity under various conditions. There is a significant need for enzymes that can function at high temperatures in a variety of industrial

processes, to replace expensive equipment and energy-intensive operations. Such enzymes are referred to as thermophilic enzymes and their ability to tolerate high temperatures is described as thermostability.

In this thesis, we investigated the structural features of thermophilic enzymes and the key factors that contribute to their thermostability. Using X-ray crystallography, the most widely used technique to determine the three-dimensional (3D) structure of proteins, we elucidated the structural details of four enzymes to gain a deeper understanding of their function and properties. These enzymes include a β-glucosidase from *Chaetomium thermophilum* (*Ct*BGL*)*, a native lytic polysaccharide monooxygenase from *Thermoascus aurantiacus* (n*Ta*AA9A), a superoxide dismutase from *Chaetomium thermophilum* (*Ct*SOD), and a gamma-carbonic anhydrase from *Caloramator australicus* (γ-*Ca*CA).

Understanding the thermostability of thermophilic enzymes can help to explain their unique characteristics and pave the way for developing new enzymes with enhanced capabilities. The production and use of enzymes with improved properties can positively affect the environment and reduce the need for energy-intensive processes. Moreover, novel thermophilic enzymes with improved properties have broad applications in various fields including biocatalysis, biofuel production, and pharmaceuticals. The findings of this thesis can therefore contribute to the development of sustainable and environmentally friendly processes that can support the achievement of the SDGs and ensure a better future for all.

1.1 Enzymes

The term ‘enzyme’ was first coined in the late 19th century by the German physiologist Wilhelm Kühne while he was experimenting with fermentation. The word ‘enzyme’ has got Greek roots: ‘en’ means ‘in’ and ‘zyme’ means ‘yeast’. Therefore, scientists believed that an ‘enzyme’ was a ‘substance’ found in yeast and responsible for fermentation to occur. However, in modern times, the term has broadened and is defined as ‘the chemical substance (protein) produced by the living organisms to catalyze biological chemical reactions by lowering the activation energy through stabilizing the transition state’. In 1980s, some ribonucleic acids (RNAs) involved in gene expression, were found to possess catalytic activity as well. These RNAs were also labelled as enzymes and are known as ribozymes [3,4]. However, enzymes discussed in this thesis are of protein nature and composition.

1.1.1 Enzymatic mechanism

In a simple catalytic reaction, enzymes act on a substance called ‘substrate’ and yield reaction products (Figure 1). Enzymes are very specific in nature, that is, each

enzyme acts on a specific substrate. They are also characterized by high efficiency and can increase the normal reaction rates by 100 million to even 10 billion times [5].

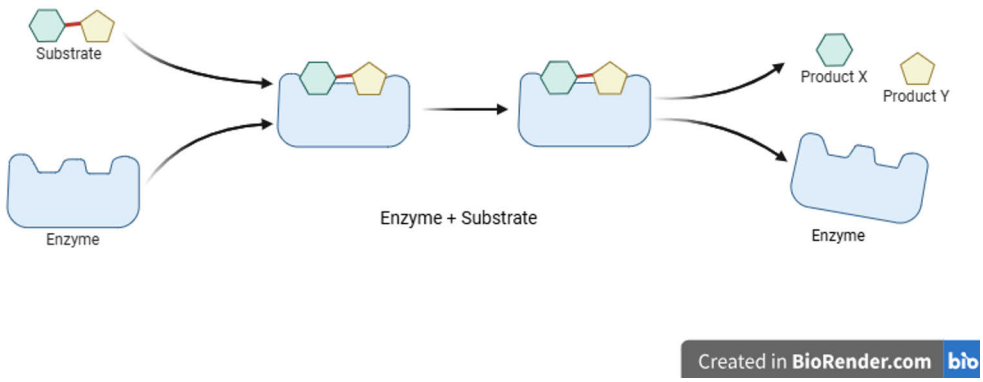


Figure 1. Schematic diagram of a simple enzymatic reaction.

In 1894, Emil Fischer, a Nobel laureate, proposed that enzymes and substrates have exact complementary geometric shapes, such that they can only fit into one another. This is also known as a 'lock and key' model [5]. This simple model has served for many years to explain enzyme function. However, it is now known that enzymes are not static entities, but they can easily adopt various conformational stages to facilitate the substrate binding and achieve catalysis by lowering energy required for the conversion of the substrate to a product [6,7]. Stein *et al.* observed that enzyme- substrate binding events with RMSD smaller than 1.5 Å follow Fisher's model (65%) while others with RMSD higher than 1.5 Å follow either a conformational model (13%) or an induced-fit model (0.8%) [8].

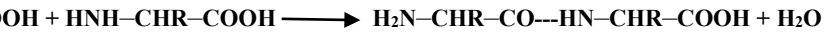
The site of the enzyme that interacts with substrate is known as the 'active site' and its residues as 'active site residues'. Substances that directly or indirectly prevent the binding of the substrate to the active site of the enzyme are known as 'inhibitors'. Inhibitors could be competitive or non-competitive depending on their binding to the enzyme. Contrary to the inhibitors, there are substances that may activate the enzyme and are known as 'activators'. Inhibitors and activators may influence, regulate, and control enzymatic activity. Moreover, an enzyme may have a non-protein part attached to it, called 'co-factor', which could be organic or inorganic in nature and assists in the enzymatic reaction. An enzyme with or without its cofactor is referred to as holoenzyme or apoenzyme, respectively [9].

1.1.2 Classification of enzymes

In 1955, the International Congress of Biochemistry in Brussels established a commission for enzyme nomenclature called the Enzyme Commission (EC). This commission published its first version in 1961, and classified enzymes based on their catalytic reactions. According to EC, enzymes were placed into seven different classes and each class was assigned with a specific EC number i.e., [EC 1]- Oxidoreductase, [EC 2]- Transferase, [EC 3]- Hydrolase, [EC 4]- Lyase, [EC 5]- Isomerase, [EC 6]- Ligase, and [EC 7]- Translocase. These classes were further divided to elaborate each type of enzyme within the class, for example, amylases belong to the class [EC 3] and are further divided into subclasses of α-amylases, β-amylases, and γ-amylases as [EC 3.2.1.1], [EC 3.2.1.2], and [EC 3.2.1.3], respectively [5,10,11]. The first digit of an EC number refers to the type of reaction or enzyme class, the second and third digit to the molecular group, bond, or product involved, and the fourth digit to the specific member of the class related to its metabolites and/or cofactors involved [12].

1.1.3 Factors affecting enzyme activity

Enzymes that belong to proteins are made up of amino acids that are linked to each other via peptide bonds. These bonds are covalent in nature and are formed by the loss of a water molecule between the carboxyl group of one amino acid and the amino group of the next one in the polypeptide chain.



Peptide bonds exhibit planarity due to partial double-bond character (40 %) caused by the resonance of the carbonyl and amide group of the contributing amino acids [13]. The resonance stabilization, trans configuration, and planar structure of the peptide bond contributes towards the protein's stability. Apart from the peptide bond, hydrophobic interactions are important for protein folding and stability. Hydrophobic interactions may contribute up to 60 ± 4 % towards protein stability by allowing the protein to decrease in surface area and avoiding unnecessary water interactions [14]. In addition to peptide bonds and hydrophobic interactions, and as in all protein structures, ionic bonds, hydrogen bonds, van der Waals forces, and disulfide bonds are also present in enzymes. All these interactions ensure the natural shape, structure, and function of the enzyme. However, some of these bonds are vulnerable and may break under a certain range of conditions. The breakage of bonds results in enzyme's denaturation and loss of natural shape, structure, and function.

Each enzyme has an optimum operating condition where it exhibits its peak catalytic activity. Several factors including temperature and pH are at play in configuring conformational changes in an enzyme. Under acidic/alkaline conditions,

because of pH variation, the enzymatic activity can vary significantly. Similarly, temperature affects catalytic activity of the enzyme. For instance, high temperature may render the enzyme denatured. The thermal stability of an enzyme can be assessed by subjecting the enzyme to a range of temperatures for a fixed period of time and then assessing its activity at optimum conditions. The temperature at which thermal denaturation begins varies from one enzyme to another. Generally, enzymes do not experience a thermal denaturation below 30 °C. However, thermal denaturation may become significant above 40 °C [3]. Enzyme thermostability also depends on the enzyme's source such as enzymes produced from halophiles or thermophiles are naturally tolerant to high salt concentration or high temperature, respectively. In general, enzymes produced by microorganisms are more thermostable than those found in mammals (Table 1) [15,16].

Table 1. Comparison of microbial and mammalian sourced enzymes.

Feature	Microbial	Mammalian
Origin	Microorganisms such as fungi and bacteria.	Animal tissues, organs, or cells.
Production	Mostly genetically modified organisms are used for industrial scale production, and offer high yield, and cost-effectiveness.	Their production is expensive, tedious and sometimes comes with ethical challenges.
Purity and Specificity	They offer high degree of engineered purity and specificity.	They can also offer high degree of purity and specificity but risk of contamination from the animal tissues is also higher
Performance	Their performance in various physiological conditions is challenging. However, they are robust and stable in wide range of pH and temperature conditions.	They often perform better in physiological conditions with high rates of activity. However, they are often sensitive to the environmental conditions and found less stable compared to their microbial counterparts.
Applications	They have numerous industrial applications such as in textile, food and beverage industry, and biofuel production.	These enzymes have applications in pharmaceutical industry as biomarkers and therapeutic agents.

1.1.4 Microbial enzymes

1.1.4.1 Fungi

Fungi are widespread in nature and have recently emerged as a valuable source of different types of enzymes. Most fungal enzymes are produced extracellularly; hence, they can be economically produced in large quantities. Moreover, fungal enzymes require simple purification methods, which make them suitable for

commercial purposes. Fungal enzymes have applications in medicine, agriculture, and bioremediation. They are also used in baking, brewing, cheesemaking, detergents, textiles, and paper production. Currently, more than 50% of all commercially produced enzymes are of fungi origin [17].

1.1.4.2 Bacteria

Bacterial enzymes have certain distinct features that make them more suitable and viable than the plant- or animal-derived enzymes. Their key characteristics are as follows [3].

- They are easy to extract and purify.
- They can be produced in batches and their quality is assessed using various biochemical assays.
- They can provide uniform batch yields since they are produced in controlled environments.
- They are generally more stable in processes that require high temperatures.
- They can be genetically engineered with desired traits.

1.1.4.3 Bacterial vs Fungal thermozymes

Thermophilic bacteria thrive in extreme environmental conditions such as hot springs and/or geothermal vents while adapting and maintaining functionality, Therefore, enzymes produced from such bacteria are thermophilic or hyperthermophilic in nature. On the other hand, fungi inhabit moderate environmental conditions, resulting in comparatively less thermostable enzymes [18]. Currently, most thermozymes are obtained from bacteria and fungi. Both sources offer valuable and potential thermophilic enzymes that have multidimensional uses and a wide range of applications. However, there are some significant differences between them. Examples are given below [19,20].

- Bacterial thermozymes have an optimum temperature of above 70 °C whereas fungal thermozymes have an optimum temperature of around 50–70 °C.
- Since bacteria belong to prokaryotes, thermozymes from bacteria are used in processes where their use in eukarya is minimal to none. On the other hand, fungi are the only eukarya that are found to survive above 61 °C. Thermozymes from fungi are therefore ranked as significantly valuable,

because they offer high compatibility with native eukaryotic enzymatic systems.

- Bacterial thermozymes such as amylases, lipases, and proteases are used in high-temperature industrial processes. Fungal thermozymes, such as cellulases, xylanases, and amylases are used in industrial processes where the temperature is moderate to high.

1.1.5 Industrial Enzymology

The study of industrial and commercial enzymes and their applications is referred to as industrial enzymology. Industrial enzymes have evolved steadily over the last century. For instance, in 1890, a commercial production unit for takadiastase (a mixture of amylolytic and proteolytic enzymes) was established in the United States of America (USA) by the Japanese scientist Jokichi Takamine. Takadiastase was obtained by cultivating *Aspergillus oryzae,* a filamentous fungus, on rice or wheat barns. Takadiastase became a commercial hit in the USA for dyspepsia treatment. Similarly, in 1913, in France, August Boidin and Jean Effront found that a bacterial heat-stable enzyme (α-amylase) was produced by *Bacillus subtilis*. This enzyme found applications in the textile industry, especially in the removal of starch during cotton production [3,21].

In 1930 and subsequent years, scientists began using fungal pectinases and other hydrolases in the preparation of fruit products. In the 1960s, acid hydrolysis for hydrolysing starch was replaced by glucoamylase. In the 1970s, proteases were used in detergents and glucose isomerase was used as a catalyst in high-fructose syrups for reversible isomerization of D-glucose to D-fructose. In the 1990s, lipases were incorporated into washing powders and several immobilized enzyme processes were developed for industrial use.

Global enzymes market witnessed an uphill growth from £110 M in 1960 to £2000 M in 2010. By the end of 2010, two companies, Novozyme (Danish) and DuPont (US) with a 47% and a 21% market share, respectively, supplied more than two-thirds of the global enzyme market. There are currently about 40–50 industrial enzymes that are produced at industrial scale, i.e., in tonnes/annum. Among these industrial enzymes, hydrolases (proteases), amylases, cellulases, and lipases account for the largest share in the global enzyme market [3].

Currently, in the fields of commerce and technology, enzymes are categorized into four types, as listed in Table 2 [3].

Table 2. Application-based classification of enzymes and representative examples.

Category	Enzyme	Reaction	Source	Application
Industrial Catalysts	Acid/Alkaline proteases	Protein digestion	*Aspergillus niger / Bacillus species*	Milk coagulation/ Detergents and washing powder
Therapeutic agents	L-Asparaginase	Removal of L-asparagine from tumors	*E. coli*	Chemotherapy
Analytical reagents	Glucose oxidase/peroxidase	Glucose oxidation / De-oxidation	*Aspergillus niger / Horseradish*	Detection and quantification of glucose in blood / hormones and antibodies
Manipulative tools for genetic engineering	DNA polymerases	DNA synthesis	*Thermus aquaticus*	DNA amplification in PCR

1.1.6 Extremozymes and thermozymes

The word 'extremophile' is a combination of two different words: '*extremus*', a Latin word meaning 'utmost/outermost', and '*philos*', a Greek word meaning 'friend'. In the scientific literature, the term 'extremophiles' refers to biological organisms found in the extremities of nature. The extremities can be of any form, such as temperature, pH, pressure, salinity, and radiation (Table 3). Most of these extremophiles are found in just one extremity, but some, such as *Bacillus licheniformis*, can tolerate more than one extreme habitat and are known as 'polyextremophiles'. Extremophiles could be viruses, bacteria, fungi, archaea, prokaryotes and/or eukaryotes [22–24].

Extremophiles that thrive at high temperatures are known as thermophiles (45–80 °C) and hyperthermophiles (above 80 °C). In lower temperatures, they are categorized as mesophiles (15–45 °C) and psychrophiles (-5–20 °C). Enzymes obtained from such extremophiles are known as thermozymes, hyperthermozymes, mesozymes, and psychrozymes, respectively. The degree of enzymatic stability against temperature (thermotolerance) increases gradually from psychrozymes to the hyperthermozymes [25].

Table 3. Categories of extremophiles and representative microorganisms.

Extremity	Termed as	Examples
Temperature	Psychrophiles / Thermophiles	*Cryptococcus laurentii / Thermatoga maritima*
pH	Acidophiles/Alkaliphiles	*Pithomyces chartarum / Alcaligenes faecalis*
Pressure	Piezophiles	*Colwellia marinimaniae*
Salinity	Halophiles	*Aspergillus sydowii*
Radiation	Radiophiles	*Deinococcus radiodurans*

1.1.7 Enzyme Thermostability

Enzyme thermostability refers here to the property of enzymes to operate at high temperatures while maintaining their catalytic activity. There is not a fully defined mechanism yet available to explain enzyme and in general protein thermostability. Although Arrhenius equation provides protein folding/unfolding kinetics, there are several other factors involved in thermostability. This is because a protein carries numerous features such as, a unique sequence, ion pairs, hydrogen bonds, hydrophobic interactions, disulphide bridges, packing, folding and unfolding entropy, and intra-chain interactions [26–29]. Every protein has a distinctive set of features that affect its thermostability. For example, a higher percentage of charged amino acid residues (Asp, Glu, Arg, and Lys) has been observed in various thermophilic enzymes. These charged residues lead to increased ion interactions, eventually influencing the resistance of the enzyme to high temperatures [30,31]. Increased number of charged residues, particularly at the surface of the proteins, is crucial for improving protein thermostability [32,33]. Similarly, a stronger hydrophobic interior has been proposed as major thermostability contributing factors in other enzymes [34].

Enzyme flexibility also plays a role in thermostability. A comparative analysis between mesophiles and thermophiles suggested that thermophiles had a higher proportion of amino acid residues such as Asp, Glu, Arg, Lys, and Tyr as compared to the Ala, Asn, Gln, Thr, Ser, and Val residues [33,35]. The percentage of these amino acid residues may affect enzyme flexibility, which in turn imparts enzyme thermostability [36]. Likewise, Pro and Gly residue percentages in an enzyme manoeuvre thermostability as they reduce or induce backbone entropy, respectively [28].

Thermostability in an enzyme can be achieved either by natural selection or by the available scientific means, that is, protein engineering. Although researchers have been devising and reporting different protein engineering strategies in the last decade, site-directed mutagenesis and directed evolution have been the most used approaches in practice. Table 4 illustrates the key differences between the two approaches [37,38].

Table 4. Key features of directed evolution and site-directed mutagenesis.

Directed Evolution (DE)	Site-directed Mutagenesis (SDM)
DE is an iterative process of generating variant libraries and screening.	SDM creates mutants by incorporating, substituting and/or deleting a specific codon of a gene of interest.
Mutants of desired traits obtained in each cycle serve as template for the next cycle.	PCR is employed to amplify the desired fragment containing the mutation.
No prior knowledge of protein structure is required.	Prior knowledge of protein structure is required.
DE experiments provide the full evolutionary trajectory of a given trait.	Disadvantages are low success rate, time-consuming, and high cost.
In DE experiments, the development of a suitable screening is the major challenge.	Low protein yield, expression variability, and incorrect folding are also observed in SDM.

Currently, artificial intelligence (AI) and computational methods are also in practice to improve thermostability profiles of enzymes. AI can be used to guide directed evolution, automate the design process, develop efficient models, and predict enzyme structures. Machine learning (ML) and artificial neural networks (ANNs) are the key drivers in developing AI-based programs. For instance, FuncLib, an automated program for designing multipoint mutations at the active site of protein and implemented as a web server was developed using computational methods. FuncLib algorithm utilizes phylogenetic analysis, and Rosetta design calculations for yielding the best possible outcomes. Similarly, MutCompute is another machine learning-based program for protein mutagenesis prediction and 3D visualisation [39,40].

1.1.7.1 Determining thermostability

There are certain methods both *in vitro* and *in silico* available to measure and/or predict thermostability. *In vitro* methods include calorimetric and spectrometric techniques. Differential scanning calorimetry (DSC) is an example of a calorimetric technique in which the enthalpy of unfolding of a protein is directly determined along with the melting temperature (Tm) and rate constant of thermal denaturation (k) [41,42]. Differential scanning fluorimetry (DSF), also known as thermofluor, is another thermal shift assay technique like DSC. However, in DSF, specialized fluorogenic dyes, such as SYPRO™ Orange, are used and changes in fluorescence are measured at thermal unfolding of the protein. DSF is a high-throughput technique and requires very small amount of sample [43]. Circular dichroism (CD) spectroscopy is a technique used to study the unfolding of a protein structure using ultraviolet light (UV). The radiation used in CD can be far-UV or near-UV to measure the unfolding of the secondary or tertiary structure of a protein, respectively. [44]. Both techniques (DSF and CD) are in general practice, however, limitations such as the high protein purity required in DSF and detection limits in CD may affect enzyme activity/stability data profile.

In silico methods include computational and machine learning methods for predicting the thermostability of a protein. Researchers have developed several prediction models based on the structure, sequence, and molecular dynamics simulations of proteins in solution. Computational programs available can perform extensive protein structural homolog searches, analyze protein folding patterns, evaluate amino acid compositions, quantify electrostatic interactions, and predict mutation hotspots and melting temperatures. ProThermDB (https://web.iitm.ac.in/bioinfo2/prothermdb/index.html) is one such thermodynamic database for proteins and mutants. *In silico* approaches for predicting thermostability are more robust, less time-consuming, and cost-effective. However, achieving high

prediction accuracy and efficiency remains a challenge for developers. Most available programs could only claim moderate accuracy owing to the complex and multifactorial nature of thermostability [45,46].

1.1.8 Thermozymes of Interest

In this thesis, we have investigated four thermophilic enzymes that belonged to the EC 3, EC 1, and EC 4 class of the EC system. This section discusses the given EC class characteristics and applications in detail.

1.1.8.1 Hydrolases

Hydrolases are enzymes that catalyze a hydrolysis reaction and belong to the class EC 3. These enzymes have been further divided into 13 subclasses, that is, EC 3.1 – EC 3.13, based on the chemical bonds they hydrolyse. For example, enzymes of subclass EC 3.1, EC 3.2, EC 3.3, and EC 3.4 cleave an ester, glycosidic, ether, or peptide bond and are therefore known as esterases, glycosidases, ether hydrolases, and proteases/peptidases, respectively. A hydrolysis reaction of the peptide bond is given below.

Hydrolases break biopolymers into smaller molecules and are involved in processes, such as digestion, transport, excretion, and regulation. Hydrolases hold approximately two-thirds of the share and are the largest contributor to the global market for industrial enzymes. Notably, glucosidases, proteases, and lipases account for more than 70% of all enzyme sales [47,48].

1.1.8.1.1 Glucosidases

Glucosidases belong to the family of glycoside hydrolase 3 (GH3) in subclass EC 3.2.1.21. These enzymes are involved in the hydrolysis of cellulose, which is the most abundant and renewable non-fossil carbon source on Earth. Cellulose is composed of repeated cellobiose units linked by β-1,4-glycosidic bonds. This bonding in cellulose confirms the crystallinity of the structure with very few amorphous regions. Thus, a set of synergistic cellulases is required for complete digestion of cellulose into glucose units. This set consists of endoglucanases, exoglucanases, and β-glucosidases. Endoglucanases act on amorphous regions of the cellulose polymer and expose the reducing/non-reducing ends of cello-oligosaccharides for the exoglucanases to act and yield cellobiose. β-glucosidases further catalyse cellobiose to glucose monomers [49,50].

Glucosidases are categorized as α- and β-glucosidases, depending on the orientation of the glycosidic bond they cleave. They are abundant in nature; α-glucosidases are predominant in mammals while β-glucosidases are ubiquitous and can be found in animals, plants, bacteria, and fungi [51]. Filamentous fungi have also been recorded as the major producers of β-glucosidases. β-Glucosidases can be successfully expressed in yeast (*S. cerevisiae* and *P. pastoris*), bacteria (in *E. coli* with pET expression vectors) and fungi (e.g., *Aspergillus sp.* and *Trichoderma sp.*) [52].

β-Glucosidases are emerging as significant catalysts in the conversion of lignocellulosic plant biomass to biofuels (ethanol, butanol and hydrocarbons). Li *et al*. [53] reported cloning and characterization of a β-glucosidase from rumen microbes of cattle feeding with *Miscanthus sinensis* (Chinese silver grass) plant. This plant has been reported as an ideal for biofuel production. β-glucosidases have an established role in wine, tea, and other beverage industries. They are employed for catalyzing a reaction that releases aromatic compounds from glycosidic precursors in fruit juices and fermenting products [54]. A β-glucosidase isolated from *Thermomicrobium roseum* and expressed in *E. coli* has been reported to be stable at pH range of 5.5 – 10.0 and temperature above 70 °C. It has been found active in high-temperature cellulose hydrolysis to reduce sugars from seawater [52]. Owing to their high pH sensitivity for determining biochemical changes, β-glucosidases are also used as soil biochemical indicators [55].

1.1.8.2 Oxidoreductases

Oxidoreductases are enzymes that catalyze oxidation/reduction reactions and belong to the class EC 1 in the EC system. They are further divided into subclasses based on the substrate upon which they act. For example, oxidoreductases of EC 1.1.1, EC 1.3.1, and EC 1.5.1 subclass enzymes act on the CH-OH, CH-CH, and CH-NH groups of donors, respectively. Oxidases, dehydrogenases, reductases, dismutases, and oxygenases are different types of oxidoreductases [11,48,56]. A simple redox reaction is given below.

$$\mathbf{AH + B \longrightarrow A + BH\ (reduced),} \qquad \mathbf{A + O \longrightarrow AO\ (oxidized)}$$

1.1.8.2.1 Superoxide Dismutases (SODs)

SODs are antioxidant metalloenzymes and belong to the subclass EC 1.15.1.1 of oxidoreductases. Copper, zinc, manganese, iron, and nickel are commonly found metals in SODs. Generally, SODs are named based on the metal they contain, for example, SODs containing iron or nickel are referred to as FeSODs or NiSODs, respectively. The metal at the active site acts as a mediator to transport electrons

from one oxygen molecule to another, and the oxidation-reduction catalysis reaction is carried out. During cellular metabolism, oxygen undergoes a series of reductions to produce superoxide (O_2^-), hydroxyl radicals (HO•), and hydrogen peroxide (H_2O_2). These unstable molecules are also known as highly reactive oxygen-containing species (ROS). ROS cause oxidative stress and are involved in damaging intracellular targets such as proteins, lipids, and DNA. Therefore, living organisms have developed antioxidant enzymes as a defence strategy, including SODs to detoxify the ROS [48,57].

SODs are found in both prokaryotes and eukaryotes. SODs in eukaryotes are divided into three classes based on their cellular localization and metal cofactor. SOD1 and SOD2 enzymes contain Cu^{+2}/Zn^{+2} and Mn^{+2} as metal cofactors and are found in the cytoplasm and mitochondrial matrix, respectively. However, SOD3 enzymes are found in the extracellular fluid with Cu^{+2}/Zn^{+2} as metal cofactors. SOD1 are 32 kDa homodimer enzymes, and its catalytic activity is mainly attributed to the presence of copper, whereas the presence of zinc is attributed to protein folding and stability of the enzyme's molecular structure. SOD2 are 96 kDa homotetramer enzymes with manganese as the metal cofactor required for their catalytic activity. SOD3 are 135 kDa homotetramer enzymes. They are found in extracellular fluids such as blood, lymph, and synovial fluid. In most species, they are found as dimers linked with disulfide bridges [58].

1.1.8.2.2 Lytic Polysaccharide Monooxygenases (LPMOs)

LPMOs are auxiliary activity (AA) enzymes that aid other Carbohydrate-Active Enzymes (CAZy) in achieving maximum yield of cellulose hydrolysis. They belong to the oxidoreductase class (EC 1) in the EC system. According to the CAZy database, AA enzymes are classified into 17 families (AA1-AA17), 9 families of lignolytic enzymes, 7 families (AA9-AA11, and AA13-AA16) of LPMOs, and an indistinct AA8 family. [59]. Members in each of these families share sequence similarity, biochemical characterization, and/or three-dimensional structure [60].

LPMOs are found abundant both in bacteria and saprophytic fungi. However, the majority of fungal LPMOs fall into 5 LPMO families: AA9, AA11, AA13, AA14, AA16 [61,62]. LPMOs are metalloenzymes as they require a metal ion at the active site to perform catalytic activity. For example, members of the AA9 family are copper-dependent LPMOs (formerly GH61) and catalyse xylan and cellulose polysaccharides through hydroxylating or dehydrogenating activity at -C1 or -C4 of the glycosidic bonds [59,63].

LPMOs have numerous industrial and biotechnological applications. Lignocellulosic biomass is abundantly available in nature and can offer an alternative in the form of biofuels. However, owing to the recalcitrant nature of this complex

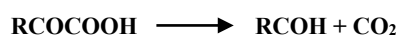
polysaccharide, a cocktail of enzymes is required for complete bioconversion. LPMOs act on crystalline polysaccharide chains and disrupt their structure, making them more accessible to hydrolytic enzymes for catalytic reactions. LPMOs from AA9 and AA10 families are therefore considered as boosters of lignocellulosic biomass conversion due to their catalytic efficiency on crystalline polysaccharides [64].

AA9 LPMOs can have synergistic interactions with other enzymes such as cellulases. Synergistic interactions are observed because cellulose oxidation by AA9 LPMOs exposes more ends of the cellulose polysaccharide for cellulase enzymes to act upon. For example, an LPMO isolated from *Gloeophyllum trabeum* (*Gt*GH61) confirmed the synergistic effects with endoglucanase (*Gt*Cel5B) and xylanase (*Gt*Xyl10G) on lignocellulosic biomass conversion [65,66].

Another example of synergistic interactions can be seen in case of *Trichoderma reesei*, an industrial cellulase producer, was found to express low levels of three AA9 LPMOs in its secretome. It was suggested that the introduction of exogenous AA9 LPMOs would significantly increase the catalytic efficiency of the enzyme solution from *T. reesei* [66]. This was confirmed by Harris *et al.* [67], as they halved the cellulase required for saccharification by expressing *Ta*AA9A in *T. reesei.* Moreover, the addition of AA9 LPMOs can also increase the glucose yield in lignocellulosic biomass conversion [67].

1.1.8.3 Lyases

Lyases, also known as transeliminases, are enzymes involved in the cleavage of C-C, C-O, and C-N bonds. These enzymes may also cleave other bonds, such as those between C-S and N-O. Lyases do not utilize hydrolysis or oxidation processes for their reaction catalysis and belong to class 4 in the EC system. A simple reaction catalyzed by lyases is given below.



They are further divided into subclasses ranging from 4.1 to 4.99, based on the type of bond they cleave. For example, enzymes of subclass 4.1 catalyze the breakage of C-C bonds and are further categorized as decarboxylases (EC 4.1.1), aldehyde-lyases (4.1.2), and other C-C lyases (EC 4.1.99). Similarly, enzymes of subclass 4.2 catalyze reactions involving cleavage of the C-O bond. Its sub-subclasses are formed based on eliminated groups of the reactions they catalyze, for example water (hydrolyases, EC 4.2.1), alcohol from polysaccharides (EC 4.2.2), phosphate (EC 4.2.3.), or other group (EC 4.2.99). Lyases are also categorized based on the substrate on which they act. For example, polygalacturonate lyase and polymethylgalacturonate lyase act on pectate and pectin, respectively [68,69].

1.1.8.3.1 Carbonic anhydrases (CAs)

Carbonic anhydrases are lyase enzymes that catalyze the reversible hydration and dehydration of carbon dioxide and bicarbonate, respectively. These enzymes are classified as EC 4.2.1.1 and are considered as the most efficient enzymes with a catalytic turnover rate of 600 000 substrate molecules (CO_2 and H_2O) into the products ($HCO3^-$ and H^+) in every second per molecule of the enzyme. The reaction catalyzed by CA is given below.

$$CO_2 + H_2O \rightleftharpoons H_2CO_3$$

CAs are metalloenzymes with zinc as the metal cofactor required for the catalytic activity. Besides zinc, in most cases, the active site has three histidines as catalytic residues. Some CAs may also contain cysteine and/or glutamine as catalytic residues. From an evolutionary perspective, CAs can be categorized into three families: α-CAs, β-CAs, and γ-CAs. α-CAs are generally found in vertebrates, eubacteria, algae, and in the cytoplasm of green plants. β-CAs are predominantly found in eubacteria, algae, and chloroplast of mono- and di-cotyledons. γ-CAs are found in methanogenic bacteria in hot springs. CAs have also been discovered in marine diatoms (δ- and ζ-CAs) and *Plasmodium species* (η-CAs) [3,70–72].

The growing population and the rampant use of fossil fuels are considered key reasons for the elevated CO_2 concentration in the atmosphere. This not only causes the global temperature to rise but also depletes the ozone layer. CAs are gaining significant importance because of their ability to capture and store atmospheric carbon. Although the carbon capture and storage (CCS) process is challenging owing to the high temperatures and the presence of metals and ionic concentrations, several thermostable CAs have been proposed and tested, some examples of which are given in Table 5 [72,73].

Table 5. CAs in CCS.

CAs	Source	Acronym	Optimum temperature	Retained CO_2 hydration activity
α-CAs	*Sulfurihydrogenibium yellowstonense*	*Ssp*CA	95 °C	100 °C for 3 h
	Sulfurihydrogenibium Azorense	*Saz*CA	80 °C	100 °C for 3 h
β-CAs	*Methanobacterium thermoautotrophicum*	Cab	65 °C	75 °C for 15 mins
γ-CAs	*Methanosarcina thermophila*	*Mt*Cam	Not Determined	55 °C for 15 mins

CCS systems, if implemented, could reduce 14 % of CO_2 emissions by 2050 [73]. Several solutions have been suggested to achieve efficient CCS systems, such as biomineralization and biocatalytic gas-liquid membrane contractors for industrial applications. During biomineralization, CO_2 reacts with minerals such as Ca^{+2} or Mg^{+2} and is stored in the form of mineral carbonates [73,74]. A psychrophilic and halotolerant α-CA isolated from *Photobacterium profundum* SS9 (*Ppr*CA) was found to be a potential accelerator of CO_2 biomineralization in the presence of Ca^{+2} ions under alkaline conditions. *Ppr*CA also exhibits efficient CO_2 hydration in both acidic and alkaline pHs [75]. On the other hand, Fu *et al.* recently designed MemZyme technology in which mixture of commercially available CAs (a CA from *Desulfovibrio vulgaris* from Codexis, Inc*,* and bovine CAs from Sigma-Aldrich) were embedded in an ultrathin nanoporous silica membrane for efficient CO_2 sequestration [76].

2 Aims

In this thesis, we aimed to provide structural insights into enzymes found in thermophilic organisms, identify their thermostability contributing factors, and evaluate their thermostability by comparative analysis with other enzymes of the same class. The studied enzymes were chosen based on the availability of their native/mutant form, biochemical characterization data, and their potential significance to biotechnology towards green chemistry and sustainable development.

The specific aims are given below:

- To determine and analyze the three-dimensional structure of a β-glucosidase from the thermophilic fungus *Chaetomium thermophilum* (*Ct*BGL).
- To determine and analyze the three-dimensional structure of an AA9 LPMO from the thermophilic fungus, *Thermoascus aurantiacus* (n*Ta*AA9A).
- To determine and analyze the three-dimensional structure of a superoxide dismutase from the thermophilic fungus *Chaetomium thermophilum* (*Ct*SOD).
- To determine and analyze the three-dimensional structure of a carbonic anhydrase from the thermophilic bacterium *Caloramator australicus* (γ-*Ca*CA).

3 Materials and Methods

3.1 Expression and purification

3.1.1 *Ct*BGL expression and purification

Chaetomium thermophilum is a spore-forming fungus, generally found in the upper layer of the soil growing at rotten organic compounds, bovine dung, and compost. A β-glucosidase from *Chaetomium thermophilum (Ct*BGL) was expressed in *Pichia pastoris* GS115 cells and purified by ion-exchange chromatography. The purity of the protein was assessed by SDS-PAGE [77]. *Ct*BGL activity was determined with salicin using the Miller's method. The purified recombinant *Ct*BGL was a 119 kDa glycoprotein with optimum catalytic activity at pH 5.0 and 60 °C [78].

3.1.2 n*Ta*AA9A expression and purification

Thermoascus aurantiacus is also a spore-forming fungus, generally found in hot springs and decaying organic debris. *T. aurantiacus* strain was isolated from fresh horse dung. n*Ta*AA9A was purified from *T. aurantiacus* (native *Ta*AA9A) by ion-exchange chromatography on a DEAE-Sepharose column (GE Healthcare, Chicago, IL, USA), followed by gel filtration on an Enrich SEC650 column (Bio-Rad, Hercules, CA, USA). The concentration and purity of n*Ta*AA9A were measured using Lowry's method and SDS-PAGE, respectively. The band of interest of purified n*Ta*AA9A was cut out after visualization on SDS-PAGE [77,79–81]. The amino acid sequence of the excised n*Ta*AA9A protein band was determined by LC-MS/MS using nano-LC combined with a mass spectrometer [82].

3.1.3 *Ct*SOD expression and purification

Superoxide dismutase from *Chaetomium thermophilum* (*Ct*SOD) *cz1* gene was expressed in *P. pastoris* GS115 cells and purified with a single ion-exchange chromatography step. The purity of the recombinant *Ct*SOD was assessed using SDS-PAGE [77]. *Ct*SOD activity was measured using the method of Stewart and Bewley [83].

3.1.4 γ-*Ca*CA expression and purification

Caloramator australicus is an anaerobic, gram-positive, sporogenic, rod-shaped, thermophilic bacterium. It has an optimum growth pH and temperature of 7 and 60 °C, respectively. It has been reported to be widely distributed on the globe, especially in heated subsurface aquifers in Australia and India, and in methanogenic sludges. A putative sequence with 95% sequence identity to the γ-type carbonic anhydrase from *Caloramator australicus* (γ-*Ca*CA) was obtained by 16S microbial community profiling at three sites from the geothermal spring of Polichnitos (Lesvos, Greece). The sequence was amplified by PCR, cloned, and expressed in *E. coli* as a 6-His tagged protein. The PCR-amplified sequence was cloned into the T7 expression vector to generate a 6xHis-tag plasmid. This plasmid was used to transform the *E. coli* expression host. The protein was purified using immobilized metal ion affinity chromatography on a Ni-NTA-Sepharose affinity column, and the purity was determined using SDS-PAGE [77]. The 6xHis-tag was left on while purification and crystallization.

3.2 Crystallization

The studied enzymes have not been crystallized before, therefore at least four commercial crystallization screens of 96 conditions each were initially setup to establish promising crystallization conditions for each enzyme. These conditions were further refined, and crystals were improved in size and quality suitable for X-ray data collection.

3.2.1 *Ct*BGL crystallization

The purified *Ct*BGL protein solution was concentrated to ~10 mg/mL using Amicon Ultra Centrifugal Filters (10,000 MW cut-off) (Millipore, MA, USA) in 10 mM HEPES–NaOH, pH 7.0 buffer. Hanging-drop vapor diffusion method was employed to obtain *Ct*BGL crystals at 16 °C using a well solution of 35–45% v/v MPD (Sigma-Aldrich, St. Louis, MO, USA). The drops were prepared by mixing 2 µL of the protein solution with an equal volume of the well solution. *Ct*BGL crystals grew as octahedra (Figure 2) to a maximum size of approximately 0.06 × 0.06 × 0.08 mm^3 within a period of 1 month.

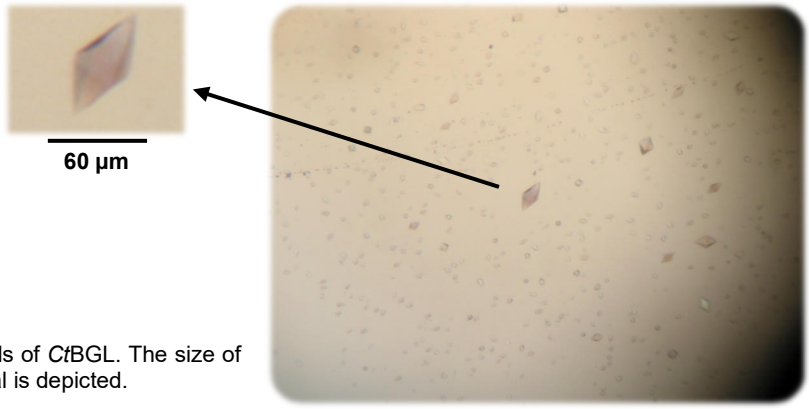


Figure 2. Crystals of *Ct*BGL. The size of the largest crystal is depicted.

3.2.2 n*Ta*AA9A crystallization

n*Ta*AA9A protein concentration of up to 10 mg/mL in buffer was achieved with 10 mM NaOAc, 0.002% NaN_3 w/v, and pH 4.8. The vapor-diffusion hanging drop method was used to create crystals at 16 °C using a well solution of 0.2 M ammonium sulfate, 0.1 M HEPES-NaOH pH 7.5, and 25% w/v PEG 3350. The drops were prepared by mixing 2 µL of the well solution with 2 µL of the protein solution and equilibrated against 0.8 mL of reservoir solution. n*Ta*AA9A crystals started to appear within a week in the form of clusters. Although single crystals were difficult to be isolated from the cluster, X-ray data collection was performed (Figure 3).

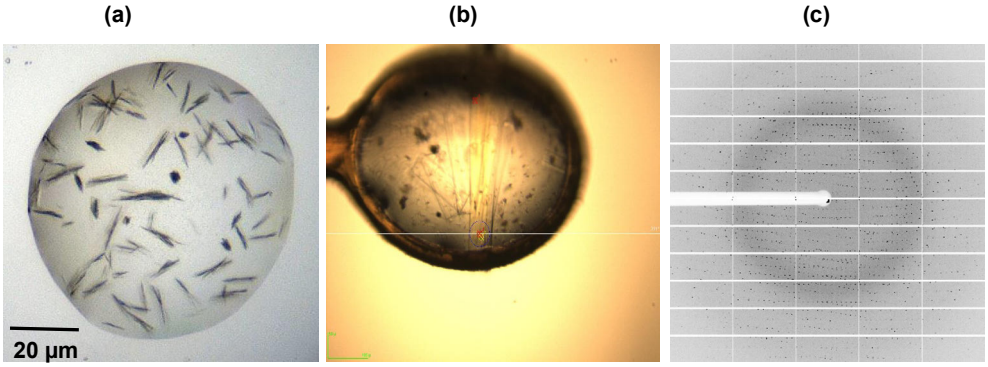


Figure 3. **(a)** n*Ta*AA9A crystals, **(b)** crystal mounting, and **(c)** X-ray diffraction image.

3.2.3 *Ct*SOD crystallization

Purified *Ct*SOD was concentrated to 10.5 mg ml^{-1} by ultra-filtration using AmiconR Ultra Centrifugal Filters (10 000 MW cut-off) (Millipore) in 20 mM Tris-HCl buffer (pH 8.2). *Ct*SOD crystals were grown using the hanging-drop vapor-diffusion

method at 16 °C using NaK phosphate 1.2-1.4 M, pH 8.2, as the precipitant. The drops were prepared by mixing 2 µL of the well solution with 2 µL of the protein solution and equilibrated against 0.8 mL of reservoir solution. Single fine-quality crystals were obtained and mounted for X-ray data collection.

3.2.4 γ-*Ca*CA crystallization

γ-*Ca*CA protein was concentrated to 14 mg/ml in buffer HEPES 10 mM (pH 7.0), NaCl 150 mM, and NaN_3 0.002 w/v, before crystallization. Crystals were obtained in the presence of 20% PEG 4K and 0.2 M sodium formate using the hanging-drop vapor diffusion method at 16 °C. The drops were prepared by mixing 2 µL of the well solution with 2 µL of the protein solution and equilibrated against 0.8 mL of reservoir solution. The obtained crystals were rod-like in shape (Figure 4).

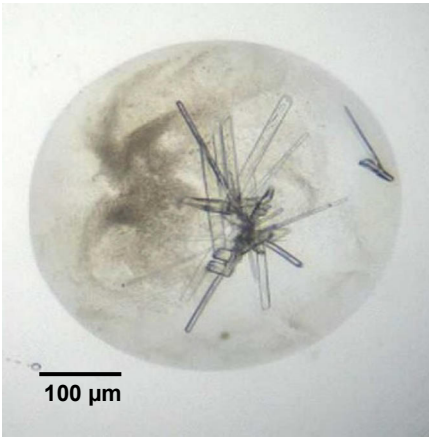


Figure 4. Typical crystals of γ-*Ca*CA.

3.3 Data collection and processing

3.3.1 *Ct*BGL data collection and processing

*Ct*BGL X-ray diffraction data were collected at EMBL-Hamburg (c/o DESY) on the X13 (λ =0.8123 Å) beamline using a MARCCD detector. No extra cryoprotectant was required because the crystallization's MPD content provided the required cryoprotection. A full dataset of one hundred and fifty images was collected from single crystal at cryogenic (100 K) temperature with a rotation range of 0.45° and exposure time of 2 s per image. XDS [84] was used for data processing and scaling. *Ct*BGL crystals were found to belong to the tetragonal space group $P4_12_12$ with unit cell dimensions of a=b=121.9 Å, and c=264.9 Å. Assuming one molecule in the crystallographic asymmetric unit, the Matthews coefficient V_M [85], is 5.4 A^3/Da, corresponding to a solvent content of ~77% .

3.3.2 n*Ta*AA9A data collection and processing

n*Ta*AA9A diffraction data up to 1.86 Å resolution were collected at ESRF (Grenoble, France) facility at cryogenic temperature (100 K) with 10% v/v glycerol

present as a cryoprotectant. Later, the resolution of the diffraction data was extended up to 1.36 Å resolution at EMBL-Hamburg (beamline P13 at the PETRA III ring) at wavelength (λ) of 0.9762 Å. A full dataset of 3600 images was collected from a single crystal at cryogenic (100 K) temperature using a rotation range of 0.1° and exposure time of 0.04 s per image. The XDS package [84] was used for data processing, and AIMLESS from the CCP4 suite [86,87] was used to scale the processed data. n*Ta*AA9A crystals were found to belong to the space group $P2_12_12_1$ with unit cell dimensions of a=37.7 Å, b=64.2 Å, and c=88.5 Å. Assuming one molecule in the asymmetric unit, the Matthews coefficient V_M [85] was 2.13 A^3/Da, corresponding to ~42% of solvent content.

3.3.3 *Ct*SOD data collection and processing

*Ct*SOD X-ray diffraction data were obtained on station X11 (λ =0.8047 Å) of the EMBL-Hamburg (c/o DESY). A single crystal flash-cooled at cryogenic temperature (100 K) using 20% (w/v) glycerol as a cryoprotectant was used for data collection. A flat-panel detector MAR555 with a crystal-to-detector distance of 350 mm was used to gather 200 images with a rotation range of 0.5° per image and exposure time of 4 s per image. The data were processed and scaled using the XDS [84] package. The crystal was found to belong to the $P6_1$ space group with unit-cell measurements of a=b=89.11 Å, and c=310.37 Å. Assuming eight molecules in the asymmetric unit, the Matthews coefficient, V_M [85] is 2.31 A^3/Dalton corresponding to a solvent content of ~47%.

3.3.4 γ-*Ca*CA data collection and processing

γ-*Ca*CA diffraction data were collected up to 1.11 Å resolution on the P13 beamline at PETRA III synchrotron in DESY, Hamburg. A single γ-*Ca*CA crystal at cryogenic temperature (100 K) was used to collect 7200 diffraction images with a rotation range of 0.05° and exposure time of 0.01 s per image. The crystal was found to belong to the $P2_12_12_1$ space group with unit cell dimensions of a=57.02 Å, b=82.84 Å, and c=98.50 Å. Assuming three molecules in the asymmetric unit, the Matthews coefficient, V_M [85], was 1.99 A^3/Dalton , corresponding to ~39% solvent content.

3.4 Structure determination and refinement

3.4.1 *Ct*BGL structure determination and refinement

Phaser [88], as implemented in PHENIX 1.15.2_3472 [89], was used to obtain the initial phases through molecular replacement. The crystal structure of *Aspergillus*

aculeatus β-glucosidase in complex with castanospermine (PDB ID 4IIF: sequence identity 61.5%) was used as a search model. Side chains were pruned with Sculptor [90] based on sequence alignment considerations. The search was initially performed under the assumption of two molecules in the asymmetric unit, but no solution was obtained. Consequently, the search was restricted to one molecule (high solvent content), and a single solution in space group $P4_12_12$ was found based on the statistics for one molecule (TFZ = 30.9). For refinement, maximum likelihood as the target function was implemented in PHENIX 1.15.2_3472 [89], using simulated annealing (1000 K). The refinement was alternated with model visualization and rebuilding using Coot 0.8.9 [91]. Since the data was of low resolution, structural overfitting to data was avoided by using tight restraints. The R_{free} (comprising 5% of the reflections excluded from the refinement) value was monitored during the refinement [92]. Rebuilding at places with poor electron density and ambiguous atomic positions was performed using high-resolution structures. Data collection and refinement statistics are summarized in Table 6.

3.4.2 n*Ta*AA9A structure determination and refinement

For structural determination of n*Ta*AA9A, the 3D structure of the recombinant *T. aurantiacus* AA9A (r*Ta*AA9A) was used as a search model (PDB ID: 2YET). Phaser [88] was used to obtain the initial phases with molecular replacement. A single solution with a TFZ score of 33.8 for one molecule in the asymmetric unit was obtained. PHENIX (v. 1.19.2) [89] was used for refinement with simulated annealing and maximum likelihood as the target function. R_{free} (calculated using 5% of the data excluded from the refinement) value was monitored during the refinement. Water molecules were also added to the structure using automated procedures as implemented in PHENIX [89]. Coot 0.9 [91] was used for visualization, rebuilding and displaying electron density maps. A glycosylation site was identified at Asn138 based on the electron density difference map. Data collection and refinement statistics are summarized in Table 6.

3.4.3 *Ct*SOD structure determination and refinement

*Ct*SOD initial phases were obtained through molecular replacement using Phaser [88] as implemented in PHENIX [89]. Based on a sequence identity of ~70%, the crystal structure of *Saccharomyces cerevisiae* Cu,Zn-SOD (PDB ID: 1F1G) was used as a template for molecular replacement. Sculptor [90] was used to construct the search model and prune the side chains according to the sequence alignment. Assuming 8 molecules in the asymmetric unit, a solution with a Z-score of 62.4 with all 8 molecules located was obtained. The initial model was subjected to automated

model building using ARP/wARP 7.1 [93]. The model was inspected and subsequently refined with REFMAC [94] in the CCP4 suite [87]. Further refinement was performed using PHENIX 1.17.1-3660 [89] with simulated annealing and maximum likelihood as the target function. Water molecules were also added to the structure using automated procedures as implemented in PHENIX [89]. R_{free} (5% of the reflections) calculations were carried out to monitor the progress of the refinement. Coot 0.9 [91] was used for visualization, rebuilding, and displaying the electron density maps. Data collection and refinement statistics are summarized in Table 6.

3.4.4 γ-*Ca*CA structure determination and refinement

Initial phases were obtained through molecular replacement using Phaser [88] as implemented in PHENIX 1.20.1 [89]. The carbonic anhydrase from *Clostridioides difficile* (PDB ID: 4MFG; sequence identity 57.7 %) was used as a template. Sculptor [90] was used to prune the side chains of the template to develop a suitable search model for MR. A unique initial solution was obtained with a TFZ score of 18.4. This solution was subjected to further refinement iterations using simulated annealing in PHENIX [89] and maximum likelihood as the energy target. Coot 0.9 [91] was used for visualization and rebuilding of the structure. R_{free} was based on 5% of the total reflections. During the refinement iterations, key values such as $R_{free,}$ Ramachandran plot, and difference map peaks were monitored. When the R_{free} stopped decreasing in refinement iterations, anisotropic refinement was employed for non-hydrogen protein atoms leading to further decrease of the R_{free} to 0.188. Data collection and final refinement statistics are shown in Table 6.

3.5 Structure validation and analysis

All refined structures (*Ct*BGL, n*Ta*AA9A, *Ct*SOD, and γ-*Ca*CA) were validated with tools implemented in PHENIX [89], Coot 0.9 [91], and MolProbity [95]. Privateer [96] was used for testing the stereochemistry and conformation of the sugars in glycosylation sites. UCSF Chimera 1.13.1 [97] was utilized for structure-based sequence alignment and for making high-resolution figures. PDBeFold [98] was employed to identify structural and sequence similarities between the query and the deposited PDB structures. ESPript 3 [99] was used for rendering sequence similarities and secondary structure details from aligned sequences. ESBRI (http://bioinformatica.isa.cnr.it/ESBRI/) was used for calculating the salt bridges. PDBePISA [100] was used for calculating solvent-accessible surface (SAS) area, interface area, and H-bonds. ExPASy server (http://web.expasy.org/protparam/) [101] was used for calculating the values of the charged residues. 2P2I Inspector

[102] was used for identifying protein-protein interactions at the interfaces. FASTA sequences and PDB structures were accessed from RCSB PDB [103] for comparison and analysis.

Table 6. X-ray data collection and refinement statistics for thermozymes used in this thesis.

Data Collection and Processing	***Ct*BGL**	**n*Ta*AA9A**	***Ct*SOD**	**γ-*Ca*CA**
Beamline	EMBL-DESY X13	EMBL-DESY P13	EMBL-DESY X11	EMBL-DESY P13
Wavelength (Å)	0.8123	0.9762	0.8047	0.8266
Resolution range (Å)	20.0-2.99 (3.11-2.99)	44.63-1.36 (1.41-1.36)	38.59-1.56 (1.64-1.56)	49.25-1.11 (1.15-1.11)
Space group	$P4_12_12$	$P2_12_12_1$	$P6_1$	$P2_12_12_1$
Unit cell dimensions a, b, c (Å) α, β, γ (°)	121.9,121.9, 264.9 90, 90, 90	37.7, 64.2, 88.5 90, 90, 90	89.1, 89.1, 310.37 90, 90, 120	57.0, 82.8, 98.5 90, 90, 90
No. of molecules in a.u	1	1	8	3
No. of measurements	222508 (22974)	559476 (22595)	975382 (83309)	2397040 (228028)
Unique reflections	41051 (4513)	43359 (2631)	194568 (29651)	184092 (17915)
Completeness (%)	99.3 (98.6)	92.7 (58.5)	98.0 (92.2)	99.9 (99.7)
Mean I/sigma (I)	7.8 (1.4)	9.1 (0.8)	15.0 (1.15)	12.2 (0.8)
Wilson B-factor (Å^2)	46.2	22.1	26.6	14.5
R_{meas}	0.287 (1.46)	0.218 (2.99)	0.008 (0.15)	0.086 (4.01)
$CC_{1/2}$	0.978 (0.492)	0.998 (0.251)	0.999 (0.416)	0.999 (0.416)
Refinement				
No. of reflections used	38861	43220	194552	183904
R_{work}/R_{free}	0.201/0.249	0.151/0.185	0.172/0.204	0.166/0.188
Number of protein residues (molecule)	836	228	1256	173
Ligands	8	4	38	17
RMSD in bonds (Å)	0.007	0.005	0.009	0.005
RMSD in angles (°)	0.94	0.87	1.05	0.89
Ramachandran favoured (%)	92.8	99.1	96.9	97.4
Outliers (%)	1.0	0.0	0.0	0.0
Clash score	10.4	2.5	5.3	5.5
Average B-factor (Å^2)	50.4	17.6	23.8	21.3
PDB IDs	**6SZ6**	**7Q1K**	**6ZS1**	**9QEV**

4 Results and Discussion

4.1 Crystal structure of *Ct*BGL

*Ct*BGL is a member of the GH3 family. As in other members of the GH3 family, *Ct*BGL also exhibits three distinct and conserved domains in its structure (Figure 5): a catalytic triose phosphate isomerase (TIM) barrel-like domain, an α/β sandwich domain and a fibronectin type III (FnIII) domain. In most cases, the TIM barrel domain along with the α/β sandwich domain houses the active site. The function of the FnIII domain has not yet been established, but it has been suggested that it stabilizes the incomplete TIM-barrel domain [104].

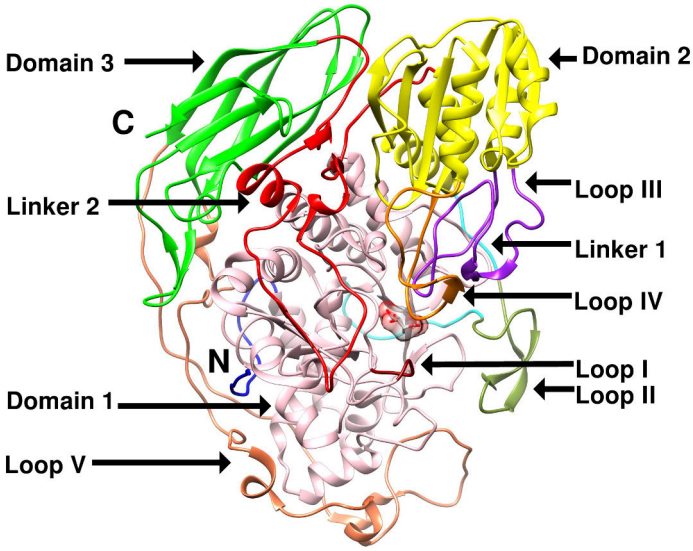


Figure 5. *Ct*BGL overall structure in ribbon representation. The domains, linkers, and loops are shown in different colors and labelled: Domain 1 (pink), Domain 2 (yellow), Domain 3 (green), Linker 1 (cyan), Linker 2 (red), Loop I–V (dark red, olive drab, purple, orange, and coral, respectively). The figure was created using UCSF Chimera [97].

Domain 1: In *Ct*BGL 3D structure, TIM domain (Domain 1) ranges from residues Trp51 to Val351. Domain 1 is the longest domain in *Ct*BGL structure, with approximately 300 residues, and it hosts the nucleophilic catalytic residue (Asp287).

Linker 1: Domain 1 is followed by the first linker region, which ranges from residues Arg352 to Glu394. This linker region connects Domain 1 with Domain 2.

Domain 2: Domain 2 is an α/β sandwich domain that consists of residues Asp395 to Tyr595. This domain is conserved among all the GH3 family enzymes. *Ct*BGL Domain 2 hosts loop III and loop IV, that comprises of residues Pro434–Val467 and Asn514–Asn536, respectively. Cysteine residues (Cys442 and Cys447) are conserved in loop III and play a role in loop III folding and stabilization by forming disulfide bridge. Loop IV hosts the catalytic acid/base residue (Glu517) and along with loop III forms a part of the active site. Residues such as Asp444, Ser458, and Glu517 of loop IV are conserved and point towards the active site.

Linker 2: Domain 2 is followed by a second linker region ranging from residues Gly596 to Thr663. This linker region connects Domain 2 to Domain 3 and plays a role in stabilizing loops I and IV.

Domain 3: Domain 3 (FnIII domain) is the third and last domain of the *Ct*BGL structure. It ranges from Phe664 to Asn867 residues that form a beta sandwich with nine β-strands. Domain 3 also includes an insertion region ranging from residues 674–770. This extended insertion region, also known as loop V, encompasses Domain I. Several conserved aromatic residues, namely, Tyr716, Tyr718, Tyr733, and Phe740, are found in loop V. An asparagine residue (Asn720) is found to be *N*-glycosylated and conserved in loop V.

4.1.1 Active site

The active site of *Ct*BGL is located at the interface between Domains 1 and 2. The structure of the active site resembles that of a shallow pocket buried mainly in Domain 1 but also surrounded by Domain 2 and the linker residues. There are two conserved catalytic residues present in members of the GH3 family at the active site: a nucleophilic catalytic residue (Asp287) and an acid/base catalytic residue (Glu517). Both catalytic residues are opposite in position but point towards each other in the shallow cavity of the active site. Asp287 is located at the N-terminus of the TIM-barrel domain, while Glu517 is located in loop IV of Domain 2 (Figure 6). A β-D-glucose (BGC) molecule was found bound at the active site. The presence of the BGC molecule was validated using *Fo–Fc* difference electron density maps. No BGC was used during the crystallization or soaking process; however, its presence in the structure suggests that it may have originated from the growth medium. An additional electron density was observed at the active site next to the BGC, which

may suggest a bound buffer molecule, MPD from the crystallization mother liquor, or a partially bound glucose molecule.

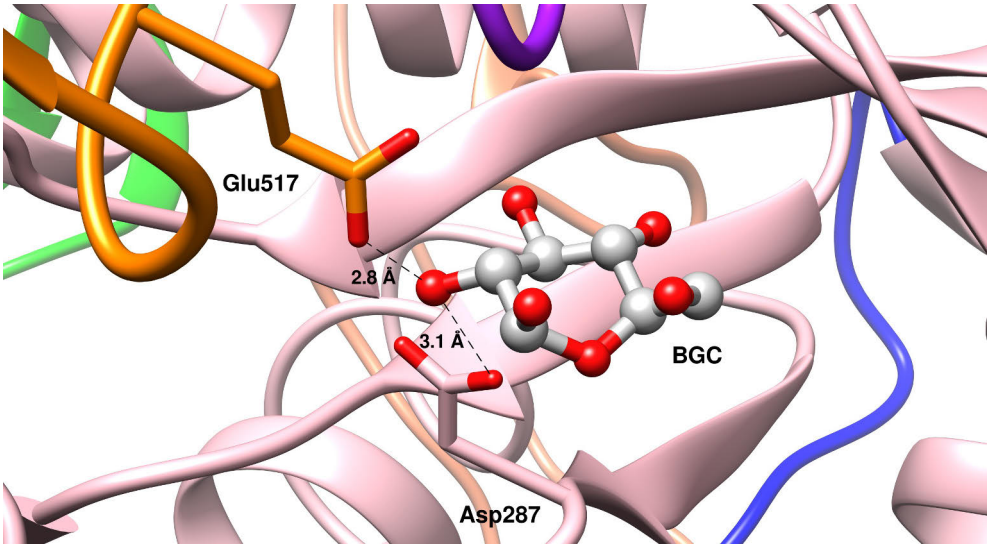


Figure 6. Active site of *Ct*BGL. The catalytic residues Glu517 (acid/base) and Asp287 (nucleophilic) are shown in orange and pink colors, respectively, by heteroatom feature. BGC is shown at the center of the active site in 'ball & stick' depiction. Hydrogen bonds formed by the catalytic residues are depicted with distances. The figure was created using UCSF Chimera [97].

4.1.2 Glycosylation

Asparagine (Asn) residues in *Ct*BGL are found involved in glycosylation and are conserved at the glycosylation sites. In total, 10 glycosylation sites were observed in the *Ct*BGL structure. All these sites were present on one face of the molecule, that is, the face bearing the active site. The key sugars involved in glycosylation were *N*-acetyl-β-D-glucosamine (NAG), β-D-mannose (BMA), and α-D-mannose (MAN). In *Ct*BGL, 27 glycans were found, ranging in length from a single sugar to multiple sugars bonded in chains. The longest glycan chain was observed in Domain 1 of the *Ct*BGL structure at the corresponding Asn329 residue, with eight sugar molecules. The largest *N*-glycan moiety displayed pi-sigma interactions and H-bonds with Domain 1 and Domain 3 residues.

Glycans participate in hydrogen-bonding interactions at the dimer interface and may help stabilize crystal packing [105]. Protein-glycan interactions are important for the binding affinity between cellulose and aromatic compounds [106–108]. Two glycosylated Asn residues, which provide protein-glycan interactions between the two chains, were found conserved at the interface of the compared β-glucosidases.

In the *Ct*BGL dimer structure, only one of these two Asn residues is present, Asn531, which is glycosylated with two monosaccharide molecules. The missing Asn residue may indicate a higher instability of the *Ct*BGL crystals and reduced intermolecular interactions in the crystal lattice. Indeed, the *Ct*BGL crystals were found to be very fragile and easily dissolved, posing a serious challenge during mounting.

4.1.3 *Ct*BGL thermostability

*Ct*BGL was thermally stable at 50 °C and retained half of its activity after incubation at 65 °C for 55 min. This enzyme also retained 29.7% of its activity after incubation at 70 °C for 10 min. For most other thermophilic fungal β-glucosidases, *Ct*BGL exhibits comparable thermostability, as previously reported [78]. In contrast, β-glucosidase from *A. fumigatus* (*Af*βG) was found to be highly thermostable and able to retain most of its activity for at least 19 h at 65 °C [78,109].

4.1.4 Structural comparison

The *Ct*BGL structure was compared with the structures of glucosidases from *Neurospora crassa* (*Nc*Cel3A, PDB ID: 5NBS), *Aspergillus aculeatus* (*Aa*Bgl1, PDB ID: 4IIF), *Rasamsonia emersonii* (*Re*Cel3A, PDB ID: 5JU6) and *Hypocrea jecorina* (*Hj*Cel3A, PDB ID: 3ZYZ) for determining structural differences and unique features. Matchmaker from UCSF Chimera [97] was employed for this purpose. Structural comparison revealed key differences in structural domains, loop and the linker region, active site, and the glycosylation sites.

Domain 1 was found to be highly conserved among the compared structures, with slight differences. The collapsed TIM-barrel model of Domain I is vital for the proper accession of the active site. It was found that near the active site, the second β-strand (Gly87 to Thr89) of the barrel was much shorter and antiparallel, which creates an active site much wider and accessible when compared with other GH3 enzyme structures with complete TIM-barrel fold.

In sequence alignment with *Ct*BGL, *Aa*Bgl1[110] showed missing residue at positions 669–675 in Domain 3 of the structure. This is why some structural changes are observed in the respective positions of both structures, especially in *Ct*BGL.

Linker region 1 is 42 residues long in *Ct*BGL when compared with *Hj*Cel3A [111], where the linker region is only 18 residues long. The *Nc*Cel3A [112] linker region is also 42 residues long (341–383) and contains a distinct 25 residue long insertion (351–376) region. This insertion region has also been termed as the hydrophobic linker region and/or loop II, which is rich in aromatic residues and plays a role in the action of the enzyme in organic solvents. Loop II region also contains residues such as Phe366, Trp367 and Trp377 that line up along one side of the

substrate binding site and help in stabilizing by interactions with *N*-glycans from neighbouring glycosylated Asn residues e.g., Asn57 in *Nc*Cel3A [113].

Domain 3 hosts loop V and differences were observed in this loop among the compared structures. Loop V was found present in *Ct*BGL, *Nc*Cel3A, *Re*Cel3A [104], and *Aa*Bgl but missing in *Hj*Cel3A. Tyrosine and phenylalanine residues such as Tyr716, Tyr718, Tyr733, and Phe740 were found highly conserved in loop V among the compared structures. However, Tyr716 and Tyr733 were missing in *Hj*Cel3A because of the missing loop V. In *Re*Cel3A and *Aa*Bgl1, Tyr718 is replaced with a Trp residue. Moreover, Phe740 is substituted by Tyr and His residues in *Re*Cel3A and *Aa*Bgl1, respectively. Missing modelled residues were also observed from 669–675 in Domain 3 of the *Aa*Bgl1 structure owing to high flexibility. This part of the structure, however, is visible in *Ct*BGL.The catalytic residues in *Ct*BGL are Asp287 (nucleophilic) and an Glu517 (acid/base). The corresponding catalytic residues in *Nc*Cel3A, *Re*Cel3A, *Aa*Bgl1 and *Hj*Cel3A were positioned as Asp276/Glu505, Asp277/Glu505, Asp280/Glu509, Asp236/Glu441, respectively. *Ct*BGL's number of glycosylation sites were compared with other proteins to evaluate the degree of glycosylation. *Re*Cel3A and *Hj*Cel3A exhibited the highest and lowest number of glycosylation sites i.e., 16 and 2, respectively [104]. The number of glycosylation sites observed in *Nc*Cel3A and *Aa*Bg11 was 7 and 8, respectively. *Ct*BGL displayed a high number of GlcNAc-type *N*-glycans, whereas *Re*Cel3A and *Aa*Bg11 showed high mannose-type *N*-glycans.

4.1.5 Thermostability contributing factors

The SAS values for the compared β-glucosidases, including *Ct*BGL, were similar, i.e. ~28100 Å^2. However, the structure of *Hj*Cel3A from the mesophilic fungus *Hypocrea jecorina* had the lowest SAS (22812 Å^2), owing to the smaller number of residues and lack of loop V.

Charged residues in protein structures may contribute to thermostability [114]. A structurally stable configuration could possibly be achieved when both negatively charged (Asp and Glu) and positively charged (Arg and Lys) residues are proportionately optimal in a protein structure. The number of charged residues showed significant differences when compared with other glucosidase structures. In particular, *Hj*Cel3A showed a reduced number of positively and negatively charged residues, despite its high thermostability with an optimum temperature of 90 °C and an unfolding temperature of about 88 °C depending on enzyme concentration [115]. These numbers were increased for β-glucosidase from other thermophilic and mesophilic β-glucosidases. In addition, *Af*βG [105], despite its high thermostability, is also characterized by a similar charged residue content, suggesting that the strength of individual ion-pair interactions may play a major role.

Glycosylation may also contribute to enzyme thermostability by promoting interactions with amino acid residues. Glycosylation has been shown to improve solubility, reduce agglomeration, and increase thermal stability of proteins [116,117].The precise mechanisms by which glycosylation affects the overall structure and function of proteins are poorly understood. Comparative analysis of protein structures revealed that *N*-glycosylation does not result in local or global structural changes. At certain temperatures, *N*-glycosylated proteins are typically less stable and aggregate more easily than glycosylated proteins. Therefore, glycosylation has been proposed to enhance the solubility, stability, and functionality of enzymes [114,118]. The same melting temperature (74.0 ± 0.2 °C) was found by thermal stability measurements of *Hj*Cel3A samples with varying levels of *N*-glycosylation, indicating that the effect of glycosylation is case-specific and that other factors may be at play [119]. Interestingly, *Tn*Bgl3B [120] lacks glycosylation but shows significant thermostability. This is probably due to the large number of charged residues. It cannot be denied that the moderate glycosylation of *Ct*BGL contributes to the limited stability of the enzyme compared to other more thermostable β-glucosidases characterized by extensive glycosylation. However, further studies are needed to better understand the role of glycosylation in β-glucosidases and protein stability.

4.2 Crystal structure of n*Ta*AA9A

The native *Ta*AA9A (n*Ta*AA9A) structure was refined to 1.36 Å resolution with R_{work} and R_{free} of 0.151 and 0.185, respectively. n*Ta*AA9A structure was determined as a single polypeptide chain of 227 amino acid residues. The structure was found to have 1761 protein atoms, 307 water molecules, 1 oligosaccharide (2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose), 1 Cu^{2+} ion, 1 Na^{+} ion, 2 NAG (N-acetyl-glucosamine), and 2 glycerol molecules.

The first and last amino acid residues in the n*Ta*AA9A protein sequence are His and Gly, respectively. However, we see aberrations in these amino acid residues. The first amino acid (His1) is methylated as 4-methyl-histidine (Hic as a 3-letter abbreviation in PDB) while the last amino acid residue (Gly) was not visible in the electron density map and thus could not be modelled due to its flexibility. The methylation of His1 could also be observed in other LPMOs, though its role is still unclear; it has been assumed to offer protection to the structure against oxidative damage. LPMOs with or without this post-translational modification of His1 methylation are catalytically active. Generally, LPMOs produced in *P. pastori*s lack His1 methylation [121,122].

In n*Ta*AA9A structure, the Cu^{2+} ion was found present at the N-terminus with a temperature factor of 13.2 $Å^2$ and occupancy of 1.0, suggesting a well-defined tightly

bound ion. The Cu^{2+} ion has Hic1 and His86 as coordinating residues. The Cu^{2+} ion is also observed in other LPMOs and is involved in the catalytic reaction. The Na^{+} ion is found present at the C-terminus of the n*Ta*AA9A structure surrounded by water molecules. Two molecules of NAG were identified based on an electron density map in the n*Ta*AA9A structure (Figure 7). Both NAG molecules were found adjacent to each other and in close proximity to Asn138 residue, suggesting a glycosylation site on Asn138. Moreover, a glycerol molecule (GOL) was also observed close to the Cu^{2+} ion and Hic1 at the N-terminus.

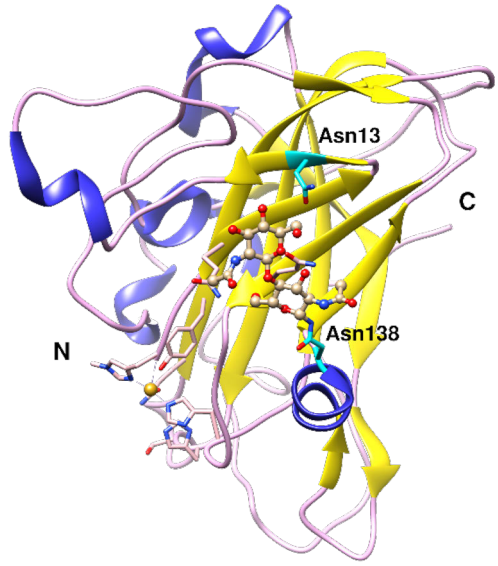


Figure 7. n*Ta*AA9A overall structure in ribbon representation. The secondary structure and ligands are depicted in different colors: alpha helix (blue), beta strands (yellow), coil (pink), Cu^{2+} (goldenrod), Na^{+} (grey), and GOL (green). Two NAG molecules from the glycosylation site are shown in 'ball & stick'. The figure was created using UCSF Chimera [97].

4.2.1 Active site

The active site of n*Ta*AA9A can be observed as a shallow cavity located close to the N-terminus of the structure. The active site was found comprised of Cu^{2+} ion and residues as Hic1, His86, Tyr175, and Gln173 (Figure 8). A GOL molecule was also found pointing towards the active site. His164 and Ser165 were also observed projecting towards the active site.

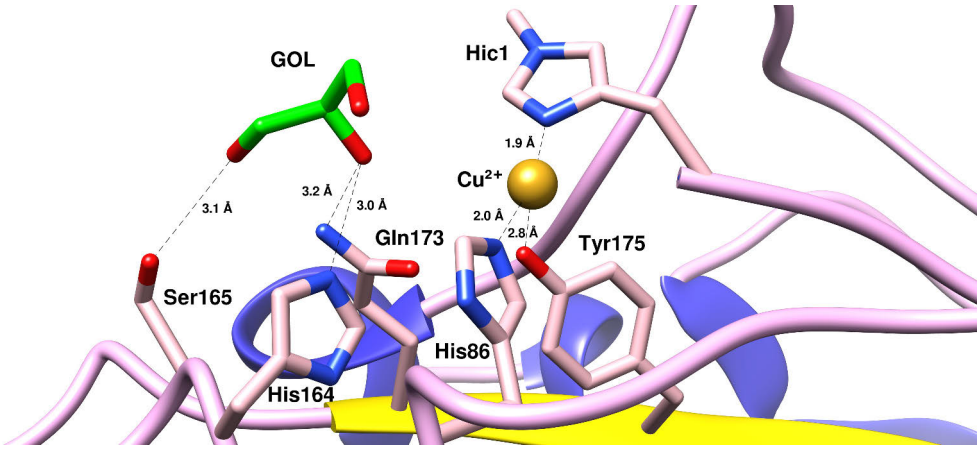


Figure 8. Active site of n*Ta*AA9A structure in ribbon representation. The secondary structure and ligands are depicted in different colors: alpha helix (blue), beta strands (yellow), coil (pink), Cu^{2+} (goldenrod), and GOL (green). Distances between copper and coordinating residues are shown. Hydrogen bonds between bound glycerol and protein residues are depicted. The figure was created using UCSF Chimera [97].

4.2.2 Structural comparison

A structural comparison between the native (n*Ta*AA9A) and recombinant (r*Ta*AA9A) [123] enzymes was performed using UCSF Chimera [97]. Structural superpositioning resulted in an RMSD value of 0.43 Å, indicating subtle differences between the compared structures. However, noticeable deviations (~0.8 – 2.4 Å) were observed in the following regions: 9–13, 25–30, 184–187, 202–203, and 213–217.

Asn138 residue was found glycosylated in both structures. In n*Ta*AA9A, two NAG molecules were successfully modelled, however a third NAG molecule was also found present, but the electron density was not enough to model it. On the other hand, only one NAG molecule was found present in the r*Ta*AA9A structure. The NAG molecules can position towards the shallow cavity of the loop and may interact with Gln5, Gln78, and Asn13 residues (Figure 9). In the case of r*Ta*AA9A, the interaction seems limited to the Gln78 since there was only one NAG molecule. Moreover, the orientation of the Asn138 side chain may also result in different positioning of the glycan moieties, for example, in r*Ta*AA9A, the NAG molecule points outwards and thus exposed to the solvent. The minor differences found in the surface residues may result in different solvent-accessible surface areas in both structures i.e., n*Ta*AA9A (9285 $Å^2$) and r*Ta*AA9A (9416 $Å^2$).

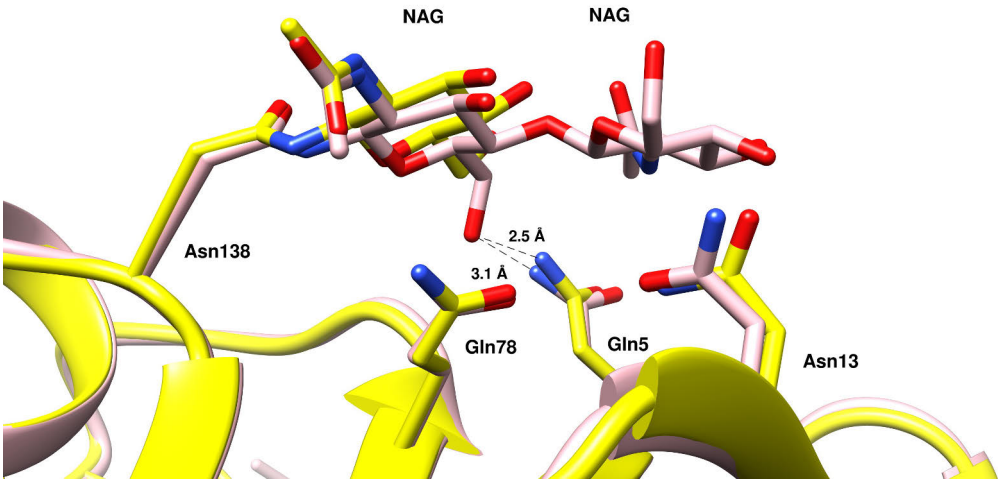


Figure 9. Comparison of the Asn138 glycosylation site between n*Ta*AA9A (pink) and r*Ta*AA9A (yellow). Both structures are depicted in ribbon representation. Distances between NAG molecules and protein residues are depicted. The figure was created using the UCSF Chimera [97].

4.2.3 Thermostability contributing factors

The thermostability of n*Ta*AA9A was determined using high-performance liquid chromatography with a refractive index detector (HPLC-RID). The successfully measured C1 and C4 oxidation activities for n*Ta*AA9A were 0.646 and 0.574 U/mg. However, for r*Ta*AA9A, these values were 0.155 and 0.153 U/mg [123]. n*Ta*AA9A was hydrolyzed with trifluoroacetic acid (TFA) and the reaction product (i.e., gluconic acid) was detected. The experiment established that n*Ta*AA9A exhibits high thermostability as after incubation for 30 min at 60, 70 and 80 °C, the enzyme retained 84.3%, 63.7%, and 35.3% of its activity, respectively. There have already been identified certain factors that may influence the thermostability of AA9A enzymes. For example, the structure of *Ta*AA9A (PDB ID: 3ZUD) led to the identification of a cluster of amino acid residues (Val90, Ser131, Leu134, and Trp141) that may improve thermostability if steered optimally [124]. Using this information, a thermostable variant of *Afu*AA9A was created (*Afu*AA9A_var) and its structure was determined at 1.87 Å resolution. *Afu*AA9A_var was found to have a Tm 7 °C higher than the native enzyme [125]. This improvement in the melting temperature was attributed to the elimination of unfavourable electrostatic interactions in the enzyme. Thermostability improvement in *Afu*AA9A also provides hints and leading clues for improving the thermostability of r*Ta*AA9A.

There are few reports of LPMO thermostability using activity assays because of the challenges in determining their quantitative activity. To date, thermostability of only three LPMOs from non-thermophilic fungi, including *Afu*AA9A from

Aspergillus fumigatus, *Ls*AA9A from *Lentinus similis*, and *Tc*AA9A from *Talaromyces cellulolyticus*, has been measured using differential scanning fluorimetry (DSF), differential scanning calorimetry (DSC), and activity assays. *Afu*AA9A exhibited a Tm of 68–69 °C, whereas *Ls*AA9A demonstrated a thermal inflection T_i of 71.8 °C. However, *Tc*AA9A lost its activity after incubation at 50 °C for 8 h [125–127].

Table 7 shows a comparison of the factors contributing to the thermostability of LPMOs. Computational programs were employed for this purpose, and theoretical measurements were recorded. For *Mc*AA9F [128], the temperature-dependent stability was calculated using the SCooP algorithm, a Gibbs-Helmholtz equation-based program [129]. This program calculates all thermodynamic quantities associated with the two-fold transition of proteins, such as melting temperature (Tm), standard folding enthalpy (H_m) measured at Tm, and standard folding heat capacity (C_p). The findings in the table suggest that n*Ta*AA9A has significant thermostability compared to *Mc*AA9F, as the former has a higher number of charged residues, Val residue percentage, surface area, salt bridges, and melting temperature.

Table 7. Comparative statistics of thermostability parameters in AA9 LPMOs.

Parameter	**n*Ta*AA9A** (7Q1K)	***Mc*AA9F** (7NTL)	***Ao*AA13** (4OPB)	***Ta*GH61** (2YET)	***Cv*AA9_A** (6YDE)	***Afu*AA9A** (6H1Z)
Asp + Glu (-)	19	16	34	19	34	18
Arg + Lys (+)	7	7	14	7	24	10
Pro:Gly	0.84	1.04	0.65	0.84	0.91	0.57
Val (%)	5.3	4.1	4.3	5.3	8.7	4.4
Amino acid residues	228	222	233	228	252	229
SAS (Å^2)	9285.0	9034.0	9265.0	9416.0	9649.0	9363.0
Interface area, Å^2	424.5	651.7	336.0	525.8	512.0	514.0
Intra-chain salt bridges	3	2	4	11	13	3
Instability Index	24.93	34.57	44.87	24.08	30.14	20.20
Melting temperature Tm (°C)	56.1	49.4	58.8	59.9	57.5	57.8

For amino acid, salt bridges, and surface area calculations, ExPASy ProtParam (https://web.expasy.org/protparam/), ESBRI (http://bioinformatica.isa.cnr.it/ESBRI/introduction.html) and PDBePISA (https://www.ebi.ac.uk/pdbe/pisa/) were used, respectively. Tm was calculated using SCooP_v1.0 (http://babylone.ulb.ac.be/SCooP/).

4.3 Crystal structure of *Ct*SOD

The structure of *Ct*SOD was refined to 1.56 Å resolution with R_{work} and R_{free} of 0.172 and 0.204, respectively. The observed RMSDs in bond lengths and bond angles were 0.009 Å and 1.05 Å, respectively. The Ramachandran plot (phi, psi) revealed that

approximately 97% of all residues were in favoured regions, 2% were in allowed regions, and none were in disallowed regions. Twenty-two residues with different conformations were identified. Based on maximum likelihood calculations as implemented in PHENIX [89], the coordinate error was 0.25 Å.

The overall structure of *Ct*SOD comprises eight molecules that are symmetrically aligned opposite each other in the asymmetric unit (Figure 10). Each molecule has a sequence length of 157 amino acid residues. The average B-factors for the eight subunits range between 19.15–22.36 $Å^2$. The average B-factor of the water molecules is 33.8 $Å^2$. A very low RMSD (0.039–0.291 Å) in Cα positions between the eight molecules suggests that each molecule is structurally identical to the other molecules in the asymmetric unit. *Ct*SOD structure, besides amino acid residues, also comprises ions such as 8 Zn^{2+} ions, 5 PO_4^{3-} anions, 15 Na^+ cations, and 4 Cu^{2+} ions. *Ct*SOD structure also includes 4 GOL molecules and 2455 water molecules. No post-translational glycosylation was detected in the *Ct*SOD structure.

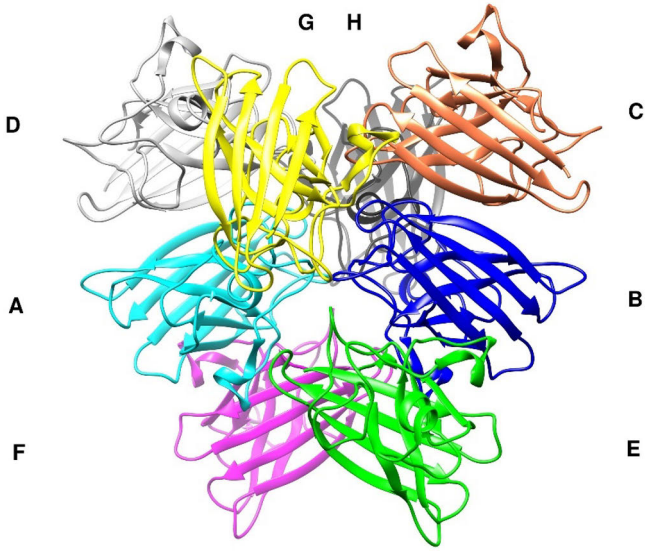


Figure 10. Overall structure of *Ct*SOD. The eight molecules are labelled from A to H and colored as cyan, blue, coral, gray, green, magenta, yellow, and black, respectively. The overall structure is depicted as a ribbon. The figure was created using the UCSF Chimera [97].

A detailed description of one of the eight molecules is shown in Figure 11. *Ct*SOD consists of two short alpha helices (α1: 57–61 and α2: 35–138) and eight β-strands in the structure. The β-strands can be seen as two groups of 4-stranded β-sheets positioned antiparallel and facing each other. These β-strands are labelled as 1a, 2b, 3c, 6d, 5e, 4f, 7g and 8h. The first β-strand (1a) is close to the N-terminus while the last β-strand (8h) is close to the C-terminus of the structure.

A conserved disulfide bond that results in a left handed spiral (LHS) as defined by Thornton [130] was found present between the Cys58 of α1 helix and Cys147 of 8h β-strand. According to the PDBePISA [100], the solvent accessibility of the Cys residues ranges from 7.6 to 12.7 Å^2 amongst the eight subunits of the *Ct*SOD structure [131].

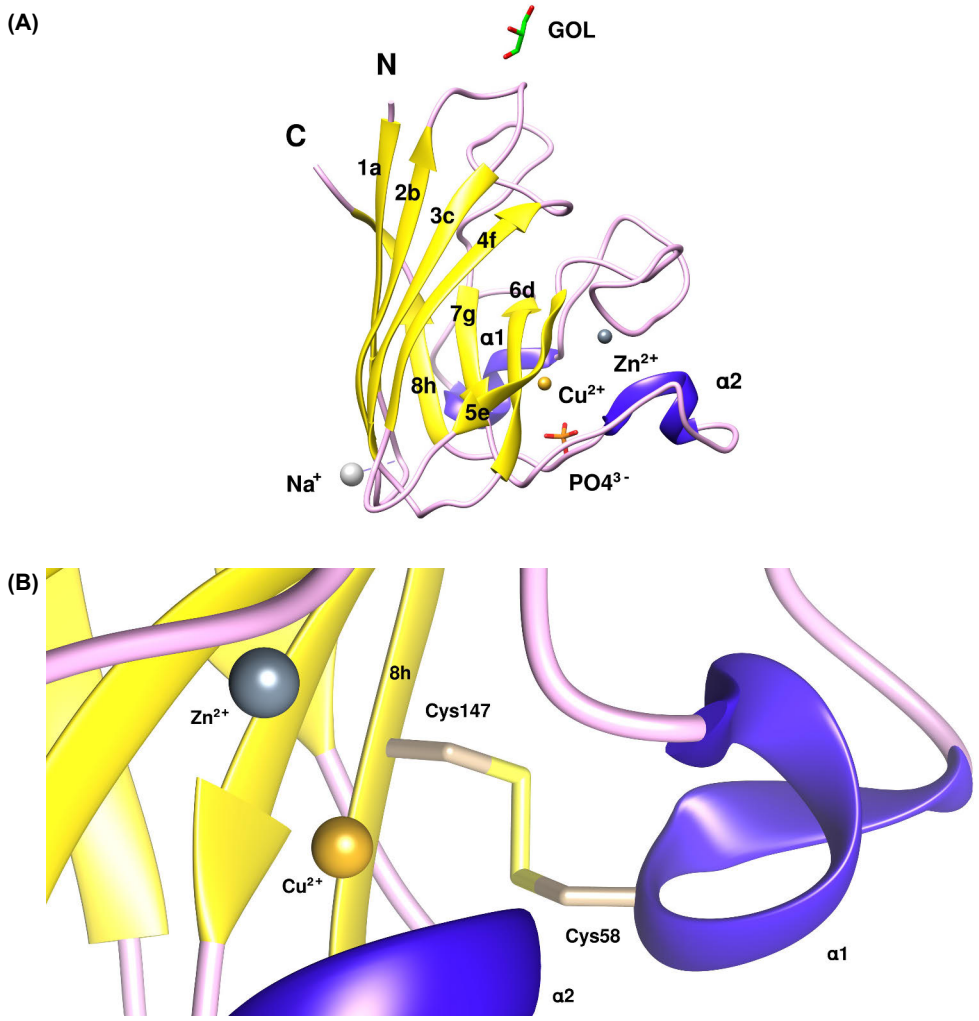


Figure 11. (A) Structure of *Ct*SOD depicted in ribbon presentation. The secondary structure and ligands are shown in different colors: alpha helix (blue), beta strands (yellow), coil (pink), Cu^{2+} (goldenrod), Zn^{2+} (stale grey), GOL (green), PO_4^{3-} (orange), and Na^+ (grey). (**B**) Disulfide bond between Cys58 and Cys147. The figures were created using UCSF Chimera [97].

4.3.1 Active site

The active site of the *Ct*SOD molecule can be seen as a buried pocket close to the lateral bottom of the structure. Two metal ions at the active site are involved in the

catalytic reaction: Cu^{2+}ion and Zn^{2+} ion. However, no copper was found in four molecules (A, B, G, and H), whereas it was present in the rest of the molecules (C, D, E, and F) of the *Ct*SOD structure. Unlike copper, zinc ions were present in all molecules of the *Ct*SOD structure.

The zinc-binding site lies on the outer and upper sides of the active-site pocket. The $Zn^{2+}$was found to have an occupancy of 1.00, and a well-defined electron density of 8σ contour level in the *Fo*-*Fc* electron density map in all subunits of the *Ct*SOD structure. An anomalous signal (at 7σ contour level) was detected during diffraction data collection near the absorption edge of Zn^{2+} ion, thus confirming its identity. The $Zn^{2+}$was found to be surrounded and coordinated by three histidines (His64, His72, and His81) and one aspartic acid (Asp84) residue (Figure 12).

The copper-binding site lies downward from the zinc-binding site and positioned buried deep at the active site. Copper ions were found in subunits C, D, E, and F with occupancies of 0.44, 0.39, 0.32, and 0.33, respectively. In the other four subunits (A, B, G, and H), where Cu^{2+} ions were absent, the site was occupied by water molecules. The Cu^{2+} ions were surrounded and coordinated by four histidines: His64, His47, His49, and His121 (Figure 12). Interestingly, His64 was commonly present in both the zinc-binding site and the copper-binding site because of the unique central position and orientation between the two binding sites. Therefore, it ligates with both the zinc and copper ions with a distance of 2.0 Å and 2.6 Å, respectively. On the lower side of the active site, an electrostatic loop was identified (residues 121–144) that connects two β-strands (7g and 8h). This electrostatic loop also holds the α2 helix (135–138) and has been implicated in Cu^{2+} uptake [132].

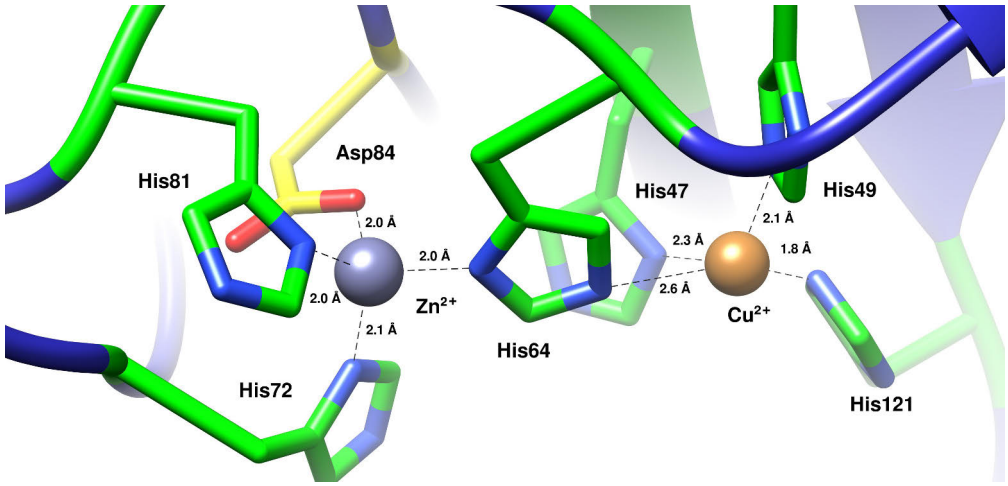


Figure 12. Active site of *Ct*SOD. Copper- and zinc-binding sites are depicted along with their contributing residues. The figure was created using UCSF Chimera [97].

This loop consists of several charged residues, including Asp125, Asp126, Glu133, Glu134, Lys137, and Arg144. The loop was found to be in the same conformation in all subunits of the *Ct*SOD structure, suggesting that its position does not affect copper binding.

4.3.2 Structural comparison

The structure of *Ct*SOD was compared with that of other SODs for comparative analysis. For this purpose, structure-based sequence alignment was performed using a secondary structure matching (SSM) tool [133]. The result identified a Cu,Zn-SOD from yeast (1SDY) with 69% sequence identity and 0.58 Å RMSD in Cα atoms as the closest structure to the *Ct*SOD. It was followed by Cu,Zn-SODs from *Xenopus laevis* (PDB ID: 1XSO) with 58% sequence identity and 0.65 Å RMSD in Cα atoms, *Mus musculus* (PDB ID: 3GTT) with 57.5% sequence identity and 0.80 Å RMSD in Cα atoms, *Sedum alfredii* (PDB ID: 4RVP) with 55.4% sequence identity and 0.80 Å RMSD in Cα atoms, and from *H. sapiens* (PDB ID: 2ZKY) with 53.5% sequence identity and 0.69 Å RMSD in Cα atoms.

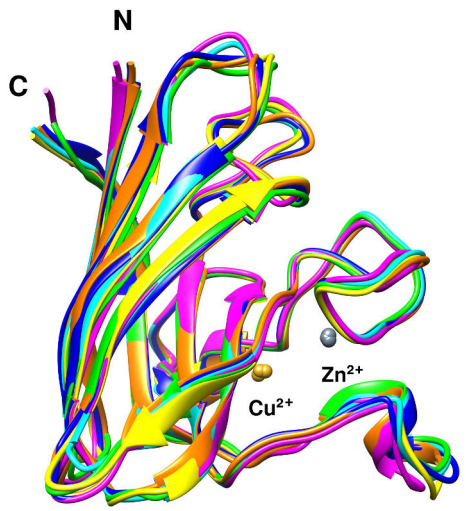


Figure 13. Structural comparison of *Ct*SOD with other Cu,Zn-SODs using SSM for structural superposition. Cu,Zn-SODs are colored as follows: *Ct*SOD (green), 1SDY (yellow), 1XSO (cyan), 2ZKY (blue), 3GTT (orange), and 4RVP (magenta). The figure was created using UCSF Chimera [97].

The structural comparison revealed that the Cu,Zn-SOD structures are widely conserved. Their active site pocket region, along with the copper- and zinc-binding sites, are conserved (Figure 13). The active site residues, including histidine residues (His47, His49, His64, His72, His81, and His121), an aspartic acid residue (Asp84),

and Cys residues forming the disulfide bond, were also found to be conserved. However, some differences were observed in their electrostatic loop. In the consensus sequence alignment, Asp and Leu residues were observed at positions 131 and 134 except in 1SDY and 4RVP, where Gly and Glu residues were observed at the given positions, respectively. Similarly, variations were observed at residues 122, 123, 124, 132, 133, 137, and 143 in the electrostatic loop, as shown in the following Figure 14.

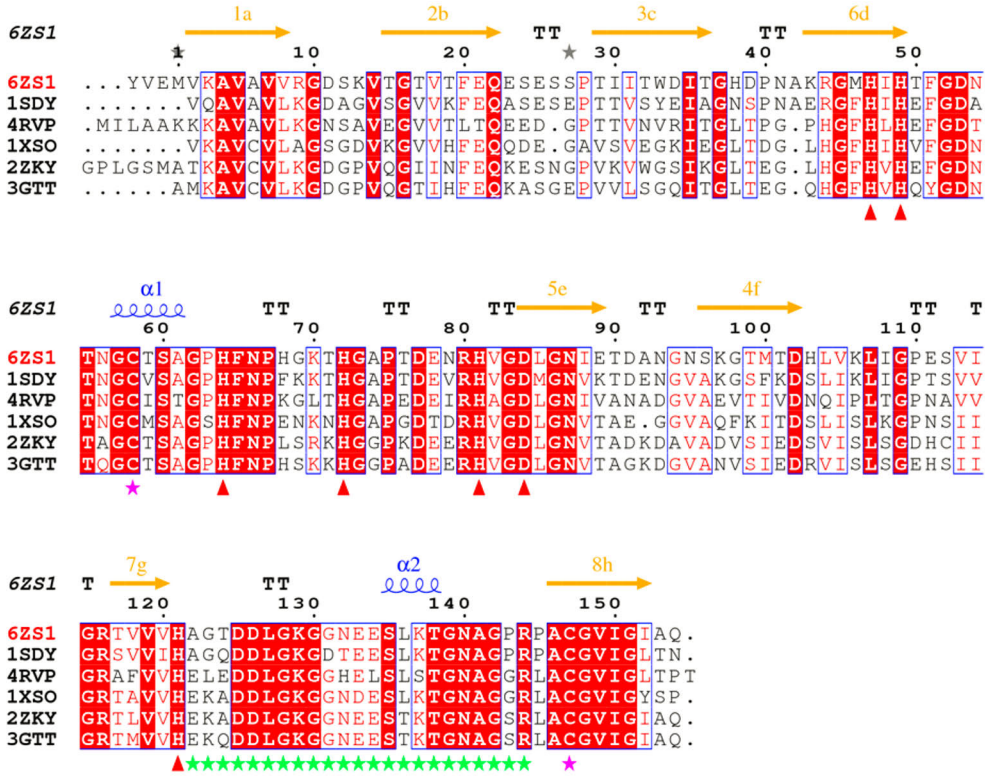


Figure 14. Structure-based sequence alignment of *Ct*SOD (PDB ID: 6ZS1). Highly conserved residues are shown in red background columns with white letters. Alpha helices are shown in blue colour while beta strands are shown in orange colour above the sequence alignment. Cys residues forming a disulphide bond are marked with 'magenta' colored stars underneath the consensus column. Active site residues including six His and one Asp residues are highlighted with 'red triangles' at the bottom of the respective columns. The electrostatic loop is marked with green-colored stars below the alignment. The figure was created with ESPript 3.0 [134].

4.3.3 Thermostability contributing factors

*Ct*SOD was found to be thermostable at 50 °C and 60 °C, retaining 93.6% of its activity after 60 minutes incubation at 60 °C. The possibility of *Ct*SOD to form oligomers, may contribute to its thermostability. Oligomerization has been suggested

to contribute to thermostability, for example, in L-isoaspartyl-O-methyltransferase from *Sulfolobus tokodaii* and *Pyrococcus furiosus* ornithine carbamoyltransferase [135,136]. A hyperthermostable Fe-SOD(*tc*SOD) was reported to form high order oligomers [137]. Interestingly, to further increase the thermostability of other Cu,Zn-SODs, an N-terminal domain found in a thermophilic SOD from *Geobacillus thermodenitrificans* NG80-2 has been used as a fusion partner [138].

Importantly, the Thr:Ser ratio was observed to be significantly high in *Ct*SOD with a similar trend reported previously in family 11 xylanases from *C. thermophilum* [139]. Table 8 shows representative thermostability parameters among different Cu,Zn-SODs [140–146].

Table 8. Comparison of different parameters in Cu,Zn-SODs to evaluate their thermostability. PDBePISA [100], ESBRI [147] and Expasy-Protparam [101] tools were used in fetching the data and formatting this table.

Parameters	**6ZS1**	**1SDY**	**1XSO**	**2ZKY**	**3F7L**	**3GTT**	**4RVP**
Organism	*C. thermophilum*	*S. cerevisiae*	*Xenopus laevis*	*H.sapiens*	*Alvinella pompejana*	*Mus musculus*	*Sedum Alfredia*
Asp + Glu (-)	19	19	19	21	22	18	18
Arg + Lys (+)	13	14	12	15	14	13	10
Val%	8.9	11.1	10	8.8	10.1	9.2	10.1
A.A residues	157	153	150	159	152	153	158
Intra chain salt bridges (chain A)	4	1	2	7	3	5	0
SAS ($Å^2$)	7554	7603	7254	7607	7420	7680	7290
Interface area ($Å^2$)	728.7	633.8	666.8	699.3	703.7	721.1	680.3
No. of H-bonds at interface	4	7	7	4	8	6	8
Pro:Gly ratio	0.35	0.36	0.15	0.23	0.18	0.19	0.37
Thr:Ser ratio	2.13	0.90	0.78	0.73	3.0	0.9	3.75

4.4 Crystal structure of γ-*Ca*CA

The structure of γ-*Ca*CA was refined to 1.11 Å resolution, the highest recorded resolution to date for any γ-type carbonic anhydrase. γ-*Ca*CA was crystallized with three molecules (A, B, C) in the asymmetric unit (Figure 15). The molecules exhibit minor differences in RMSD values (RMSD for A-B, B-C and C-A pairs were 0.134 Å, 0.128 Å, and 0.128 Å, respectively) owing to the flexibility at the C-terminus having 6xHis-tags. Each molecule has two distinctive parts. The first part starting from the N-terminus (1–143 residues long), has a triangular cylindrical shape with

β-strands. The second part has a long α-helix (144–173 residues long) at the C-terminus (Figure 15). Each molecule lies parallel to the next neighboring molecule and joins it via active site coordination containing zinc as the metal ion. Overall, the trimer structure of γ-*Ca*CA has 3 Zn^{2+} ions (one per molecule), 12 Na^{+} cations and 2 formate (FMT) ions. Both FMTs lie at the interface of chain A and chain C. FMT1 was found s in proximity with the zinc ion while FMT2 was observed located beneath but away from the FMT1, near to the start of the α-helix.

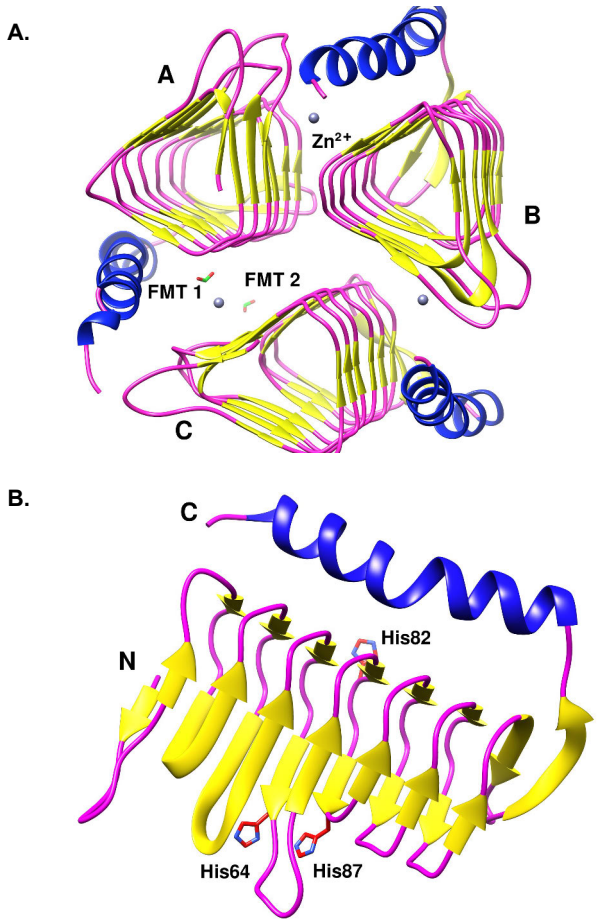


Figure 15. (A) Overall structure of γ-*Ca*CA. **(B)** Chain B of γ-*Ca*CA. The three catalytic His residues are shown.

4.4.1 Active site

The location of the active site was positioned in the middle of the first part (as discussed in the previous paragraph) of each molecule lying outside towards the neighboring molecule. The catalytic active site of γ-*Ca*CA was formed with a Zn^{2+} ion,

three highly conserved catalytic residues i.e., His64, His82, His87, and proposedly a fourth residue (Tyr159). Two histidines (His64 and His82) come from one molecule while the third histidine (His87) and Tyr159 come from the neighboring molecule around the central Zn^{2+} ion. Hence, three active sites are present, each at the interface of neighboring chains. The interaction distances at the active site amongst the participating residues are measured and depicted in Figure 16.

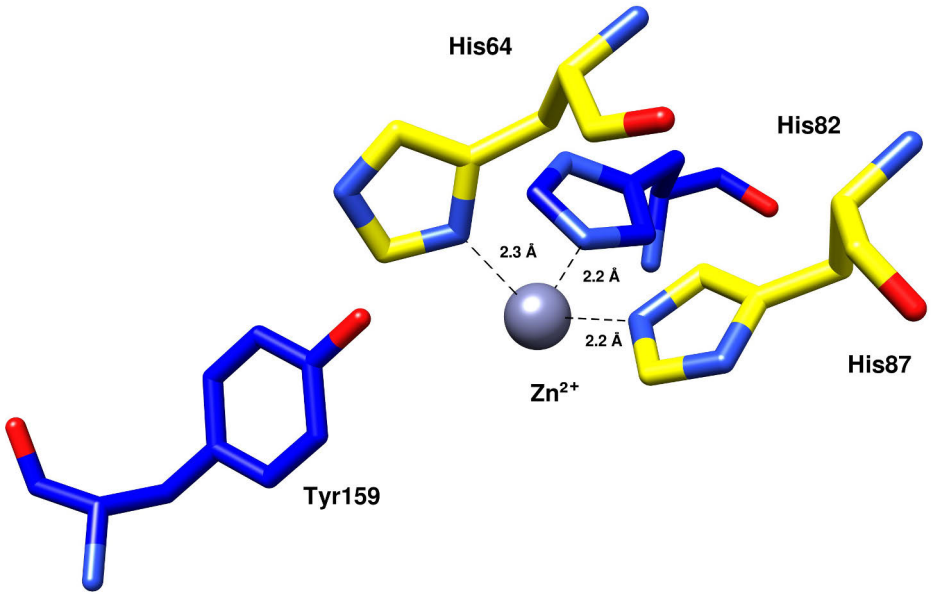


Figure 16. Zn^{2+} interactions at the active site of γ-*Ca*CA. His64 and His87 from chain B are colored in yellow while His82 and Tyr159 from chain C, are colored in blue. The figure was created using UCSF Chimera [97].

4.4.2 Structural comparison

γ-*Ca*CA structure was compared with other γ-type CA structures to evaluate structural differences. For this purpose, PDBeFold [98] and PDBePISA [100] were employed for searching structures with sequence similarities to the γ-*Ca*CA. The γ-type CA structures selected for comparison were, CA from *Thermus thermophilus* HB8 (TtCA, PDB ID: 6IVE), CA from *E. coli* (YrdA; PDB ID: 3TIO) and CA from *Methanosarcina thermophila* (Cam; PDB ID: 1THJ), CA from *Burkholderia pseudomallei* (BpsCA; PDB ID: 7ZW9), CA from *Brucella abortus* (RicA; PDB ID: 4N27), and CA from *Geobacillus Kaustophilus* (Cag; PDB ID: 3VNP). The sequence similarities of γ-*Ca*CA with other CAs were observed as with Cag (51.79%), YrdA (43.45%), RicA (41.67%), BpsCA (39.29%), TtCA (37.5%), and with Cam (22.62%). Key findings noticed upon comparison are: i). All the compared

CA structures have zinc (Zn^{2+}) as the catalytic metal present at the active site. ii). The active site has three highly conserved His catalytic residues. iii). The root mean square deviation (RMSD) among the compared CA structures was between 0.71–1.45 Å, suggesting differences in the structures as illustrated in Figure 17. iv) The highest RMSD (1.45 Å) observed with Cam can be attributed to its significant structural differences with γ-*Ca*CA at the N-terminus, β1-β2 loop, and β8-β9 loop. A structure-based sequence alignment is shown in Figure 18.

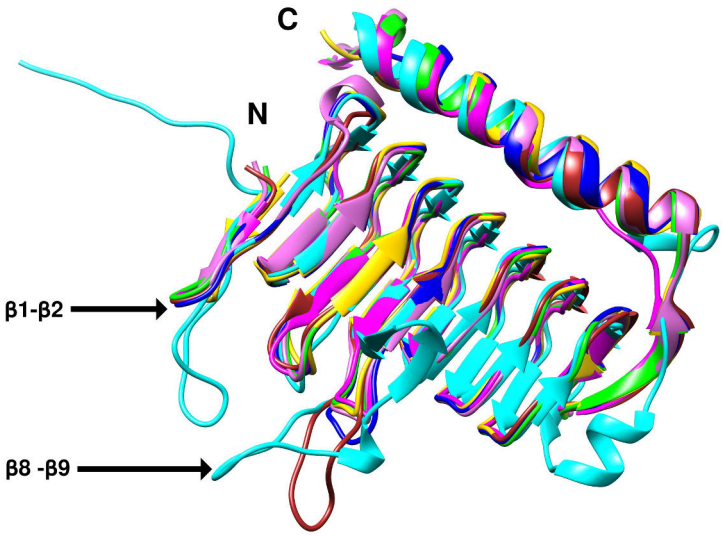


Figure 17. Structural comparison among the γ-type CAs. γ-*Ca*CA is colored blue and compared with TtCA (magenta), YrdA (brown), BpsCA (green), RicA (orchid), Cag (gold), and Cam (cyan). The figure was created using UCSF Chimera [97].

4.4.3 Thermostability contributing factors

Carbonic anhydrases extracted from thermophilic organisms exhibit high thermostability for example *Methanobacterium thermoautotrophicum* was found stable up to 75 °C [148]. The thermal stability of the γ-*Ca*CA was assessed to evaluate the effect of different pH values on structural stability by employing differential scanning fluorimetry (DSF) to monitor melting temperature (Tm). γ-*Ca*CA exhibited Tm values higher than 60°C in the pH range between 5.5 and 9, suggesting it a thermostable enzyme. Several reasons for its thermostability were proposed and evaluated, However, no concrete pattern or formula was discovered. γ-*Ca*CA structural details lead to some clues that could be considered to explain its thermostability.

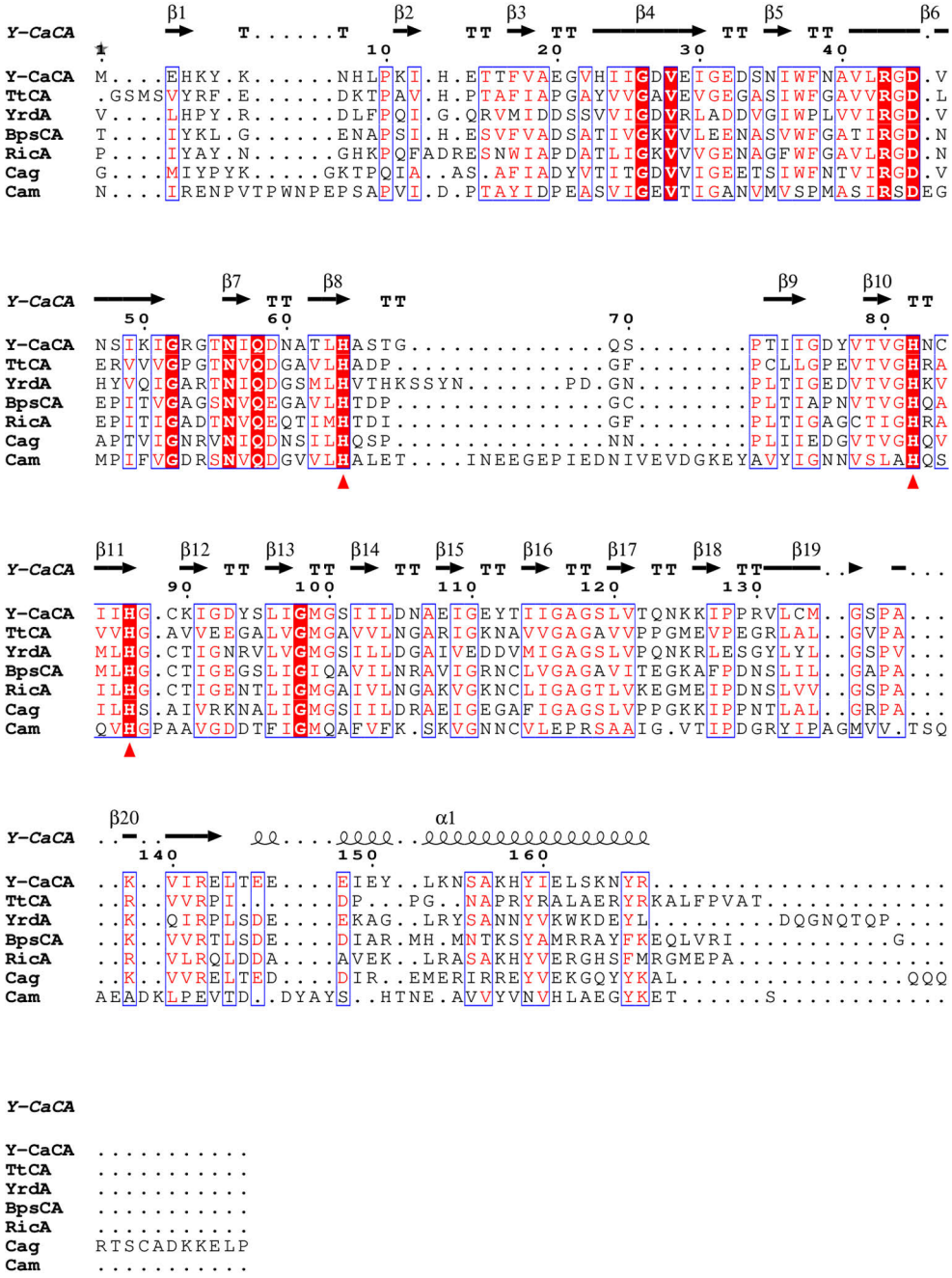


Figure 18. Structure-based sequence alignment of the compared CA structures. Conserved residues are highlighted in red columns. Conserved catalytic residues (His) at the active site are marked by red triangles underneath.

Hydrophobic clusters (HCs) were identified using ProteinTools [149]. In general, Ile, Leu, and Val residues contribute to form hydrophobic clusters. Hydrophobic clusters conform to protein as water resistant and result in higher protein stability. In total, 11 hydrophobic clusters were observed in γ-*Ca*CA among which 3 were the largest clusters (cluster no. 0, 5, and 9) in chains A, B, and C respectively (Figure 19). These large clusters are located inside the trigonal beta-strands pyramid of each chain and occupy an area of 5330.0 Å for cluster no. 0, 5216.1 Å for cluster no. 5 and 5236.4 Å for cluster no. 9.

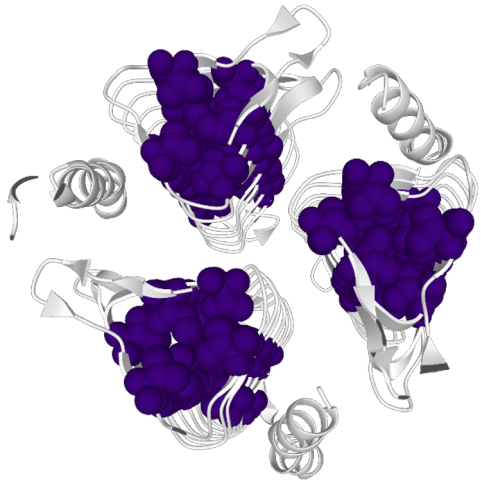


Figure 19. The largest hydrophobic cluster in each chain of γ-*Ca*CA are colored in dark blue. The figure was obtained from ProteinTools [149].

The differences among the charged residues, hydrophobic residues, salt bridges, hydrogen bonds, isoelectric point, and predicted melting temperature have been recorded in Table 9. The numeric values of these factors, when compared, suggest γ-*Ca*CA as a potentially thermostable *γ*-CA.

Table 9. Comparison of the thermostability contributing factors among γ-CAs.

	γ-*Ca*CA (9QEV)	**TtCA (6IVE)**	**YrdA (3TIO)**	**RicA (4N27)**	**BpsCA (7ZW9)**	**Cag (3VNP)**	**Cam (1THJ)**
Mol. wt (kDa)	19.2	17.5	20.0	20.8	22.4	20.0	23.0
Number of residues	173	171	183	195	210	183	214
Asp + Glu (-)	20	15	23	18	22	21	34
Arg + Lys (+)	16	15	16	16	17	21	10
Pro: Gly	0.29	0.64	0.45	0.35	0.33	0.65	0.94
Val + Leu + Ile	25.8	26.2	27.9	21.6	21.9	26.2	23.4
Surface area (SAS) Å^2	8627	7996	8149	8522	8111	8065	10527
Interface area Å^2	863	1214	859	953	949	1035	1348
Intra-chain salt bridges	16	42	55	28	3	17	3
H-bonding interactions	15	16	10	14	15	11	22
Predicted Solubility	0.59	0.58	0.69	0.46	0.45	0.58	0.91
Isoelectric point (pI)	6.3	8.1	5.3	7.1	6.5	7.9	4.1

Numbers in this table were fetched from Expasy-ProtParam [101], PDBePISA [100], ESBRI [147], and Protein-Sol [150].

5 Conclusions and Future Perspectives

In conclusion, four thermozymes were investigated in this research work to determine their 3D structure and probable thermostability contributing factors. A summary of findings is given below:

5.1 *Ct*BGL

*Ct*BGL structure was determined at a resolution of 2.99 Å (PDB ID: 6SZ6). The structure exhibited a conserved three-domain structure (TIM, α/β sandwich, and FnIII domain) with variations in loop and linker regions. *Ct*BGL's catalytic residues were identified Asp287 (nucleophilic) in the TIM domain and Glu517 (acid/base) in the α/β sandwich domain. *Ct*BGL structure revealed a new glycosylation site at position Asn504. In comparison with other β-glucosidases, *Ct*BGL was found moderately glycosylated. The glycosylation pattern and configuration of charged residues were attributed as the most probable contributing factors to the *Ct*GBL's thermostability.

5.2 n*Ta*AA9A

n*Ta*AA9A structure was determined at a resolution of 1.36 Å (PDB ID: 7Q1K). It revealed a characteristic LPMOs β-sandwich fold with antiparallel β-sheets, twisted around and connected through loops. n*Ta*AA9A was found active in cleaving cellulose and xylan substrates to C1- and C4-oxidized products. Like in other LPMOs, His1 was found methylated and a catalytic Cu^{2+} ion was identified close to the N-terminal in n*Ta*AA9A. When compared for glycosylation, n*Ta*AA9A was found glycosylated with at least two NAG molecules at Asn138, pointing towards a shallow groove whereas only one NAG molecule was observed in r*Ta*AA9A, pointing outwards and solvent exposed. A higher thermostability of n*Ta*AA9A was also confirmed by TLC, MALDI-TOF MS, and HPLC-RID. Variations in the surface loops and glycosylation patterns were observed as the possible thermostability contributing factors.

5.3 *Ct*SOD

*Ct*SOD structure was determined at a resolution of 1.56 Å (PDB ID: 6ZS1). The structure was crystallized with eight molecules (A–H) in the asymmetric unit. Each molecule has a pair of 4-stranded antiparallel β-sheets, and 2 short α-helices connected through the linker regions. *Ct*SOD active site revealed Cu^{2+} and Zn^{2+} bindings sites. Zn^{2+} was observed in all eight molecules while Cu^{2+} was present only in C, D, E, and F molecule. Hydrophobic and charged residues were suggested to play a role in the oligomerization of *Ct*SOD. *Ct*SOD structure exhibits a disulfide bond between Cys58 and Cys147. The presence of this disulfide bond and electrostatic interactions in *Ct*SOD were assumed as potential contributing factor to its thermostability.

5.4 γ-*Ca*CA

γ-*Ca*CA structure was determined at a resolution of 1.11 Å. γ-*Ca*CA was crystallized as a trimer in the asymmetric unit. Each molecule has a triangular cylindrical region with downwards running β-strands starting from the N-terminus and connected through the linker region to the long α-helix ending at the C-terminus. γ-*Ca*CA active site has Zn^{2+} as the catalytic metal ion coordinating with three highly conserved His residues (His64, His82, and His87). The active site was formed between the neighbouring molecules as two histidines (His64 and His82) contribute from the same molecule while His87 from the neighbouring molecule. Structural comparison of γ-*Ca*CA with other γ-CAs showd subtle differences in the loops and the linker regions. Thermal shift assay revealed Tm values higher than 60 °C in the pH range between 5.5 and 9 for γ-*Ca*CA. Hydrophobic clusters, charged residues, and salt bridges were proposed as possible contributing factors to its thermostability. Based on the γ-*Ca*CA structure,an engineered form of γ-*Ca*CA (γ-*Ca*CAmut) with high thermostability (>100 °C) has been created that is currently under investigation with a preliminary structure available. Both structures will be compared in detail for a better understanding of structure-function-stability interplay in γ-*Ca*CA.

Designing novel and efficient thermozymes require in-depth understanding of their 3D structure. The results obtained in this thesis will be used to further understand protein thermostability for better enzyme manipulation to meet green chemistry and sustainable future requirements.

Acknowledgements

PhD research is a challenging journey that entails uphill tasks and phases filled with moments of achievement, gratitude, confusion, dullness, depression, and hope. Numerous actors and factors influence this journey, such as the research supervisor, funding, lab and resources, collaborative research groups, and family. I have been privileged to find myself surrounded by an ideal research supervisor, an exemplary research work environment, the best lab with the latest equipment, and supportive peers. Although there have been inert periods with family bindings or no research funding, all were pleasantly upheld with my research mentor's support and guidance.

I would like to express my gratitude to my research mentor, counsellor, and beacon, Prof. Tassos Papageorgiou. Throughout my research work, I found him motivating, accommodating, patient, and benevolent. His unwavering support and insightful guidance at every phase of my research work have kept me propelling and advancing. His recurring assistance and probe-welcoming nature set an impeccable example of supervision. I hope to grasp and retain his noble traits to enlighten the research tracks of my upcoming endeavours.

I would like to thank my research director Prof. Jyrki Heino for providing a conducive environment for my research work. His support and direction have been a requisite to achieve my research goals. I would also acknowledge the support and assistance of my faculty coordinators Nina Lehtimäki, Sari Järvi, and Aiste Ivanauskaite. I am thankful to all my mentors to date specially Prof. Furqan Hashmi and Prof. Martti Tolvanen. I am also thankful to Nirmal, Polychronis, and Herbert for their valuable scientific discussions and research inputs. I pay warm regards to my friend Imran Iqbal for always standing and bearing with me.

I would gratefully acknowledge Turku Bioscience Centre and Turun Yliopistosäätiö for the financial support. Turku Bioscience Centre is also acknowledged for offering exceptional infrastructure support. I am thankful to Sirkku Grönroos and Paula Luoma for their administrative support. Mikael Wasberg, Pasi Viljakainen, and Juha Strandén are acknowledged for offering their technical support.

Finally, I would like to exalt my family for their love, support, and devotion.

List of References

1. Sadigov R. Rapid Growth of the World Population and Its Socioeconomic Results. *ScientificWorldJournal*. 2022;2022:8110229. doi:10.1155/2022/8110229
2. Kleespies MW, Dierkes PW. The importance of the Sustainable Development Goals to students of environmental and sustainability studies—a global survey in 41 countries. Humanities and Social Sciences Communications. 2022;9(1):218. doi:10.1057/s41599-022-01242-0
3. Robinson PK. Enzymes: principles and biotechnological applications. Essays Biochem. 2015;59:1-41. doi:10.1042/bse0590001
4. Arsalan A, Younus H. Enzymes and nanoparticles: Modulation of enzymatic activity via nanoparticles. Int J Biol Macromol. 2018;118(Pt B):1833-1847. doi:10.1016/j.ijbiomac.2018.07.030
5. Gurung N, Ray S, Bose S, Rai V. A broader view: microbial enzymes and their relevance in industries, medicine, and beyond. Biomed Res Int. 2013;2013:329121. doi:10.1155/2013/329121
6. Saleh T, Kalodimos CG. Enzymes at work are enzymes in motion. Science. 2017;355(6322):247-248. doi:10.1126/science.aal4632
7. Schwartz SD. Protein Dynamics and Enzymatic Catalysis. J Phys Chem B. 2023;127(12):2649-2660. doi:10.1021/acs.jpcb.3c00477
8. Stein A, Rueda M, Panjkovich A, Orozco M, Aloy P. A Systematic Study of the Energetics Involved in Structural Changes upon Association and Connectivity in Protein Interaction Networks. Structure. 2011;19(6):881-889. doi:10.1016/j.str.2011.03.009
9. Lewis T, Stone WL. Biochemistry, Proteins Enzymes. In: StatPearls. StatPearls Publishing; 2024. Accessed February 23, 2024. http://www.ncbi.nlm.nih.gov/books/NBK554481/
10. Balaji S. The transferred translocases: An old wine in a new bottle. Biotechnol Appl Biochem. 2022;69(4):1587-1610. doi:10.1002/bab.2230
11. Gasteiger E, Hoogland C, Gattiker A, et al. Protein Identification and Analysis Tools on the ExPASy Server. Walker JM, ed. The Proteomics Protocols Handbook. Published online 2005:571-607. https://doi.org/10.1385/1-59259-890-0:571
12. Enzyme nomenclature and classification: the state of the art. The FEBS Journal. 2023;290(9):2214-2231. https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/febs.16274
13. Godbey WT. Chapter 2 - Proteins. In: Godbey WT, ed. An Introduction to Biotechnology. Woodhead Publishing; 2014:9-33. doi:https://doi.org/10.1016/B978-1-907568-28-2.00002-2
14. Pace CN, Fu H, Fryar KL, et al. Contribution of Hydrophobic Interactions to Protein Stability. J Mol Biol. 2011;408(3):514-528. doi:10.1016/j.jmb.2011.02.053
15. Gurung N, Ray S, Bose S, Rai V. A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond. BioMed Research International. 2013;2013(1):329121. doi:10.1155/2013/329121
16. Nigam PS. Microbial Enzymes with Special Characteristics for Biotechnological Applications. Biomolecules. 2013;3(3):597-611. doi:10.3390/biom3030597
17. El-Gendi H, Saleh AK, Badierah R, Redwan EM, El-Maradny YA, El-Fakharany EM. A Comprehensive Insight into Fungal Enzymes: Structure, Classification, and Their Role in Mankind's Challenges. J Fungi (Basel). 2021;8(1):23. doi:10.3390/jof8010023

18. England JL, Shakhnovich BE, Shakhnovich EI. Natural selection of more designable folds: A mechanism for thermophilic adaptation. Proceedings of the National Academy of Sciences. 2003;100(15):8727-8731. doi:10.1073/pnas.1530713100
19. de Oliveira TB, Gomes E, Rodrigues A. Thermophilic fungi in the new age of fungal taxonomy. Extremophiles. 2015;19(1):31-37. doi:10.1007/s00792-014-0707-0
20. Pietikäinen J, Pettersson M, Bååth E. Comparison of temperature effects on soil respiration and bacterial and fungal growth rates. FEMS Microbiol Ecol. 2005;52(1):49-58. doi:10.1016/j.femsec.2004.10.002
21. Vachher M, Sen A, Kapila R, Nigam A. Microbial therapeutic enzymes: A promising area of biopharmaceuticals. Current Research in Biotechnology. 2021;3:195-208. doi:10.1016/j.crbiot.2021.05.006
22. Ibrar M, Ullah MW, Manan S, Farooq U, Rafiq M, Hasan F. Fungi from the extremes of life: an untapped treasure for bioactive compounds. Appl Microbiol Biotechnol. 2020;104(7):2777-2801. doi:10.1007/s00253-020-10399-0
23. Coker JA. Recent advances in understanding extremophiles. F1000Res. 2019;8:F1000 Faculty Rev-1917. doi:10.12688/f1000research.20765.1
24. Beeler E, Singh OV. Extremophiles as sources of inorganic bio-nanoparticles. *World J Microbiol Biotechnol*. 2016;32(9):156. doi:10.1007/s11274-016-2111-7
25. Li WF, Zhou XX, Lu P. Structural features of thermozymes. *Biotechnol Adv*. 2005;23(4):271-281. doi:10.1016/j.biotechadv.2005.01.002
26. Sadeghi M, Naderi-Manesh H, Zarrabi M, Ranjbar B. Effective factors in thermostability of thermophilic proteins. *Biophys Chem*. 2006;119(3):256-270. doi:10.1016/j.bpc.2005.09.018
27. Vieille C, Zeikus GJ. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev*. 2001;65(1):1-43. doi:10.1128/MMBR.65.1.1-43.2001
28. Hait S, Mallik S, Basu S, Kundu S. Finding the generalized molecular principles of protein thermal stability. *Proteins*. 2020;88(6):788-808. doi:10.1002/prot.25866
29. De Vendittis E, Castellano I, Cotugno R, Ruocco MR, Raimo G, Masullo M. Adaptation of model proteins from cold to hot environments involves continuous and small adjustments of average parameters related to amino acid composition. *J Theor Biol*. 2008;250(1):156-171. doi:10.1016/j.jtbi.2007.09.006
30. Trivedi S, Gehlot HS, Rao SR. Protein thermostability in Archaea and Eubacteria. *Genet Mol Res*. 2006;5(4):816-827.
31. Panja AS, Bandopadhyay B, Maiti S. Protein Thermostability Is Owing to Their Preferences to Non-Polar Smaller Volume Amino Acids, Variations in Residual Physico-Chemical Properties and More Salt-Bridges. *PLOS ONE*. 2015;10(7):e0131495. doi:10.1371/journal.pone.0131495
32. Suhre K, Claverie JM. Genomic correlates of hyperthermostability, an update. *J Biol Chem*. 2003;278(19):17198-17202. doi:10.1074/jbc.M301327200
33. Szilágyi A, Závodszky P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure*. 2000;8(5):493-504. doi:10.1016/s0969-2126(00)00133-7
34. Taylor TJ, Vaisman II. Discrimination of thermophilic and mesophilic proteins. *BMC Struct Biol*. 2010;10 Suppl 1(Suppl 1):S5. doi:10.1186/1472-6807-10-S1-S5
35. Paz A, Mester D, Baca I, Nevo E, Korol A. Adaptive role of increased frequency of polypurine tracts in mRNA sequences of thermophilic prokaryotes. *Proc Natl Acad Sci U S A*. 2004;101(9):2951-2956. doi:10.1073/pnas.0308594100
36. Parthasarathy S, Murthy MR. Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Eng*. 2000;13(1):9-13. doi:10.1093/protein/13.1.9
37. Kumar S, Dangi AK, Shukla P, Baishya D, Khare SK. Thermozymes: Adaptive strategies and tools for their biotechnological applications. *Bioresour Technol*. 2019;278:372-382. doi:10.1016/j.biortech.2019.01.088

38. Cobb RE, Chao R, Zhao H. Directed Evolution: Past, Present and Future. *AIChE J*. 2013;59(5):1432-1440. doi:10.1002/aic.13995
39. Khersonsky O, Lipsh R, Avizemer Z, et al. Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol Cell*. 2018;72(1):178-186.e5. doi:10.1016/j.molcel.2018.08.033
40. Shroff R, Cole AW, Morrow BR, et al. A structure-based deep learning framework for protein engineering. *BioRxiv*. Published online 2019:833905.
41. Saqib AAN, Siddiqui KS. How to calculate thermostability of enzymes using a simple approach. *Biochem Mol Biol Educ*. 2018;46(4):398-402. doi:10.1002/bmb.21127
42. Qin Z, Balasubramanian SK, Wolkers WF, Pearce JA, Bischof JC. Correlated parameter fit of arrhenius model for thermal denaturation of proteins and cells. *Ann Biomed Eng*. 2014;42(12):2392-2404. doi:10.1007/s10439-014-1100-y
43. Elgert C, Rühle A, Sandner P, Behrends S. Thermal shift assay: Strengths and weaknesses of the method to investigate the ligand-induced thermostabilization of soluble guanylyl cyclase. *J Pharm Biomed Anal*. 2020;181:113065. doi:10.1016/j.jpba.2019.113065
44. Miles AJ, Janes RW, Wallace BA. Tools and methods for circular dichroism spectroscopy of proteins: a tutorial review. *Chem Soc Rev*. 2021;50(15):8400-8413. doi:10.1039/d0cs00558d
45. Jia L, Yarlagadda R, Reed CC. Structure Based Thermostability Prediction Models for Protein Single Point Mutations with Machine Learning Tools. *PLoS One*. 2015;10(9):e0138022. doi:10.1371/journal.pone.0138022
46. Sterpone F, Melchionna S. Thermophilic proteins: insight and perspective from in silico experiments. *Chem Soc Rev*. 2012;41(5):1665-1676. doi:10.1039/c1cs15199a
47. Shukla E, Bendre AD, Gaikwad SM, Shukla E, Bendre AD, Gaikwad SM. Hydrolases: The Most Diverse Class of Enzymes. In: *Hydrolases*. IntechOpen; 2022. doi:10.5772/intechopen.102350
48. Espina G, Atalah J, Blamey JM. Extremophilic Oxidoreductases for the Industry: Five Successful Examples With Promising Projections. *Front Bioeng Biotechnol*. 2021;9:710035. doi:10.3389/fbioe.2021.710035
49. Wilson DB. Cellulases and biofuels. *Curr Opin Biotechnol*. 2009;20(3):295-299. doi:10.1016/j.copbio.2009.05.007
50. Phitsuwan P, Laohakunjit N, Kerdchoechuen O, Kyu KL, Ratanakhanokchai K. Present and potential applications of cellulases in agriculture, biotechnology, and bioenergy. *Folia Microbiol (Praha)*. 2013;58(2):163-176. doi:10.1007/s12223-012-0184-8
51. Bai H, Wang H, Sun J, et al. Production, purification and characterization of novel beta glucosidase from newly isolated Penicillium simplicissimum H-11 in submerged fermentation. *EXCLI J*. 2013;12:528-540.
52. Yang W, Su Y, Wang R, et al. Microbial production and applications of β-glucosidase-A review. *Int J Biol Macromol*. 2024;256(Pt 2):127915. doi:10.1016/j.ijbiomac.2023.127915
53. Li Y, Liu N, Yang H, et al. Cloning and characterization of a new β-glucosidase from a metagenomic library of rumen of cattle feeding with Miscanthus sinensis. *BMC Biotechnol*. 2014;14:85. doi:10.1186/1472-6750-14-85
54. Singh G, Verma AK, Kumar V. Catalytic properties, functional attributes and industrial applications of β-glucosidases. *3 Biotech*. 2015;6(1):3. doi:10.1007/s13205-015-0328-z
55. Chae Y, Cui R, Woong Kim S, An G, Jeong SW, An YJ. Exoenzyme activity in contaminated soils before and after soil washing: ß-glucosidase activity as a biological indicator of soil health. *Ecotoxicol Environ Saf*. 2017;135:368-374. doi:10.1016/j.ecoenv.2016.10.007
56. McDonald AG, Tipton KF. Enzyme nomenclature and classification: the state of the art. *FEBS J*. 2023;290(9):2214-2231. doi:10.1111/febs.16274
57. Shahi ZKM, Takalloo Z, Mohamadzadeh J, Sajedi RH, Haghbeen K, Aminzadeh S. Thermophilic iron containing type superoxide dismutase from Cohnella sp. A01. *Int J Biol Macromol*. 2021;187:373-385. doi:10.1016/j.ijbiomac.2021.07.150

58. Zheng M, Liu Y, Zhang G, Yang Z, Xu W, Chen Q. The Applications and Mechanisms of Superoxide Dismutase in Medicine, Food, and Cosmetics. *Antioxidants (Basel)*. 2023;12(9):1675. doi:10.3390/antiox12091675
59. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research*. 2021;50(D1):D571-D577.
60. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*. 2013;6(1):41. doi:10.1186/1754-6834-6-41
61. Sharma S, Modi R, Kaur A. Lytic polysaccharide monooxygenases producing microbes: A key indicator for biomass-degrading enzymes. *Biocatalysis and Agricultural Biotechnology*. 2024;60:103337. doi:10.1016/j.bcab.2024.103337
62. Várnai A, Hegnar OA, Horn SJ, Eijsink VGH, Berrin JG. Fungal Lytic Polysaccharide Monooxygenases (LPMOs): Biological Importance and Applications. In: Zaragoza Ó, Casadevall A, eds. *Encyclopedia of Mycology*. Elsevier; 2021:281-294. doi:10.1016/B978-0-12-819990-9.00019-6
63. Forsberg Z, Sørlie M, Petrović D, et al. Polysaccharide degradation by lytic polysaccharide monooxygenases. *Curr Opin Struct Biol*. 2019;59:54-64. doi:10.1016/j.sbi.2019.02.015
64. Corrêa TLR, dos Santos LV, Pereira GAG. AA9 and AA10: from enigmatic to essential enzymes. *Appl Microbiol Biotechnol*. 2016;100(1):9-16. doi:10.1007/s00253-015-7040-0
65. Jung S, Song Y, Kim HM, Bae HJ. Enhanced lignocellulosic biomass hydrolysis by oxidative lytic polysaccharide monooxygenases (LPMOs) GH61 from Gloeophyllum trabeum. *Enzyme Microb Technol*. 2015;77:38-45. doi:10.1016/j.enzmictec.2015.05.006
66. Zhang R. Functional characterization of cellulose-degrading AA9 lytic polysaccharide monooxygenases and their potential exploitation. *Appl Microbiol Biotechnol*. 2020;104(8):3229-3243. doi:10.1007/s00253-020-10467-5
67. Harris PV, Welner D, McFarland KC, et al. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry*. 2010;49(15):3305-3316. doi:10.1021/bi100009p
68. McDonald AG, Boyce S, Tipton KF. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res*. 2009;37(Database issue):D593-597. doi:10.1093/nar/gkn582
69. Puchart V, Šuchová K, Biely P. Chapter 6 - Importance of accessory enzymes in hemicellulose degradation. In: Goldbeck R, Poletto P, eds. *Polysaccharide-Degrading Biocatalysts*. Foundations and Frontiers in Enzymology. Academic Press; 2023:139-176. doi:https://doi.org/10.1016/B978-0-323-99986-1.00005-3
70. Lindskog S. Structure and mechanism of carbonic anhydrase. *Pharmacology & Therapeutics*. 1997;74(1):1-20. doi:10.1016/S0163-7258(96)00198-2
71. Bertini I. *Biological Inorganic Chemistry: Structure and Reactivity*. University Science Books; 2007.
72. Di Fiore A, Alterio V, Monti SM, De Simone G, D'Ambrosio K. Thermostable Carbonic Anhydrases in Biotechnological Applications. *Int J Mol Sci*. 2015;16(7):15456-15480. doi:10.3390/ijms160715456
73. Sharma A, Chiang RA, Manginell M, et al. Carbonic Anhydrase Robustness for Use in Nanoscale CO2 Capture Technologies. *ACS Omega*. 2023;8(41):37830-37841. doi:10.1021/acsomega.3c02630
74. Manyumwa CV, Bishop ÖT. In Silico Investigation of Potential Applications of Gamma Carbonic Anhydrases as Catalysts of CO2 Biomineralization Processes: A Visit to the Thermophilic Bacteria Persephonella hydrogeniphila, Persephonella marina, Thermosulfidibacter takaii, and Thermus thermophilus. *Int J Mol Sci*. 2021;22(6):2861. doi:10.3390/ijms22062861

75. Somalinga V, Foss E, Grunden AM. Biochemical characterization of a psychrophilic and halotolerant α-carbonic anhydrase from a deep-sea bacterium, Photobacterium profundum. *AIMS Microbiol*. 2023;9(3):540-553. doi:10.3934/microbiol.2023028
76. Fu Y, Jiang YB, Dunphy D, et al. Ultra-thin enzymatic liquid membrane for CO2 separation and capture. *Nat Commun*. 2018;9(1):990. doi:10.1038/s41467-018-03285-x
77. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970;227(5259):680-685. doi:10.1038/227680a0
78. Xu R, Teng F, Zhang C, Li D. Cloning of a gene encoding β-glucosidase from Chaetomium thermophilum CT2 and its expression in Pichia pastoris. *J Mol Microbiol Biotechnol*. 2011;20(1):16-23. doi:10.1159/000322606
79. Li DC, Lu M, Li YL, Lu J. Purification and characterization of an endocellulase from the thermophilic fungus Chaetomium thermophilum CT2. *Enzyme and Microbial Technology*. 2003;33(7):932-937. doi:10.1016/S0141-0229(03)00245-X
80. Chen C, Chen J, Geng Z, Wang M, Liu N, Li D. Regioselectivity of oxidation by a polysaccharide monooxygenase from Chaetomium thermophilum. *Biotechnol Biofuels*. 2018;11:155. doi:10.1186/s13068-018-1156-2
81. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measurement with the Folin phenol reagent. *J Biol Chem*. 1951;193(1):265-275.
82. Chen J, Guo X, Zhu M, Chen C, Li D. Polysaccharide monooxygenase-catalyzed oxidation of cellulose to glucuronic acid-containing cello-oligosaccharides. *Biotechnology for Biofuels*. 2019;12(1):42. doi:10.1186/s13068-019-1384-0
83. Stewart RR, Bewley JD. Lipid peroxidation associated with accelerated aging of soybean axes. *Plant Physiol*. 1980;65(2):245-248. doi:10.1104/pp.65.2.245
84. Kabsch W. XDS. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 2):125-132. doi:10.1107/S0907444909047337
85. Matthews BW. Solvent content of protein crystals. *Journal of Molecular Biology*. 1968;33(2):491-497. doi:10.1016/0022-2836(68)90205-2
86. Agirre J, Atanasova M, Bagdonas H, et al. The ıt CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallographica Section D*. 2023;79(6):449-461. doi:10.1107/S2059798323003595
87. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. 1994;50(Pt 5):760-763. doi:10.1107/S0907444994003112
88. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007;40(Pt 4):658-674. doi:10.1107/S0021889807021206
89. Adams PD, Afonine PV, Bunkóczi G, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 2):213-221. doi:10.1107/S0907444909052925
90. Bunkóczi G, Read RJ. Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 4):303-312. doi:10.1107/S0907444910051218
91. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004;60(Pt 12 Pt 1):2126-2132. doi:10.1107/S0907444904019158
92. Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*. 1992;355(6359):472-475. doi:10.1038/355472a0
93. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc*. 2008;3(7):1171-1179. doi:10.1038/nprot.2008.91
94. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997;53(Pt 3):240-255. doi:10.1107/S0907444996012255

95. Chen VB, Arendall WB, Headd JJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 1):12-21. doi:10.1107/S0907444909042073
96. Agirre J. Strategies for carbohydrate model building, refinement and validation. *Acta Cryst D*. 2017;73(2):171-186. doi:10.1107/S2059798316016910
97. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612. doi:10.1002/jcc.20084
98. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*. 2004;60(Pt 12 Pt 1):2256-2268. doi:10.1107/S0907444904026460
99. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(Web Server issue):W320-324. doi:10.1093/nar/gku316
100. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372(3):774-797. doi:10.1016/j.jmb.2007.05.022
101. Gasteiger E, Hoogland C, Gattiker A, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Humana Press; 2005:571-607. doi:10.1385/1-59259-890-0:571
102. Basse MJ, Betzi S, Morelli X, Roche P. 2P2Idb v2: update of a structural database dedicated to orthosteric modulation of protein-protein interactions. *Database (Oxford)*. 2016;2016:baw007. doi:10.1093/database/baw007
103. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Research*. 2000;28(1):235-242. doi:10.1093/nar/28.1.235
104. Gudmundsson M, Hansson H, Karkehabadi S, et al. Structural and functional studies of the glycoside hydrolase family 3 β-glucosidase Cel3A from the moderately thermophilic fungus Rasamsonia emersonii. *Acta Crystallogr D Struct Biol*. 2016;72(Pt 7):860-870. doi:10.1107/S2059798316008482
105. Agirre J, Ariza A, Offen WA, et al. Three-dimensional structures of two heavily N-glycosylated Aspergillus sp. family GH3 β-D-glucosidases. *Acta Crystallogr D Struct Biol*. 2016;72(Pt 2):254-265. doi:10.1107/S2059798315024237
106. Yamaguchi H, Nishiyama T, Uchida M. Binding affinity of N-glycans for aromatic amino acid residues: implications for novel interactions between N-glycans and proteins. *J Biochem*. 1999;126(2):261-265. doi:10.1093/oxfordjournals.jbchem.a022443
107. Payne CM, Jiang W, Shirts MR, Himmel ME, Crowley MF, Beckham GT. Glycoside hydrolase processivity is directly related to oligosaccharide binding free energy. *J Am Chem Soc*. 2013;135(50):18831-18839. doi:10.1021/ja407287f
108. Hudson KL, Bartlett GJ, Diehl RC, et al. Carbohydrate-Aromatic Interactions in Proteins. *J Am Chem Soc*. 2015;137(48):15152-15160. doi:10.1021/jacs.5b08424
109. Kim KH, Brown KM, Harris PV, Langston JA, Cherry JR. A proteomics strategy to discover beta-glucosidases from Aspergillus fumigatus with two-dimensional page in-gel activity assay and tandem mass spectrometry. *J Proteome Res*. 2007;6(12):4749-4757. doi:10.1021/pr070355i
110. Suzuki K, Sumitani J ichi, Nam YW, et al. Crystal structures of glycoside hydrolase family 3 β-glucosidase 1 from Aspergillus aculeatus. *Biochem J*. 2013;452(2):211-221. doi:10.1042/BJ20130054
111. Karkehabadi S, Helmich KE, Kaper T, et al. Biochemical characterization and crystal structures of a fungal family 3 β-glucosidase, Cel3A from Hypocrea jecorina. *J Biol Chem*. 2014;289(45):31624-31637. doi:10.1074/jbc.M114.587766
112. Karkehabadi S, Hansson H, Mikkelsen NE, et al. Structural studies of a glycoside hydrolase family 3 β-glucosidase from the model fungus Neurospora crassa. *Acta Crystallogr F Struct Biol Commun*. 2018;74(Pt 12):787-796. doi:10.1107/S2053230X18015662

113. Karkehabadi S, Hansson H, Mikkelsen NE, et al. Structural studies of a glycoside hydrolase family 3 β-glucosidase from the model fungus Neurospora crassa. *Acta Crystallogr F Struct Biol Commun*. 2018;74(Pt 12):787-796. doi:10.1107/S2053230X18015662
114. Lee HS, Qi Y, Im W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Sci Rep*. 2015;5:8926. doi:10.1038/srep08926
115. Turner P, Svensson D, Adlercreutz P, Karlsson EN. A novel variant of Thermotoga neapolitana beta-glucosidase B is an efficient catalyst for the synthesis of alkyl glucosides by transglycosylation. *J Biotechnol*. 2007;130(1):67-74. doi:10.1016/j.jbiotec.2007.02.016
116. Kayser V, Chennamsetty N, Voynov V, Forrer K, Helk B, Trout BL. Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. *Biotechnol J*. 2011;6(1):38-44. doi:10.1002/biot.201000091
117. Ioannou YA, Zeidner KM, Grace ME, Desnick RJ. Human alpha-galactosidase A: glycosylation site 3 is essential for enzyme solubility. *Biochem J*. 1998;332 (Pt 3)(Pt 3):789-797. doi:10.1042/bj3320789
118. Meldgaard M, Svendsen I. Different effects of N-glycosylation on the thermostability of highly homologous bacterial (1,3-1,4)-beta-glucanases secreted from yeast. *Microbiology (Reading)*. 1994;140 (Pt 1):159-166. doi:10.1099/13500872-140-1-159
119. Mohsin I, Poudel N, Li DC, Papageorgiou AC. Crystal Structure of a GH3 β-Glucosidase from the Thermophilic Fungus Chaetomium thermophilum. *Int J Mol Sci*. 2019;20(23):E5962. doi:10.3390/ijms20235962
120. Pozzo T, Romero-García J, Faijes M, Planas A, Nordberg Karlsson E. Rational design of a thermostable glycoside hydrolase from family 3 introduces β-glycosynthase activity. *Glycobiology*. 2017;27(2):165-175. doi:10.1093/glycob/cww081
121. Quinlan RJ, Sweeney MD, Lo Leggio L, et al. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci U S A*. 2011;108(37):15079-15084. doi:10.1073/pnas.1105776108
122. Petrović DM, Bissaro B, Chylenski P, et al. Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation. *Protein Sci*. 2018;27(9):1636-1650. doi:10.1002/pro.3451
123. Yu W, Yu J, Li D. Analysis of lytic polysaccharide monooxygenase activity in thermophilic fungi by high-performance liquid chromatography-refractive index detector. *Front Microbiol*. 2022;13:1063025. doi:10.3389/fmicb.2022.1063025
124. Quinlan RJ, Sweeney MD, Lo Leggio L, et al. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci U S A*. 2011;108(37):15079-15084. doi:10.1073/pnas.1105776108
125. Lo Leggio L, Weihe CD, Poulsen JCN, et al. Structure of a lytic polysaccharide monooxygenase from Aspergillus fumigatus and an engineered thermostable variant. *Carbohydrate Research*. 2018;469:55-59. doi:10.1016/j.carres.2018.08.009
126. Tandrup T, Tryfona T, Frandsen KEH, Johansen KS, Dupree P, Lo Leggio L. Oligosaccharide Binding and Thermostability of Two Related AA9 Lytic Polysaccharide Monooxygenases. *Biochemistry*. 2020;59(36):3347-3358. doi:10.1021/acs.biochem.0c00312
127. Zhang R, Liu Y, Zhang Y, et al. Identification of a thermostable fungal lytic polysaccharide monooxygenase and evaluation of its effect on lignocellulosic degradation. *Appl Microbiol Biotechnol*. 2019;103(14):5739-5750. doi:10.1007/s00253-019-09928-3
128. Mazurkewich S, Seveso A, Hüttner S, Brändén G, Larsbrink J. Structure of a C1/C4-oxidizing AA9 lytic polysaccharide monooxygenase from the thermophilic fungus Malbranchea cinnamomea. *Acta Crystallogr D Struct Biol*. 2021;77(Pt 8):1019-1026. doi:10.1107/S2059798321006628

129. Pucci F, Kwasigroch JM, Rooman M. SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. *Bioinformatics*. 2017;33(21):3415-3422. doi:10.1093/bioinformatics/btx417
130. Thornton JM. Disulphide bridges in globular proteins. *J Mol Biol*. 1981;151(2):261-287. doi:10.1016/0022-2836(81)90515-5
131. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372(3):774-797. doi:10.1016/j.jmb.2007.05.022
132. Ciriolo MR, Battistoni A, Falconi M, Filomeni G, Rotilio G. Role of the electrostatic loop of Cu,Zn superoxide dismutase in the copper uptake process. *Eur J Biochem*. 2001;268(3):737-742. doi:10.1046/j.1432-1327.2001.01928.x
133. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst D*. 2004;60(12):2256-2268. doi:10.1107/S0907444904026460
134. Gouet P, Robert X, Courcelle E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res*. 2003;31(13):3320-3323. doi:10.1093/nar/gkg556
135. Tanaka Y, Tsumoto K, Yasutake Y, et al. How oligomerization contributes to the thermostability of an archaeon protein. Protein L-isoaspartyl-O-methyltransferase from Sulfolobus tokodaii. *J Biol Chem*. 2004;279(31):32957-32967. doi:10.1074/jbc.M404405200
136. Clantin B, Tricot C, Lonhienne T, Stalon V, Villeret V. Probing the role of oligomerization in the high thermal stability of Pyrococcus furiosus ornithine carbamoyltransferase by site-specific mutants. *Eur J Biochem*. 2001;268(14):3937-3942. doi:10.1046/j.1432-1327.2001.02302.x
137. Wang S, Dong ZY, Yan YB. Formation of high-order oligomers by a hyperthemostable Fe-superoxide dismutase (tcSOD). *PLoS One*. 2014;9(10):e109657. doi:10.1371/journal.pone.0109657
138. Li M, Zhu L, Wang W. Improving the thermostability and stress tolerance of an archaeon hyperthermophilic superoxide dismutase by fusion with a unique N-terminal domain. *Springerplus*. 2016;5:241. doi:10.1186/s40064-016-1854-9
139. Hakulinen N, Turunen O, Jänis J, Leisola M, Rouvinen J. Three-dimensional structures of thermophilic beta-1,4-xylanases from Chaetomium thermophilum and Nonomuraea flexuosa. Comparison of twelve xylanases in relation to their thermal stability. *Eur J Biochem*. 2003;270(7):1399-1412. doi:10.1046/j.1432-1033.2003.03496.x
140. Mohsin I, Zhang LQ, Li DC, Papageorgiou AC. Crystal Structure of a Cu,Zn Superoxide Dismutase From the Thermophilic Fungus Chaetomium thermophilum. *Protein Pept Lett*. 2021;28(9):1043-1053. doi:10.2174/0929866528666210316104919
141. Djinović K, Gatti G, Coda A, et al. Structure solution and molecular dynamics refinement of the yeast Cu,Zn enzyme superoxide dismutase. *Acta Crystallogr B*. 1991;47 (Pt 6):918-927. doi:10.1107/s0108768191004949
142. Djinović Carugo K, Battistoni A, Carrì MT, et al. Three-dimensional structure of Xenopus laevis Cu,Zn superoxide dismutase b determined by X-ray crystallography at 1.5 A resolution. *Acta Crystallogr D Biol Crystallogr*. 1996;52(Pt 1):176-188. doi:10.1107/S0907444995007608
143. Seetharaman SV, Taylor AB, Holloway S, Hart PJ. Structures of mouse SOD1 and human/mouse SOD1 chimeras. *Arch Biochem Biophys*. 2010;503(2):183-190. doi:10.1016/j.abb.2010.08.014
144. Shin DS, Didonato M, Barondeau DP, et al. Superoxide dismutase from the eukaryotic thermophile Alvinella pompejana: structures, stability, mechanism, and insights into amyotrophic lateral sclerosis. *J Mol Biol*. 2009;385(5):1534-1555. doi:10.1016/j.jmb.2008.11.031
145. Qiu, R., Li, C., Zhai, J., Tang, L., Zhang, H., Yuan, M., Hu, X. Crystal structure of superoxide dismutase from sedum alfredii. https://doi.org/10.2210/pdb4RVP/pdb
146. Yoshikawa, S., Kukimoto-Niino, M., Ito, K., Shirouzu, M., Urushitani, M., Takahashi, R., Yokoyama, S., RIKEN Structural Genomics/Proteomics Initiative (RSGI). Crystal structure of human Cu-Zn superoxide dismutase mutant G93A. https://doi.org/10.2210/pdb2zky/pdb

147. Costantini S, Colonna G, Facchiano AM. ESBRI: a web server for evaluating salt bridges in proteins. *Bioinformation*. 2008;3(3):137-138. doi:10.6026/97320630003137
148. Smith KS, Ferry JG. A plant-type (beta-class) carbonic anhydrase in the thermophilic methanoarchaeon Methanobacterium thermoautotrophicum. *J Bacteriol*. 1999;181(20):6247-6253. doi:10.1128/JB.181.20.6247-6253.1999
149. Ferruz N, Schmidt S, Höcker B. ProteinTools: a toolkit to analyze protein structures. *Nucleic Acids Res*. 2021;49(W1):W559-W566. doi:10.1093/nar/gkab375
150. Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics*. 2017;33(19):3098-3100. doi:10.1093/bioinformatics/btx345

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2025

ISBN 978-952-02-0183-8 (Print)
ISBN 978-952-02-0184-5 (PDF)
ISSN 0082-7002 (Print)
ISSN 2343-3175 (Online)