Chapter 1

General Introduction and Review of Literature

1.1. Introduction

Lignocellulosic biomass (LCB) from plants is the most abundant biomass on earth. It mainly includes various agricultural residues, such as cereal straw, bagasse, forest residues, waste from paper and pulp industry etc. The annual worldwide production of LCB amounts to approximately 181.5 billion tons, of which 8.2 million tons is currently utilized. Traditionally, LCB is utilized for energy requirement and production of paper, textile and construction materials. With advancement in technology, LCB has been realized as a sustainable source of biofuel, bioplastics, enzymes and other value added products (Mujtaba *et al.* 2023).

LCB is primarily comprised of two carbohydrate components, cellulose (40-50 %), hemicellulose (25-30 %), and one non-carbohydrate phenolic polymer, lignin (15-20 %). Cellulose is the major structural polysaccharide of plant cell wall, and is an unbranched homopolymer of 4-O-β-D-glucopyranosyl-D-glucose units. Unlike cellulose, hemicellulose is composed of several heterogenous polymers like xylan, glucuronoxylan, arabinoxylan and xyloglucan, and its main function is to strengthen the cell wall through binding to the cellulose microfibrils. Lignin is a phenylpropanoid polymer composed of sinapyl (S), coniferyl (G) and *p*-coumaryl (H) alcohol units connected through C-C and ester linkages. It forms an impenetrable network around cellulose which reduces its surface area for hydrolysis by microbial enzymes (Tan *et al.* 2021, Wu *et al.* 2020b). Cellulose and hemicellulose are depolymerized to sugars for further conversion by fermentation and biocatalytic processes to value added products, whereas lignin can be extracted for production of aromatics (Weng *et al.* 2021, Intasian *et al.* 2021). However, LCB remains unutilized or under-utilized due to its high degree of lignification and therefore, its bioconversion is challenging for development of economic biorefineries (Mora-Sandi *et al.* 2021). Currently, the LCB refinery requires energy-intensive pretreatment steps for removal of lignin and subsequent hydrolysis with a cocktail of cellulolytic enzymes to liberate fermentable sugars. In recent years, significant progress has been achieved in removal of recalcitrant lignin through physical, chemical and physicochemical pretreatments, however, these pretreatments are energy intensive and produce microbial and enzymatic inhibitors that obstruct the downstream cellulolytic saccharification and fermentation (Sanchez-Munoz *et al.* 2022).

Microbial depolymerization of LCB is a consequence of the synergistic action of hydrolases, lignin-modifying enzymes (LME) and their associated enzymes. The CAZy database classifies carbohydrate active enzymes (CAZymes) into six categories: glycoside hydrolases (GH) that cleave glycoside bonds, glycosyl transferases (GT) that help in the formation of glycoside bonds, carbohydrate esterases (CE) that modify ester bonds, polysaccharide lyases (PL) that cleave glycoside bonds via non-hydrolytic pathways, auxiliary active (AA) enzymes that are oxidoreductases and assist the other enzymes in LCB degradation. Other than these, there are carbohydrate binding modules (CBM) that are non-enzymatic molecules and facilitate enzyme adherence to substrate (Pinard *et al.* 2015, Sutzl *et al.* 2018). Production of various biocatalysts and their quantities by different microorganisms can be further investigated for industrial benefits such as better saccharification, shorter processing intervals, low enzyme dosage and removal of feedback inhibitors (Malgas *et al.* 2017). Efficient LCB valorization requires fulfillment of the following criteria: microbial hosts that are tolerant to LCB-derived inhibitors, can withstand genetic manipulation, can act on a broad spectrum of substrates, and bioprocess optimization with a focus on maximum benefits from saccharification (Sohn *et al.* 2022). Development of a process for the efficient conversion of LCB, involving less capital and environmental harm, is both lucrative as well as a challenge for biorefineries (Yin *et al.* 2021).

The variety and quantity of enzyme production determines the rate and final efficiency of LCB degradation. Bacterial hydrolytic enzymes are of interest as they could find application in industrial valorization processes (Kumar *et al.* 2016). Thermophiles are being considered for industrial applications due to the numerous advantages they confer, such as increased solubility of cellulosic components, lesser viscosity, increased hydrolysis, lesser contamination due to mesophiles and lower enzyme dosages (Kazeem *et al.* 2017). In addition, the enzymes they produce, can maintain their structure at higher temperatures, resist denaturation in the presence of solvents, have higher reaction velocities and have better enzyme-substrate interactions (Cann *et al.* 2020).

In natural environments, microorganisms employ multiple enzymes with different substrate specificities and catalytic mechanisms for complete degradation of LCB (Kumar and Chandra 2020). To date there are three different paradigms for depolymerization of cellulose by microorganisms; firstly the free enzyme paradigm wherein enzymes are secreted as single catalytic units. Such enzymes are produced and secreted by the aerobic bacteria and fungi (Benatti and Polizeli 2023). Examples include aerobic filamentous fungi, like *Trichoderma reesei* and *Penicillium* species, and aerobic bacteria belonging to genera *Alteromonas, Bacillus, Cellulomonas, Thermobifida* and *Saccharophagus* (Balla *et al.* 2022, Hobdey *et al.* 2015). The other paradigm involves the use of cellulosomes wherein a large complex of several enzymes interacting through cohesin and dockerin modules, operate in close proximity on a large protein scaffold. Anaerobic fungi belonging to phylum Neocallimastigomycota and bacteria, like *Clostridium* sp, *Ruminicoccus* sp, *Acetivibrio cellulolyticus* have been reported to develop cellulosomes (Lillington *et al.* 2021, Gilbert 2007). The recently emerged third paradigm consisting of multimodular enzymes having multiple catalytic domain in one enzyme has been reported in the extreme thermophilic fermentative anaerobes, like *Caldicellulosiruptor morganii* and *Caldicellulosiruptor danielii* (Conway *et al.* 2018). However, due to the structural complexity of LCB the cellulolytic system of individual microorganism is usually inefficient and moreover, such system might be susceptible to metabolic repression.

LCB depolymerization in natural ligninocellulolytic habitat, like compost, bovine rumen, termite gut, forest soil, is achieved through the co-existence of diverse types of microorganisms, that have specialized themselves in lignocellulosic breakdown (Mello *et al.* 2016). Such microbial consortia utilize interdependent interactions between multiple compatible, often taxonomically diverse microorganisms to achieve collective adaptibility against fluctuating environmental conditions such as pH, moisture, temperature and are therefore preferable over single strain cultures (Bradacova *et al.* 2019). Their metagenomic analyses have revealed the microbial diversity, and ligninocellulolytic genes for comprehensive understanding of the depolymerization pathway (Moraes *et al.* 2018). Microbial consortia developed from natural LCB degradation environment by “top down” strategy using enrichment culture with ligninocellulolytic substrates as carbon source has improved LCB deconstruction potential, however, such consortia are very complex and show variation in microbial diversity and function. The problem can be addressed by constructing synthetic consortia with limited number of culturable microorganisms of wherein metabolic co-operation between the members bring about effective deconstruction. Moreover, synthetic consortia between cellulase-producing and non-cellulase producing microorganisms can be used to achieve substantial aerobic disintegration of polymeric natural matter with greater saccharification productivity (Chukwuma *et al.* 2021, Lin 2022). Additionally, microbial consortia adapted to grow at higher temperature become enriched with thermotolerant microorganisms producing thermotolerant enzymes with greater abilities to penetrate into crystalline cellulose, higher specific activity and stability; and thus allowing more flexibility to process configuration.

Vermicomposting consists of four stages: a preliminary stage (20-40) °C where waste degradation is initiated, a microbial metabolism-induced caloric stage (40-65) °C, a stabilization stage where earthworms are added and bacterial growth continues, and a completion stage where a constancy in composition is achieved (Abdel-Rahman *et al.* 2016, Barthod *et al.* 2018). The Red Wiggler earthworm, *Eisenia fetida*, is added to hasten the process due to the rich microbial content of its intestines and the improvement of humus quality by its excrement. The native bacteria of composting environments include both aerobes and anaerobes (Cai *et al.* 2018) whose functionality and survival in compost habitat is influenced by their environmental conditions (Antunes *et al.* 2016). The metabolic changes in the original mesophilic population lay the conditions for thermophiles to develop (Haseena *et al.* 2016). Although temperatures increase during the caloric stage, some of these lignocellulolytic microorganisms develop a tolerance to it and thrive. These microorganisms are capable of producing enzymes that would degrade plant refuse at elevated temperatures (Premlatha *et al.* 2017).

Rice serves as the staple food for more than half of the global population and as per Food and Agriculture Organization (FAO) estimation it covers an area of 164 million hectare worldwide with a production of 760 million tons (FAO 2022) (Ramos *et al.* 2023). Rice straw (RS), containing cellulose (36.2-47 %), hemicellulose (16-35 %) and lignin (5.6-36.1 %), is one of the most readily available agroresidues in the world yet it is grossly underutilized as a LCB. It is often burnt for disposal thereby leading to the emission of green house gases, imparting to environmental pollution (Ning *et al.* 2021, Hassan *et al.* 2021). Although RS can be channeled as an industrial substitute of sugary substrates for valorization (Panjiar *et al.* 2020) into animal feed, biofuels, paper and pulp, mulch and fertilizers (Goodman 2020), it shows relatively more resistance to microbial degradation primarily due to its relatively higher silica and ash content.

Present investigation was undertaken with the aim to develop a thermotolerant lignocellulolytic bacterial consortium producing thermotolerant cellulolytic and hemicellulolytic GHs with the capability to saccharify pretreated RS. The RS degradative ability of the consortium was explored through taxonomic and functional characterization of consortium metagenome. The culturable lignocellulolytic bacteria were isolated from the consortium and were phylogenetically characterized. Purification and characterization of the cellulase produced by one of the isolates was followed by gene cloning of the cellulase from the isolate. A co-culture was developed from culturable lignocellulolytic bacteria isolated from the consortium, the synergy of co-culture in the degradation of RS was evaluated and its cultural parameters were optimized for enhanced cellulase production for the purpose of saccharification of pretreated RS.

1.2. Objectives

Development of a lignocellulose degrading consortium from environmental sample,

Biochemical analysis of the consortium for production of cellulolytic GHs and depolymerization of agro-residues,

Sequencing of the consortium and analysis of the sequence for genes encoding GH and their phylogenetic origin,

Isolation and screening of lignocellulolytic bacteria from the consortium and their phylogenetic characterization,

To determine the effect of co-culturing of selected isolates on cellulase production and optimization of process parameters for enhanced cellulase production by the co-culture,

Application of cellulolytic enzymes produced by the co-culture in saccharification of agro-residues, analysis of saccharification products; and comparison of compositional differences of agroresidue before and after treatment.

Purification and characterization of cellulase from the bacterial isolate

Isolation of gene encoding cellulase from the isolate and sequence analysis.

1.3. Review of literature

Lignocellulosic biomass is constituted of agroresidues: plant parts that have been left behind after harvest such as stalks, straw, dry leaves, husk, pods and wild grasses (**Table 1.1**) which can form a sustainable alternative to our existing feedstocks (Bhuvaneshwari *et al.* 2019). They are imperishable as they can be provided on a large scale basis due to their generation in bulk. Plant refuse is typically composed of cellulose, hemicellulose and lignin in varying amounts (**Fig. 1.1**).

1.3.1. Structure of lignocellulose

LCB is the most abundant, low cost, renewable feedstock, for the development of value added products and bioenergy. Lignocellulose is the main structural component of plants cell wall and it is mainly consists of intertwining matrix of three polymers: cellulose (15-30 %), hemicellulose (25-40 %) and lignin (20-40 %) (**Fig. 1.2**). Among them cellulose and hemicellulose are linear carbohydrate polymers that can be converted to fermentable sugar, whereas lignin, a complex, hydrophobic, branched heteropolymer of phenylpropanoid units, can be extracted for production of aromatics. The structural details of cellulose, hemicellulose and lignin are as follows:

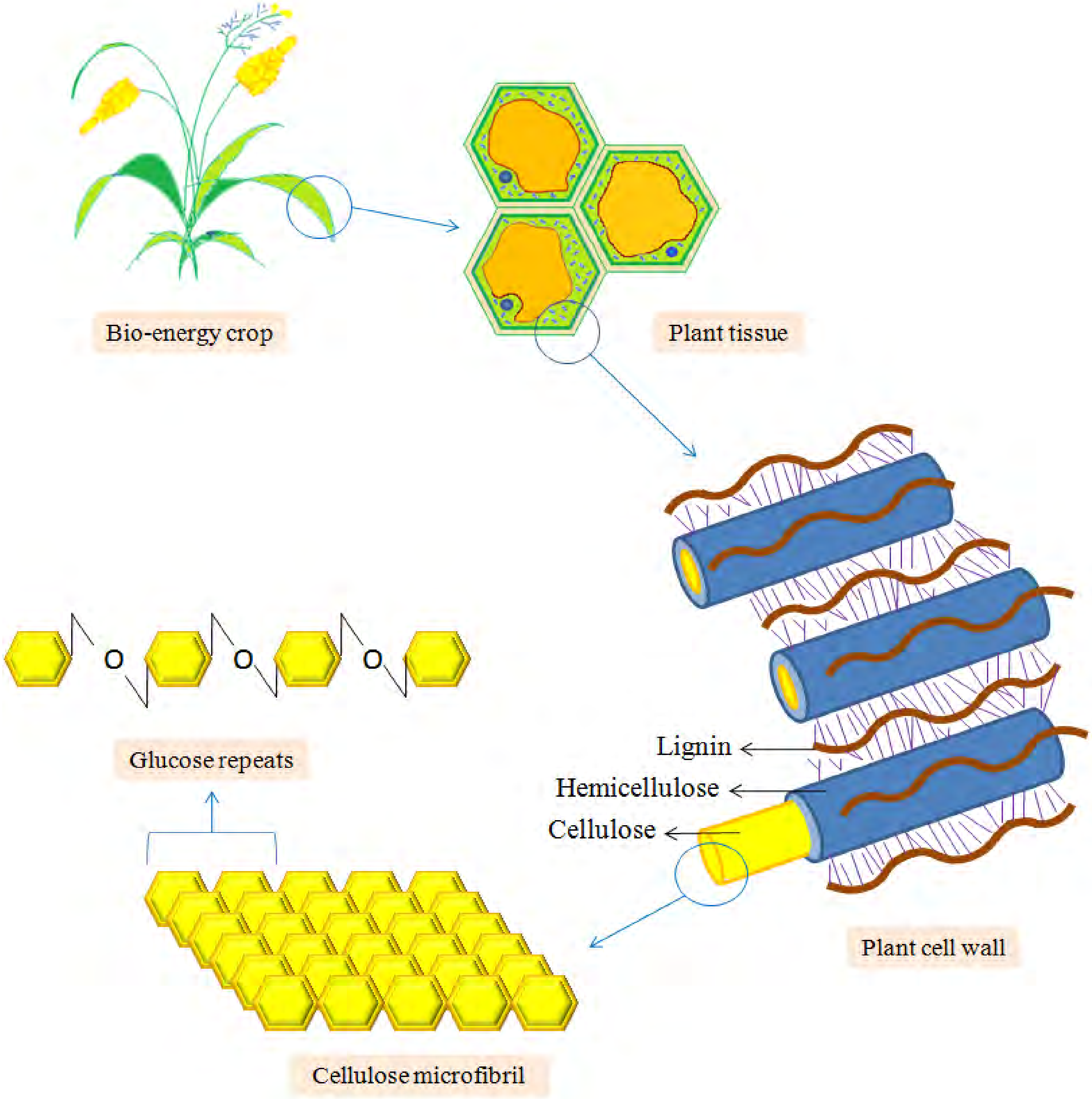

Fig. 1.1. **Structural organization of lignocellulosic biomass**

1.3.1.1. Cellulose

The linear homopolymer cellulose is made of D-glucose repeats linked through β-1,4-glycosidic bonds (Abejon *et al.* 2018). Its structure consists of a non-reducing terminus with a -C4-OH group and a reducing terminus with a -C1-OH group. The number of >C=O and -COOH groups is variable and can be altered by various chemical treatments (Gupta *et al.* 2019). These polysaccharide chains are arranged in parallel to each other to form cellulose microfibrils that are held together by intermolecular hydrogen bonds and van der Waals interactions. The microfibrils are further bundled together to form macrofibrils of cellulose.

Cellulose differs between plant species mostly due to its degree of polymerization (DP) which is the number of glucose units that make up one polymer molecule. Generally, the range for the DP of cellulose is reported to be between 800-10000 units. The cellulose molecule has four isoforms, with the most common form being Type I. Upon mercerization (alkali treatment) it converts into Type II which has greater thermodynamic stability. Both Type I and Type II would give Type III cellulose upon ammoniacal treatment, and Type III upon heating would give Type IV cellulose (Lavanya *et al*. 2011, Gautum *et al.* 2010).

1.3.1.2. Hemicellulose

Hemicellulose is a branched heteropolysaccharide composed of both pentoses and hexoses that is easily degradable through heat, chemicals or biocatalysts (Abejon *et al.* 2018). Important structural constituents include β-glucans, xyloglucans, xylans, mannans and mannoglycans (Tufail *et al.* 2018). A hemicellulose molecule may contain from 70 to 200 monomeric units of D-xylose, D-mannose, D-glucose, D-galactose, L-arabinose and 4-O-methyl-α-D-glucuronic acid. They are linked to lignin via covalent linkages and to cellulose via intermolecular hydrogen bonding (Xu *et al.* 2018). The contents of a hemicellulosic chain are dependent on its source, such as majority of xylans in grasses or glucomannans in conifers. A decrease of hemicellulosic content to half can not only increase cellulosic availability but also reduce inhibitor formation (Kucharska *et al.* 2018). Hemicellulose has a non-specific, amorphous nature, is highly solvable and provides little mechanical strength to plant tissues. Although it has lesser DP (<200) and molecular weight than cellulose, it is interconnected in nature through mostly β-1,3 bonds and form a matrix-like material where cellulosic fibrils may lie (Igbinadolor and Onilude 2013).

1.3.1.3. Lignin

Lignin forms a mechanical barrier to hydrolases, by preventing them from reaching the underlying cellulosic layers, through a process called non-productive binding (Su *et al.* 2018). Lignin is a complex, hydrophobic, branched heteropolymer of phenylpropanoid units which has the second highest availability on this planet after cellulose (Tang *et al*. 2020). It is made up of three phenylpropanoid subunits: *p*-hydroxyphenyl (H) derived from *p*-coumaryl alcohol, guaiacyl (G) from coniferyl alcohol and syringyl (S) from sinapyl alcohol, with a varying percentage of aliphatic and aromatic residues, and functional groups (-COOH, >C=O, $-CH_3OH$ and -OH) (Lu *et al.* 2017). Lignin is generated through the integration of phenoxyl ions, released by the dehydrogenation of these alcoholic subunits. A variety of end groups such as acyl, aryl glycerol, para-coumaric acid, ferulic acid, hydroxy cinnamaldehyde and hydroxycinnamyl alcohol have so far been identified depending on its end products. Due to its heterogeneity, the molecular weight of lignin is variable and usually ranges within 6700-23500 (Katahira *et al.* 2018). The diverse subunits present in this racemic macromolecule are linked through either less stable aryl-ether links (β-O-4), or through highly stable carbon-carbon bonds such as β-1, β-5, β-β' and 5-5'. The degree of condensation imparted by the presence of these highly resistant bonds is countered by the syringyl unit (S)/ guaiacyl unit (G) ratio, since the presence of a $-CH_3OH$ group at the C5 position of the syringyl (S) ring structure, prevents condensation. Therefore a higher S/G ratio would decrease the chemical inertness and in turn increase susceptibility towards various pretreatment procedures (Jardim *et al*. 2020).

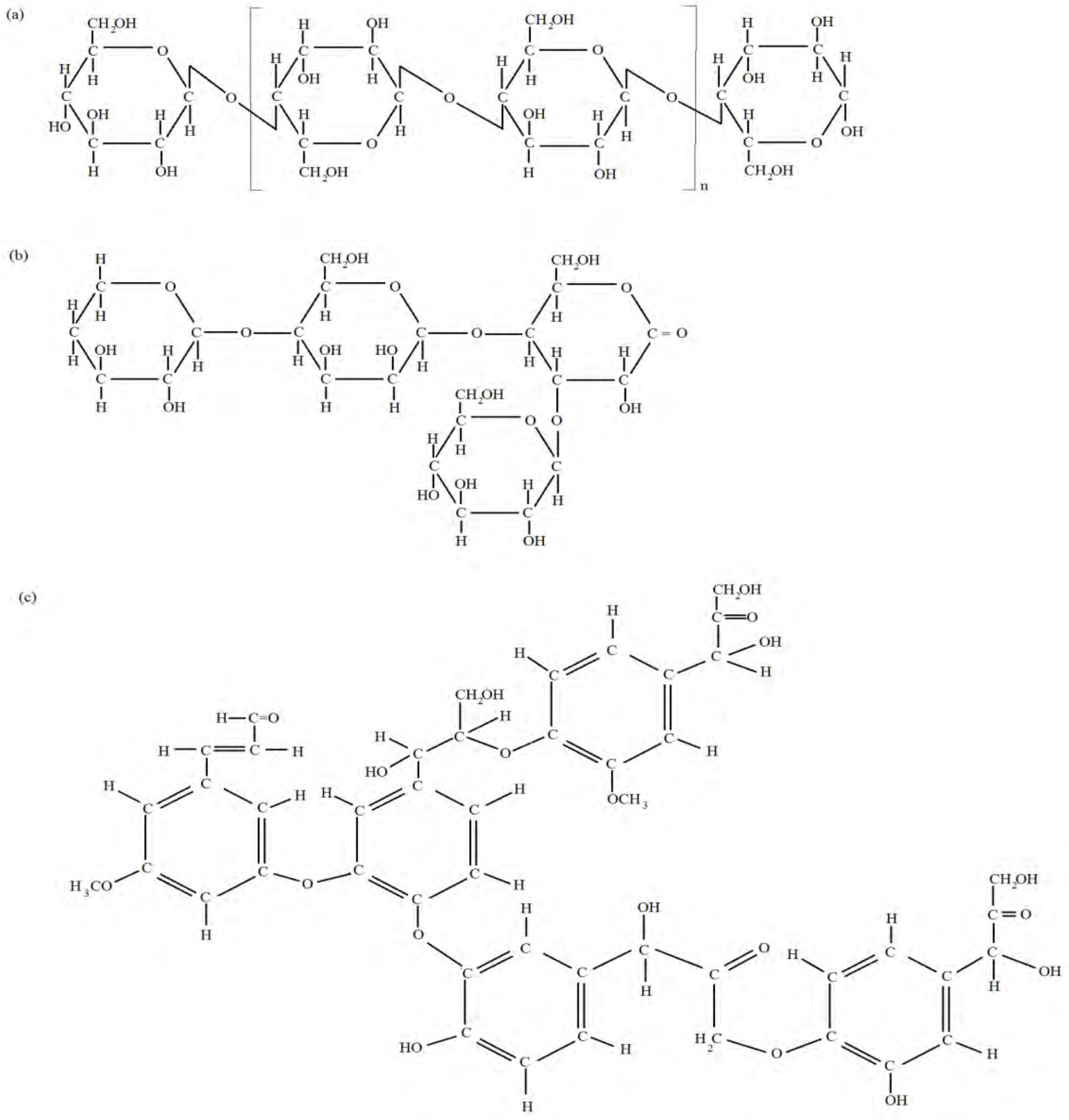

Fig. 1.2. Lignocellulosic biomass structural components (a) cellulose, (b) hemicellulose and (c) lignin

Most lignocellulosic biomass is constituted of agroresidues and the composition of cellulose, hemicellulose and lignin has been found to vary moderately between different types of agroresidues. Some of them are listed below:

Table 1.1. Agroresidues and their chemical composition

Source	Cellulose (%)	Hemicellulose (%)	Lignin (%)	References
Rice straw	(36.2-47)	(16-35)	(5.6-36.1)	(Ning *et al.* 2021)
Rice husk	(32.7-41.5)	(14-29.3)	(18.1-33.6)	(Ning *et al.* 2021)
Wheat straw	(30-43.4)	(19.5-45.2)	(7.5-22.2)	(Ning *et al.* 2021)
Jojoba hulls	17.29	1.34	35.29	(Akl *et al.* 2019)

Jatropha hulls	28.14	6.62	7.34	(Akl *et al.* 2019)
Sugarcane bagasse	(38-45)	(17-33.3)	(4-33.5)	(Ning *et al.* 2021)
Softwood	(18-38)	(15-33)	(30-60)	(Diez *et al.* 2020)
Hardwood	(43-47)	(25-35)	(16-24)	(Diez *et al.* 2020)
Sweet sorghum straw	37.74	28.07	21.48	(Dong *et al.* 2019)
Hemp	(37-46)	(14-20)	(15-22)	(Ji *et al.* 2021)
Corn fiber	(15-20)	(25-35)	(2-3)	(Kaur *et al.* 2020)
Jute	(58-63)	(20-24)	(12-15)	(Liew *et al.* 2017)
Miscanthus	(30-50)	(10-40)	(5-30)	(Xu *et al.* 2020)
Palm fruit waste	55.9	6.1	23.9	(Chin *et al.* 2021)
Barley straw	(31-45)	(27-38)	(14-19)	(Igbinadolor and Onilude 2013)
Pineapple leaf	(70-82)	18	(5-12)	(Igbinadolor and Onilude 2013)
Banana waste	(60-65)	(6-8)	(5-10)	(Igbinadolor and Onilude 2013)
Nut shells	(25-30)	(25-30)	(30-40)	(Kumar and Sharma 2017)

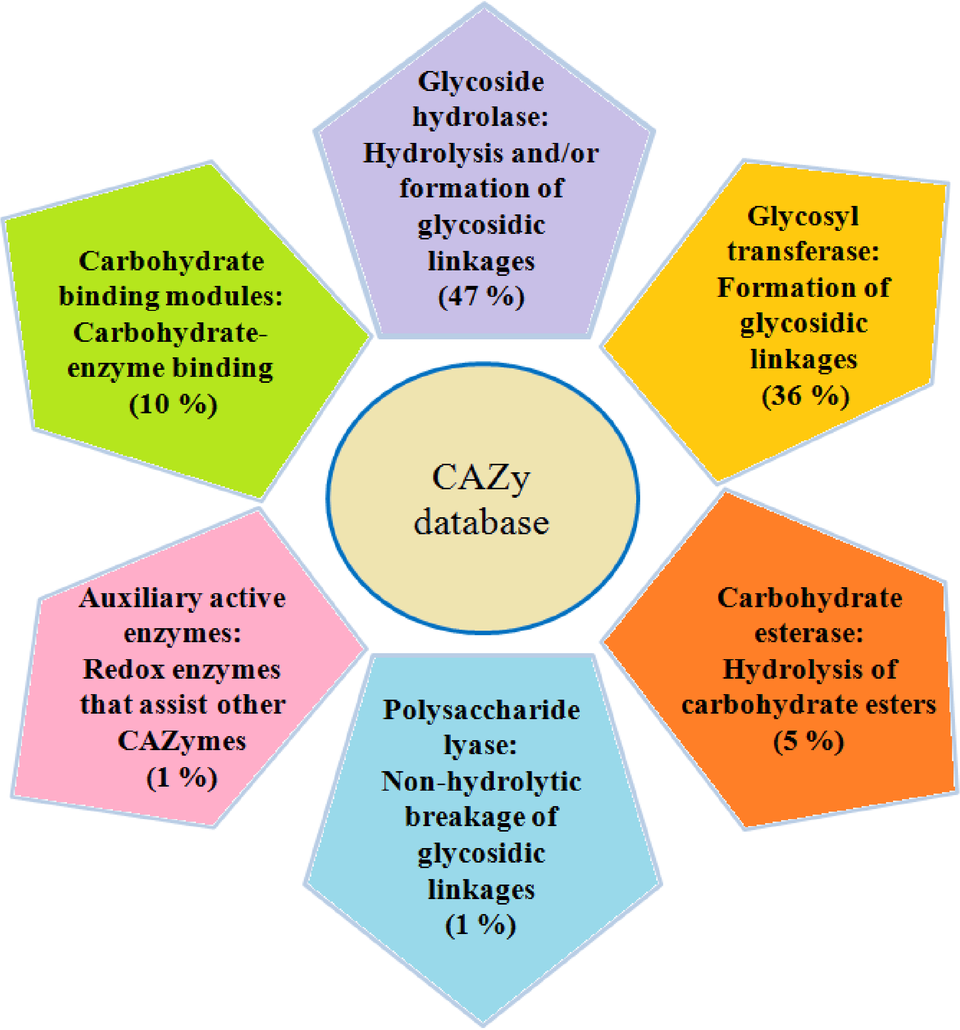

Fig. 1.3. Classification of Carbohydrate Active Enzymes (CAZymes) by the CAZy database

1.3.2. Carbohydrate Active Enzymes (CAZymes)

CAZymes participate in the construction and breakdown of carbohydrates and conjugated compounds. They have been formed from a restricted number of progenitor molecules through the modifications of their substrate specificity and thereby the products formed (Cantarel *et al*. 2009). CAZyme families contain one or more than one keystone proteins which have sequential similarities with the rest of the proteins of that family. The CAZyme database (CAZyme) (http://www.cazy.org/) (1999) facilitated by CAZypedia, records the catalytic processes and conformations of hundreds of carbohydrate-active protein families which have been classified according to conserved peptide sequences (Lombard *et al.* 2014, Barrett *et al.* 2020). CAZymes have been classified into six super families: glycoside hydrolases, glycosyl transferases, carbohydrate esterases, carbohydrate binding modules, polysaccharide lyases and auxillary active enzymes (**Fig. 1.3**) (Yu *et al.* 2021).

1.3.2.1. Glycoside hydrolase (GH)

GHs are widely present in all living organisms with the abilities to process carbohydrates. They are involved in degradation of polysaccharides, like starch, cellulose, hemicellulose and chitin. GHs also catalyze deglycosylation and reglycosylation of reactions required for the maturation of glycoproteins. GHs cleaving O-glycosydic bonds are classified as EC 3.2.1, while those breaking protein N-glycosydic bonds belong to the class EC 3.2.2. The molecular weights of GHs vary between 20-100 kDa for monomers with majority of them ranging 40-70 kDa. GHs distinctly are exo-glycosidase acting at the non-reducing end and endo-glycosidase cleaving at internal sites along an oligosaccharide either in a chemical environment-specific manner or at random (Stütz and Wrodnigg 2011).

GHs share a common catalytic mechanism based on nucleophile-mediated substitutions. Proton donations are performed by glutamate (Glu) and aspartate (Asp) residues in the active site and can be classified as either inversion or retention. In inversion pathway, a single -COOH group sequentially activates a nucleophilic water molecule to hydrolyze glycosidic linkages. There is formation of an oxocarbenium intermediate followed by only one nucleophilic substitution, bringing about changes of the anomeric structure. In retention, one -COOH group donates a proton to the substrate while the other -COOH group causes a β- to α- inversion resulting in the formation of a primary glycosylated intermediate. Then the first -COOH group accepts a proton from the nucleophilic water molecule and causes an α- to β- inversion (second displacement) via an oxo-carbonium ion formation (deglycosylation). Hence, retention pathway is a double-displacement reaction **(Fig. 1.4)** (Ardevol and Rovira 2015, Bhardwaj *et al.* 2019). The depolymerization of agroresidues requires catalytic action of GHs like xylanases and cellulases.

1.3.2.1.1. Xylanases

Xylanolytic enzymes can be sourced from fungi, bacteria, algae, protozoa, molluscs and arthropods, and they have applications in paper and textile processing, food, juicing, animal feed and pharmaceutical industries (Alvarez-Cervantes *et al.* 2016). They completely degrade hemicellulose through the collaborative action of many xylanolytic enzymes, such as endo-β-

1,4-xylanases (E.C.3.2.1.8), β-D-xylosidases (E.C.3.2.1.37), α-arabinofuranosidase (E.C.3.2.1.55) and acetyl xylan esterases (E.C.3.1.1.72) (Bouraoui *et al.* 2016, Bosetto *et al.* 2016).

Endo-β-1,4-xylanases

Endo-β-1,4-xylanases catalyze the cleavage of β-1,4-xylosidic linkages and release D-xylopyranosyl groups from xylans (Azzouz *et al*. 2021). They belong to the GH families 5, 8, 10, 11, 16, 30, 43, 51, 98 and 141 (Katsimpouras *et al.* 2019). Studies regarding their structure have revealed the presence of three types of conformations. GH10 endo-β-1,4-xylanases have higher molecular weight than GH11 endo-β-1,4-xylanases and contain an $(\alpha/\beta)_8$ barrel structure. GH11 xylanases have a right-handed β-jelly roll structure, which tend to give them a globular conformation (Liu *et al.* 2019). GH30 xylanases have their active sites in $(\beta/\alpha)_8$ barrels linked to 9β strands (Moreira and Filho 2016).

β-D-xylosidases

β-D-xylosidases cleave β-1,4 linkages in xylooligosaccharides specifically at the non-reducing termini to release xylose as the end product (Li *et al*. 2018b). They fall under GH families 1, 3, 30, 39, 43, 51, 52, 54, 116 and 120. β-D-xylosidases from all these families use retaining mechanism except for GH43 which uses invertion mechanism (Bosetto *et al*. 2016). Structural diversity is prominent among β-D-xylosidases in the form of $(\beta/\alpha)_8$ barrel in GH1, 3, 5, 30, 39, and 51; a five-sheet β propeller in GH43; an $(\alpha/\alpha)_6$ barrel in GH52 and 116, a β-sandwich in GH54 and a right-handed parallel β-helix in GH120. The catalytic region of β-D-xylosidases are situated inside a hollow which is abundant in negatively charged and aromatic amino acids and is therefore hydrophobic in nature (Rohman *et al.* 2019).

Acetyl xylan esterase

Acetyl xylan esterases are CAZyme belonging to the CE family and catalyze the cleavage of acetate from hemicellulose (Adesioye *et al*. 2018). Acetyl xylan esterases have been reported from CE families 1, 2, 3, 4, 5, 6, 7, 12 and 16, with serine (Ser) esterases belonging to CE1, 5 and 6, and aspartate esterases belonging to CE4 (Park *et al*. 2018a). Both have been found to deacylate β-D-xylopyranosides, except that while the former can act on both mono- and bi-acylated compounds, the latter can act on only mono-acylated compounds at positions 2 and 3 (Biely *et al*. 2014). CE3 and CE12 belong to the lipase GDSL protein family which indicates presence of Ser, glycine (Gly), aspargine (Asn) and histidine (His) in their active sites (Alalouf *et al*. 2011). Presence of an α/β hydrolytic scaffold has been noted in the active sites of many acetyl xylan esterases with a Ser-His-Asp triad or a Ser-His dyad, where Ser acts as the nucleophile (Komiya *et al.* 2017).

Arabinofuranosidases

Arabinofuranosidases are enzymes that catalyze the cleavage of α-1,2- and α-1,3- bonds and help to release α-arabinosyl groups from hemicellulosic side chains (Fortune *et al.* 2019). Arabinofuranosidases exist in many GH families such as 2, 3, 43, 51, 54, 62 (Zheng *et al.* 2018a). GH43 arabinofuranosidases are able to release arabinofuranosidyl groups from xylans with C-O-2 and C-O-3 substitutions and α-1,2-arabinofuranose groups from compounds with single or double substitutions (Ahmed *et al.* 2017). Arabinofuranosidases catalyze hydrolysis either through an inversion (GH43) which involves a single-displacement reaction or through a retainment procedure (GH3, 51, 54) which involves a double-displacement reaction, both of which are based on acid/base substitutions (Amore *et al.* 2015).

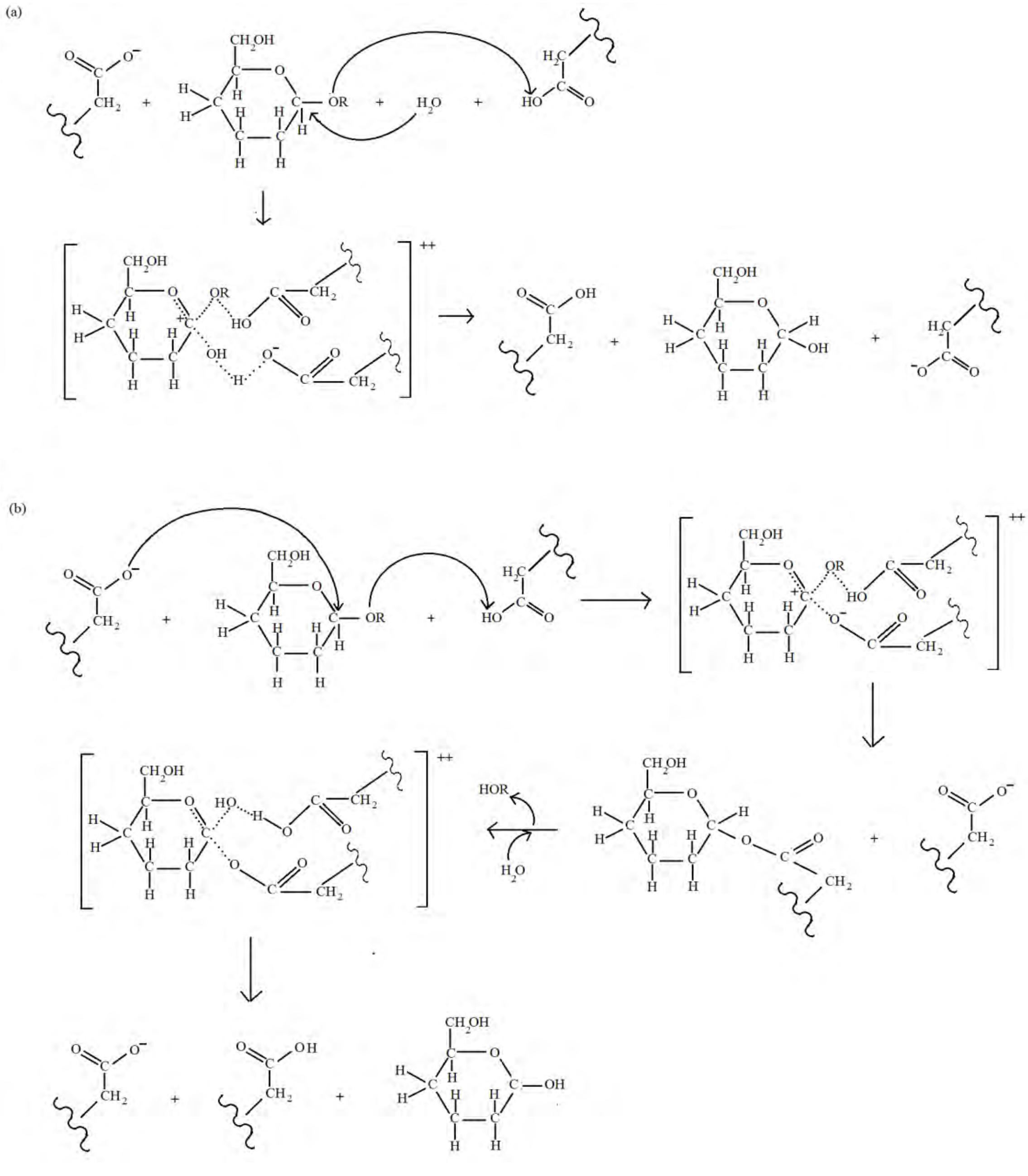

Fig. 1.4. Generalized mechanisms for enzymatic hydrolysis of glycoside (a) inverting GHs and (b) retaining GHs, figures in brackets are oxocarbenium ion-like transition states

1.3.2.1.2. Cellulases

Cellulases are spread among protozoa, fungi, and bacteria and have gained importance due to their applications in food, feed, brewery, detergent-making, paper pulping and textile industries (Cano-Ramirez *et al.* 2016, Biswas *et al.* 2020). The cellulolytic enzyme system involves three types of enzymes: endoglucanases (EC 3.2.1.4), exoglucanases or cellobiohydrolases (non-reducing end (EC 3.2.1.176) and reducing end (EC 3.2.1.91)) and β-glucosidases (EC 3.2.1.21) (Thoresen *et al*. 2021).

Cellobiohydrolase/Exoglucanase

Cellobiohydrolases or exoglucanases are cellulolytic enzymes that attack reducing or non-reducing termini of cellulosic chains and move unidirectionally along a cellulosic chain to release cello-oligosaccharides (Keller *et al*. 2020). Fungal cellobiohydrolases come from GH families 6, 7 and 48 while bacterial cellobiohydrolases come from GH families 6, 9 and 48 (Zhang and Zhang 2013). The active sites of cellobiohydrolases are formed by several helices which intertwine to form a pocket-like hollow, linked to a carbohydrate-binding domain. This domain situates the enzyme near the reducing end with the help of three hydrophobic tyrosine (Tyr) residues on its surface and a fourth one involved in van der Waals interactions (Segato *et al.* 2012).

Endoglucanases

Endoglucanases are cellulolytic enzymes that hydrolyze β-1,4-glycosidic linkages in the amorphous regions of the cellulosic chain and promote swelling and fibrillation (Bispo *et al.* 2018). Endoglucanases have earlier been reported from GH families 5, 6, 8, 9 and 45 (Alzari *et al.* 1996). All endoglucanases have a crevice in their catalytic regions, which consists of numerous attachment sites for glucose. An endoglucanase reported from *Clostridium cellulovorans* had a tryptophan (Trp) and a Tyr residue at positions 162 and 232 that formed a clamp to partly cover the active site. A singular cellulosic chain at a time would be fastened in the active site crevice due to temporary displacement of this amino acid clamp (Bianchetti *et al*. 2013).

β-glucosidases

β-glucosidases have been placed under the GH families 1, 3, 5, 9, 30 and 116. They are able to hydrolyze a wide range of bonds such as α-1,3-, α-1,4-, α-1,6-, β-1,3-, β-1,4- and β-1,6-. Other types of β-glucosidases, such as cellobiase cleaves only cellobiose, and aryl β-glucosidases cleave only when an aryl group is present (Ahmed *et al.* 2017). β-glucosidases from GH1, 5 and 30 (GH-A) have their catalytic regions within conserved $(\beta/\alpha)_8$ barrels, GH9 β-glucosidases have their's within $(\alpha/\alpha)_6$ barrels while GH3 β-glucosidases have their's distributed across a $(\beta/\alpha)_8$ barrel and an α_6/β_3 sandwich. Each region donates one carboxylate residue for enzyme activity (Cairns and Esen 2010, Srivastava *et al.* 2019).

Table 1.2. CAZymes from GH families and their characteristics

GH families	Type of CAZyme	Structural component	Catalytic mechanism
GH1	β-D-xylosidases	$(\beta/\alpha)_8$ barrel	Retention[b]
	β-glucosidase	$(\beta/\alpha)_8$ barrels	Retention[d]

GH10	Endo-β-1,4-xylanase	$(\alpha/\beta)_8$ barrel	Retention[a]
GH11	Endo-β-1,4-xylanase	β-jelly roll	Retention[a]
GH116	β-D-xylosidases	$(\alpha/\alpha)_6$ barrel	Retention[b]
	β-glucosidase	$(\alpha/\alpha)_6$ barrel[a]	Retention[a]
GH120	β-D-xylosidases	β-helix	Retention[b]
GH141	Endo-β-1,4-xylanase	β-helix[a]	not yet known[a]
GH16	Endo-β-1,4-xylanase	β-jelly roll[a]	Retention[a]
GH2	Arabinofuranosidase	$(\beta/\alpha)_8$ barrel[a]	Retention[a]
GH3	β-D-xylosidases	$(\beta/\alpha)_8$ barrel	Retention[b]
	Arabinofuranosidase	$(\beta/\alpha)_8$ barrel[a]	Retention[c], Inversion
	β-glucosidase	$(\beta/\alpha)_8$ barrel,α_6/β_3 sandwich	Retention[d]
GH30	Endo-β-1,4-xylanase	$(\beta/\alpha)_8$ barrels, β strands	Retention[a]
	β-D-xylosidases	$(\beta/\alpha)_8$ barrel	Retention[b]
	β-glucosidase	$(\beta/\alpha)_8$ barrels	Retention[a]
GH39	β-D-xylosidases	$(\beta/\alpha)_8$ barrel	Retention[b]
GH43	Endo-β-1,4-xylanase	β sheet propeller[a]	Inversion[a]
	β-D-xylosidases	β sheet propeller	Inversion[b]
	Arabinofuranosidase	β sheet propeller[a]	Inversion
GH45	Endoglucanase	catalytic pocket	Inversion[a]
GH48	Cellobiohydrolases	α-helical pocket	Inversion[a]
GH5	Endo-β-1,4-xylanase	$(\beta/\alpha)_8$ barrel	Retention[a]
	β-D-xylosidases	$(\beta/\alpha)_8$ barrel[a]	Retention[b]
	Endoglucanase	$(\beta/\alpha)_8$ barrel[a]	Retention[e]
	β-glucosidase	$(\beta/\alpha)_8$ barrels	Retention[a]
GH51	Endo-β-1,4-xylanase	$(\beta/\alpha)_8$ barrel[a]	Retention[a]
	β-D-xylosidases	$(\beta/\alpha)_8$ barrel	Retention[b]
	Arabinofuranosidase	$(\beta/\alpha)_8$ barrel[a]	Retention[c], Inversion
GH52	β-D-xylosidases	$(\alpha/\alpha)_6$ barrel	Retention[b]
GH54	β-D-xylosidases	β-sheet sandwich	Retention[b]
	Arabinofuranosidase	β-jelly roll[a]	Retention[c], Inversion
GH6	Cellobiohydrolases	α-helical pocket	Inversion[a]
	Endoglucanase	$(\beta/\alpha)_8$ barrel[a]	Inversion[a]
GH62	Arabinofuranosidase	β sheet propeller[a]	Inversion[a]
GH7	Cellobiohydrolases	α-helical pocket	Retention[a]
GH8	Endo-β-1,4-xylanase	$(\alpha/\alpha)_6$ barrel[a]	Inversion[a]
	Endoglucanase	$(\beta/\alpha)_8$ barrel[a]	Inversion[a]
GH9	Cellobiohydrolases	α-helical pocket	Inversion[a]
	Endoglucanase	$(\beta/\alpha)_8$ barrel[a]	Inversion[a]

	β-glucosidase	$(\alpha/\alpha)_6$ barrels	Inversion[d]
GH98	Endo-β-1,4-xylanase	$(\alpha/\beta)_8$ barrel[a]	Inversion[a]

[a](CAZy database), [b](Rohman *et al.* 2019), [c](Amore *et al.* 2015), [d](Srivastava *et al.* 2019), [e](Bianchetti *et al.* 2013)

1.3.2.2. Auxiliary active enzymes (AA)

Apart from cellulases and hemicellulases, there are auxiliary active enzymes that are a recently classified category of CAZymes. They may be directly or indirectly associated with lignin degradation (Sützl *et al.* 2018).

1.3.2.2.1. Lignin-degrading auxiliary enzymes (LDA)

LDA do not directly cause lignin breakdown. Instead they assist other enzymes that do so, serially, often generating peroxide ions. They include glyoxal oxidase (EC 1.2.3.5), aryl alcohol oxidase (EC 1.1.3.7), pyranose oxidase (EC 1.1.3.10), cellobiose dehydrogenase (EC 1.1.99.18), glucose oxidase (EC 1.1.3.4) and glucose dehydrogenase (EC 1.1.5.9) (Janusz *et al.* 2017, Cerutti *et al.* 2021).

Glyoxal oxidase

Glyoxal oxidase is a mono-copper metalloproteinaceous enzyme that belongs to a class of radical copper (Cu) oxidases in the AA5 family. It converts -R–HC=O groups to -R-COOH- groups in 2-oxoaldehydes, α and β-hydroxycarbonyl compounds, including methylglyoxal and glyoxal (Daou *et al.* 2019, Janusz *et al.* 2017). It liberates H_2O_2 and interacts with lignin, manganese and versatile peroxidases which directly participate in lignin degradation (Wohlschlager *et al.* 2021). They have a single highly conserved copper-binding site formed by a Cysteine-Tyrosine-Histidine-Histidine-Tyrosine (Cys-Tyr-His-His-Tyr), the last one attached by a parallel bond (Yin *et al.* 2015).

Aryl alcohol oxidase

Aryl alcohol oxidases are FAD-containing oxidative enzymes that belong to the family AA3, subfamily 2 (AA3_2). They convert allylic and aromatic 1° alcohols to their respective aldehydes, coupled with the reduction of O_2 to H_2O_2. Their amino terminal FAD-binding site is highly conserved and is comprised of alternating inter-hydrogen linked β-sheets and α-helices (Rossman fold) that fasten ADP. The carboxy terminal active site which is more varied in structure does contain a conserved His that oxidizes substrate and re-oxidizes FAD (Urlacher and Koschorreck 2021).

Pyranose oxidase

Pyranose oxidase is a flavin-dependent enzyme belonging to the AA3_4 subfamily of GMC oxidoreductases that oxidizes monosaccharides while reducing O_2 to H_2O_2. Its structural arrangement is alike aryl alcohol oxidases, but unlike fungal pyranose oxidases which are tetramers, bacterial pyranose oxidases are monomers and bind FAD in a non-covalent manner (Abrera *et al.* 2020).

Cellobiose dehydrogenase

Cellobiose dehydrogenase is an enzyme from the AA3_1 subfamily that oxidizes cellobiose, lactose, manno- and cellodextrins to lactones while reducing metals, cytochrome *c* or 2,6-dichloroindophenol (Nghi *et al.* 2021). Its structure is comprised of a FAD-domain linked to a cytochrome domain which contains heme B (Felice *et al.* 2021).

Glucose dehydrogenase and glucose oxidase

Glucose-1-oxidase reduces quinoid products (1,4-benzoquinone, 2,6-dimethoxy-1,4-benzoquinone, 2-methoxy-1,4-benzoquinone) of laccase and accelerates lignosulfonate breakdown (Ander and Marzullo 1977). This is done through the H_2O_2-activation of peroxidases. Glucose dehydrogenase oxidizes D-glucose coupled with the reduction of FAD and quinones. Both these enzymes belong to the AA3_2 subfamily (Cerutti *et al.* 2021).

1.3.2.2.2. Lignin-modifying auxiliary enzymes (LME)

Apart from LDA, there are actual ligninolytic enzymes that directly participate in the structural modification of lignin. LME include laccase (EC 1.10.3.2), lignin peroxidase (LiP) (EC 1.11.1.14) and manganese peroxidase (MnP) (EC 1.11.1.13) (Casciello *et al.* 2017).

Laccase

Laccase or benzenediol:oxygen oxidoreductase oxidizes phenolic, non-phenolic, aromatic and non-aromatic substrates containing the functional groups methyl ($–OCH_3$), amino ($-NH_2$), diamino ($-(NH_2)_2$), hydroxyindols, complex metal ions such as octacyanomolybdate ($-[Mo(CN)_8]^{4-}$), ferrocyanide ($-[Fe(CN)_6]^{4-}$), hexacyanoosmiate ($-[Os(CN)_6]^{4-}$), etc while simultaneously reducing O_2 to H_2O_2. Hydroquinone, catechol, guaiacol, syringaldazine, 2,4-dimethoxyphenol are the common substrates for laccase (Chauhan *et al.* 2017). Laccases catalyze the C_α-C_β and aryl-alkyl cleavages of β-O-4 and β-1 bonds in lignin (Youn *et al.* 1995). They are capable of completely degrading lignin by themselves without being coupled with other enzymes (Zhang *et al.* 2021). Laccases belong to the AA1 family of the CAZy database (Silva *et al.* 2021). They are multicopper enzymes with four types of catalytic Cu. Blue Cu (Type I, paramagnetic, absorbs at 610 nm) oxidizes substrate via a trigonal pyramid. It is linked equatorially to two conserved His, one conserved Cys and axially to either a methionine (Met) (bacteria) or leucine (Leu)/ phenylalanine (Phe) (fungi). A non-blue Cu (Type II, paramagnetic) containing double His forms a trinucleated cluster with Type III to reduce O_2. Type III (diamagnetic, absorbs at 330 nm) is a Cu pair with anti-ferromagnetic coupling stabilized by an -OH group. It is linked to six His residues (Arregui *et al.* 2019, Claus 2004).

Lignin peroxidase

LiP or diaryl propane oxygenase catalyzes the breakage of C-C and β-O-4' bonds through the single electron oxidation and demethylation of both phenolic and non-phenolic lignin-derived compounds, that includes aromatic alcohols (catechol, guaiacol, vanillyl alcohol), benzylmethyls, β-O-4-linked arylglycerol aryl ethers, N and benzyl-substituted anilines, to ketones and aldehydes. They are globular proteins containing an immobile heme group linked to two narrow tunnels. The protein structure is folded into proximal and distal regions through 8α (major), 8α (minor) and 3β (short-length, antiparallel) sheets (Pham et al. 2021, Falade *et al.* 2017). LiP belongs to the AA2 family of the CAZy database (Sanchez-Ruiz *et al.* 2021).

Mn peroxidase

Mn peroxidase (MnP) or Mn^{2+}:H_2O_2 oxidoreductase is a H_2O_2-activatedglycoproteinaceous enzyme that catalyzes the breakage of the C_α-C_β bond (Manyapu *et al.* 2022). It degrades chlorophenols, polycyclic aromatics (phenanthrene, pyrene, fluoranthene, anthracene), nitro-aromatic compounds (2,4,6-trinitrotoleune) and high molecular weight chlorolignins coupled with the oxidization of Mn^{2+} to Mn^{3+} (Qin *et al.* 2014). Its structure is comprised of 10α (major) and 1α (minor) and encloses a heme group. MnP has 5 disulfide linkages, one of which joins the carboxy terminal Cys341 and Cys348 residues. There is a peroxide-binding cleft formed by a conservative His46-Arg42 association, catalytic site formed by a Asp242-His173 association and a Mn^{2+}-binding cleft formed by a Glu35-Glu39-Asp179 association (Chandra *et al.* 2017). MnP belongs to the AA2 family of the CAZy database (Silva *et al.* 2021).

1.3.3. Applications of cellulolytic and hemicellulolytic enzymes

Cellulolytic and hemicellulolytic enzymes find many applications in industry. These include animal feed, food processing, detergents and textiles, biodeinking and biopulping, etc.

1.3.3.1. Animal feed

The nutritional properties of animal feed are inversely proportional to its non-starch polysaccharide (NSP) content. NSP increases feed viscosity and reduces surface area in gut (Alagawany *et al.* 2018, Moftakharzadeh *et al.* 2019). Enzyme supplementation improves feed assimilation due to breakdown of viscous materials thereby enhancing chyme flow in the gut (Kundu *et al.* 2012). Exogenic application of fibrolytic hydrolases degrades dietary fibres and polysaccharides in feed and releases monosaccharides (Lewis *et al.* 1996). Intake of partially hydrolysed feed gives a distinct rise in gut pH, which favours fibrolysis over amylolysis. Cellulase-treated feed increased the quality and quantity of milk and altered its lipid profile as in it had a diminished percentage of saturated fatty acid (Rojo *et al.* 2015). β-glucanase treatment of feed increased poultry body weight, energy intake, increased life span, plumage lustre and body weight in turkeys and caused weight gain and improved protein metabolism in piglets (Almirall *et al.* 1995, Odetallah *et al.* 2002, Bedford *et al.* 1992).

Similar results have been achieved from cows as well where a mix of cellulase and xylanase added to corn and alfalfa based feed, increased milk quantity (~110 %), lipid (120 %) and protein contents (113 %) (Schingoethe *et al.* 1999).

1.3.3.2. Food Industry

Machine pressing and artificial clarifiers can be substituted by macerating enzymes which are a combination of cellulase, hemicellulase and pectinase. Enzymatic clarification enhances the flavour of the juice through the release of various aromatic compounds (Kumar 2015). A pre-pressing enzymatic treatment partially disintegrates the cell walls, increases juice yield and brightness of the product (Trappey *et al.* 2007). Macerating enzymes have been used along with standard procedures such as hot water treatment to increase juice yield in apples by 137 % and on hydrated date pulp by 68 %. Apart from juices, they are also employed in nectar and puree manufacture as homogeneity regulators, carotenoid extraction from cashew apple wastes and in flavonoid extraction from flower petals and seeds where they offer advantages such as reduced processing time and thermal damage (Sharma *et al.* 2014, Toushik *et al.* 2017, Macedo *et al.* 2015, Sajith *et al.* 2016). Also glucanases find extensive applications in wine and breweries as they help to degrade complex polysaccharides. They are either added during the malting step to reduce viscosity and improve filtration efficiency or in the mashing step where they degrade cellulose into dextrins (Sharada *et al.* 2014, Gomaa 2018).

1.3.3.3. Detergents and textiles

Addition of cellulases to detergents requires them to be thermoalkalitolerant and be resistant to detergent components such as wetting agents, frothing agents, bleaches, enzymes, builders and brighteners (Adrio and Demain 2014, Gurkok 2019). Efficient scouring of cotton, polyester and nylon blends has been performed by xylanase in combination with pectinase, cellulase, lipase, protease while cellulases have been earlier used for depiling cotton, wool, Lyocell, viscose, and bast materials (hemp, ramie, flax, jute) (Chatha *et al.* 2017, Simic *et al.* 2015). Biostoning of denim fabric is preferred over pumice stones as it does not induce frictional pressure and is ecofriendly. It functions by retarding defibrillation to prevent binding of grime and gently clean out the dirt, help to brighten colour and prevent backstaining (Mojsov 2012). Alkaliphilic and neutrophilic cellulases with tightly packed non polar and aromatic amino acids on their surface are employed to prevent backstaining of denim. These residues create regional hydrophobicity that attracts and bind to indigo molecules thus preventing their redeposition (Anish *et al.* 2006, Gusakov *et al.* 2000).

1.3.3.4. Biodeinking

Hydrolytic treatment is particularly effective on xerographic and laser prints which are otherwise difficult to deink via chemical methods, due to being heat-infused onto the paper while printing (Lee *et al.* 2011). It is cheap, eco-friendly and offers advantages such as increased brightness, removal of fine fibrils from pulp thereby reducing viscosity, improvement of drainage efficiency, smoothness, mechanical strength and elasticity of paper through the influencing of hydrogen linkage and surface contact between fibrils (Zhao *et al.*

2006, Efrati *et al.* 2013, Akbarpour *et al.* 2018). Xylanase upon combination with laccase, gave 65.8 % removal of ink from old newsprint pulp with a 21.6 % increase in whiteness (Virk *et al*. 2013). Paper waste contains fillers, sizing starch, gums or glues and various types of coloured inks (Imamoglu *et al*. 2013). Their removal is performed via a three-step flotation consisting of loosening of ink granules from paper, their attachment to the foam and their removal along with the foam. Wetting agents are added in all steps to prevent back-staining of paper (Zhao *et al.* 2004). Cellulases are applied during the first step where they sever intermolecular links between fibrils at the sites of ink-paper binding (Mayeli and Talaeipour 2015, Dutt *et al*. 2012). Detachment of ink granules from paper decreases their hydrophobicity and eases their separation (Abraha and Kifle 2019).

1.3.3.5. Biopulping

Enzymes used for bioleeching are required to be resistant to heavy metals, organosolvents, benzoic compounds, etc which are present in the pulp (Kumar and Rani 2019). Xylanases are known to remove lipids and triglycerides which have a higher degree of hydrophobicity (Adrio and Demain 2014). Wood pulp contains both longitudinal fibres and spicules which retard drainage and exert on the machines. This was solved by cellulase treatment which improved drainage efficiency by ~20 % (Jeffries 1992). Alkaliphilic endoglucanases are usually preferred in pulp treatment. They have been noted to impart mechanical strength of pulp, reduce ink residues and increase brightness in a shorter time period (0.5 h) (Wang *et al.* 2017, Biswas *et al.* 2019, Lee *et al*. 2017b).

1.3.3.6. Valorization

Substantial amounts of energy can be generated through a process known as valorization. It is defined as the proper recycling of discarded biodegradable matter that is routinely generated from farming and commerce. Although valorization was initially considered related to organic waste management, it has regained popularity due to the conversion of agrowastes into beneficial substances including sugars, amino acids, organic compounds and food supplements, second generation biofuels, green chemicals, biofabrics, biocatalysts, biodegradable plastics and biosorbents (Arancon *et al.* 2013, Ahmed *et al*. 2017, Diaz *et al.* 2018, Diez *et al*. 2020, Tsai *et al*. 2021). Harvest residue or miscellaneous weeds can be used as raw material as they are not in competition with our demand for food. Their conversion to value added products involves three steps: pretreatment of agromaterial, enzymatic processing of the available cellulosic chains to release the 5- or 6-carbon sugars and the microbial fermentation of these released sugars (Odelade *et al.* 2016, Anwar *et al.* 2014). Monosaccharides released through the combined action of cellulases and xylanases, form the substrates for microbial fermentation (Ahmed *et al.* 2017).

1.3.4. Pretreatment

Untreated agricultural wastes are not amenable to enzymatic saccharification due to the recalcitrance imparted to them by their lignin content, the crystallinity and the DP of their cellulosic content (Xu *et al*. 2020). Lignin in particular could decrease the sugar yield to (40-50) % (Chambon *et al.* 2019). In order to facilitate sugar release, plant refuse is first subjected to a reductive preparation and a structural modification (pretreatment). The purpose of pretreatment is to decrease lignin content and make cellulosic portions accessible to the enzymes (Yang *et al*. 2016b, Alias *et al.* 2021). Efficient cellulosic hydrolysis is possible when the lignin and the hemicellulosic portions have been removed or at least significantly reduced (Zheng *et al*. 2018b). The agroresidue is prepared through chopping or grinding that decreases the fragment size and increases the surface area for hydrolysis (Gandla *et al.* 2018). Pretreatment is then carried out through physical, chemical, physico-chemical, biological or combined procedures (**Fig. 1.5**) (Chopda *et al.* 2020).

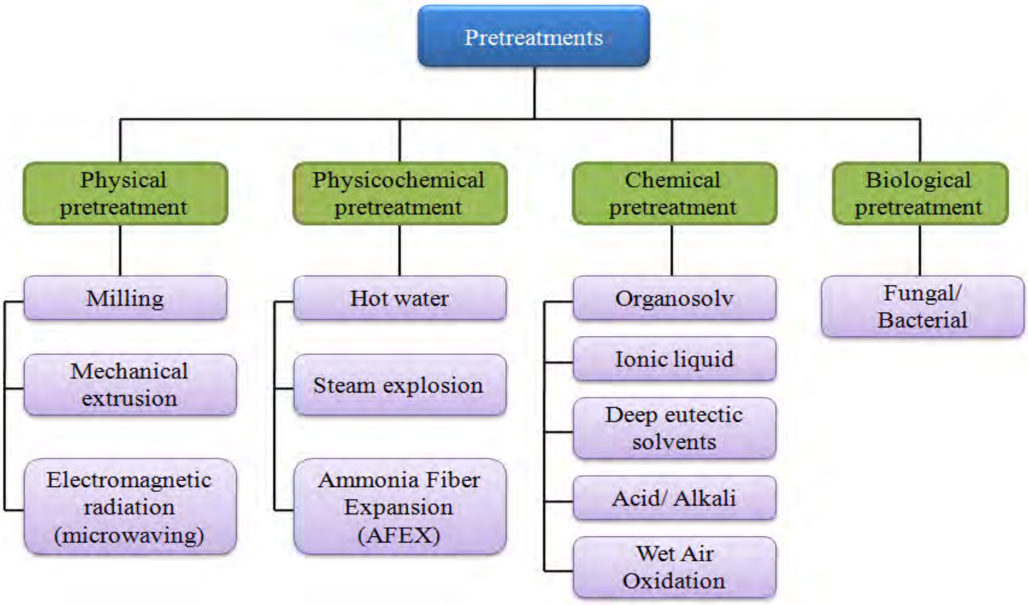

Fig. 1.5. Pretreatments of lignocellulosic biomass

1.3.4.1. Physical pretreatments

Physical pretreatments are of three main types: milling, microwaving and mechanical extrusion. They have been used on a large scale on a variety of agro-residues.

Milling

Milling can be of several types, such as ball, hammer, colloid, vibrational, two-roll centrifugal, attrition, pin and knife milling (Dey *et al.* 2020). It can enhance the rate of bio-conversion, pore formation and flow rate, reduce particle size, density and localization, while modifying cellulosic crystallinity and chain length to increase surface area for hydrolytic reactions. Despite involving no chemicals, milling (grinding) is not popular due to its energy-intensive nature (Amin *et al.* 2017, Shahi *et al.* 2020).

Mechanical extrusion

Another procedure is mechanical extrusion which involves shearing and mixing of agro-residues under high temperatures (>300 °C). It causes sequential deconstruction and reconstruction of cellulosic fibers and increases susceptibility of the residue to enzymes due to disruption of both the amorphous and the crystalline fractions of the cellulosic chain. Alike milling, mechanical extrusion is also an expensive, difficult and energy-intensive process (Maurya *et al.* 2015, Kumar and Sharma 2017).

Radiation

In addition to these two methodologies, electromagnetic radiation (microwaving) and ultrasonic waves (ultrasonic pretreatment) can also be used to penetrate and weaken the bonds between the lignocellulosic components, namely lignin. Microwaving has previously enhanced the saccharification of RS by 200 % and sugarcane bagasse by 300 % (Pattanaik *et al.* 2019).

1.3.4.2. Physico-chemical pretreatments

Physico-chemical pretreatments include hot water hydrolysis, Ammonia Fibre Expansion (AFEX) and steam explosion.

Hot water hydrolysis

Hot water (170-200 °C) hydrolysis is non-toxic, inexpensive and has been performed on rice, corn and wheat straws. It can liberate fermentable saccharides within relatively short time periods of 10-40 min, while converting hemicellulosic functional groups to acetate and organic acids which aid in further dissolution (Sun *et al.* 2020). Hot water is able to infiltrate and deconstruct plant residue by increasing the pore formation and separation of the cellulosic fibrils. Although it is efficient in increasing the availability of the cellulosic portions, it has limited ability to prevent lignin redeposition and formation of inhibitors (Bensah and Mensah 2018).

Ammonia Fibre Expansion (AFEX)

In AFEX, the ammonium and hydroxyl groups formed due to exothermic reaction of ammonia with moisture, would be used to attack hemicellulosic acetyl groups and ester bonds in lignin. Hydrated LCB would be percolated by pressurized liquid ammonia 60-100 °C for 5-30 min followed by a sudden decrease in pressure (Latif *et al.* 2018). Ammonia is recyclable and can be used for high substrate loadings (~18 %), and combined with elevated temperatures it can reduce the enzyme dosages in downstream processing while modifying hemicellulosic reducibility, substrate availability, crystallinity and length of cellulosic fibrils and lignin percentages (Avci *et al.* 2019). Corn stover subjected to AFEX at 120 °C for 30 min, retained only 56 % of its original lignin content but almost all of its sugar content (Sousa *et al.* 2016).

Steam explosion

Steam explosion or autohydrolysis subjects the biomass to repeated cycles of pressurization (0.69-47 atm) and depressurization due to steam (160-260 °C). The pores present in the biomass allow the entry of steam which would expand thereby causing the breakage of hydrogen bonds and causing more pores to form for steam to enter (Taylor *et al.* 2019). Steam explosion can be used in combination with other pretreatments, such as alkali. Steam exploded RS (170 °C, 10 min) followed by 2 % KOH removed lignin by 85 % (Banoth *et al.* 2017).

1.3.4.3. Chemical pretreatments

Chemical pretreatment mainly include organosolv, ionic liquids (IL), acid, alkali and wet oxidation pretreatments.

Organosolving

Organosolving is known to yield highly purified cellulosic substrate within a short time interval of 30-120 min. It is carried out at temperature ranging between 80 °C to 200 °C and can be performed with glycerol, acetate, acetone, gamma valerolactone, ethanol, methanol, acetone and ethylene glycol (Chin *et al.* 2021). Organic solvents act on intermolecular bonds of lignin and separate them from hemicellulose, often with the help of mediating compounds such as sulphuric acid (H_2SO_4), salicylate, oxalate and hydrochloric acid (HCl) (Capolupo and Faraco 2016).

Ionosolv

Ionosolv pretreatment utilizes IL which are aqueous solutions of thermostable salts and are preferred over organic solvents for their resistance to vaporization, inflammation and their recyclability. Over 90 % sugar yield has been achieved from sugarcane bagasse by using low dosages of 1-ethyl, 3-methyl imidazolium chloride ([Imim][OAc]) (Chambon *et al.* 2019). Lignin removals of 53 % in bamboo, 46 % in RS and 32 % in maplewood flour were achieved by 1-propyronitrile-3-benzyllimidazolium chloride ([C2CNBzim]Cl), cholinium argininate ([Ch][Arg]) and 1-butyl-3-methylimidazolium acetate ([Bmim]Ac) treatments, respectively. Addition of mediating compounds such as dimethyl formamide, dimethyl sulfoxide (DMSO), 1,3-dimethyl-2-imidazolidinone and sodium dodecyl sulphate (SDS) bring down the temperature and IL concentrations required for the reaction (Hou *et al.* 2017).

Deep eutectic solvents (DES)

DES are similar to IL and are proportionate mixtures of 2 or more solvents so that the resulting DES has a much lower freezing temperature than the components. They are being considered as a suitable replacement of IL due to reduced toxicity, cost and high biodegradability (Wang and Lee 2021). DES components are either hydrogen-bonding donors or acceptors whose interactions ensure degradation of LCB through weakening of its constituent hydrogen bonds. DES pretreatment involves prolonged incubation at 60-130 °C wherein it is not stable. Also it is moisture-absorptive (Chen and Mu 2019). Processes involving DES are planned with its drawbacks in consideration.

Acid

Acid pretreatement has involved the usage of acetic acid, sulfuric acid and phosphoric acid (Mafa *et al.* 2020). Its efficiency depends on its concentration and reaction temperature. An increase in temperature brings about a loss of saccharides and therefore, it is ideally carried out at 60-110 °C to give upto 70 % yield of glucose. Lower concentrations of acid (3-10 %) require higher temperatures such as 60 °C while higher concentrations (>30 %) can be applied at room temperature (Kucharska *et al.* 2018).

Alkali

Alkali pretreatment can be carried out under mild temperature and pressure, and exclusively removes lignin, with minimal inhibitor formation. Even though all three of sodium hydroxide (NaOH), calcium hydroxide ($Ca(OH)_2$) and potassium hydroxide (KOH) can be used to perform it, $Ca(OH)_2$ is preferred to NaOH or KOH (Romani *et al*. 2016, Lukajtis *et al.* 2018). Alkali pretreatment acts through the saponification of hemicellulosic uronic ester and acetyl linkages. Efficiency of alkali pretreatment is inversely proportional to the lignin content of the biomass and works better for hardwoods than softwoods (Chen *et al.* 2013).

Wet air oxidation (WAO)

Apart from these techniques, there is WAO. WAO is a type of pretreatment that is performed by gaseous oxygen in a temperature range of 125-320 °C under 4.93-19.74 atm pressure. It has the advantages of being exothermic, reduced risk of inhibiting compound formation, and obtaining highly pure cellulose and hemicellulose fractions (Sharma *et al.* 2015).

1.3.4.4. Biological pretreatments

Biodegradation of LCB by microbes has evolved in nature in an environmentally friendly way. Certain microorganisms display extensive biomass degradation during their growth and can be used for LCB modification. Microbes or microbial enzymes based pretreatment exhibits higher reaction rate and specificity, mild reaction conditions, lower energy requirement and facilitated access of hydrolytic enzymes to cellulosic polymers (Ummalyma *et al.* 2019). Although both bacteria and fungi can have ligninolytic activity, bacteria-mediated degradation of lignin occurs at a slower rate and to a lesser extent as compared to fungal specimen (Rico-Garcia *et al.* 2020). Among fungi, white rot fungi are more desirable despite them affecting mainly cellulose as compared to brown and soft rot fungi (Baruah *et al.* 2018). Bacteria from the phyla Actinobacteria and Eubacteria are able to degrade lignin and can be used as isolates, consortium, or co-cultures with other bacteria and fungi, and as enzymatic extracts. Consortium exhibits greater stability over pH, temperature and additives, and can be prefered over single isolates (Ferdes *et al.* 2020). Biological pretreatment has been performed by ligninolytic bacteria, such as *Bacillus sp, Pseudomonas sp, Aeromonas sp, Streptomyces sp* and by white rot fungi, such as *Phanerochaete chrysosporium, Pleurotus ostreatus, Trametes versicolor*, *Coriolus versicolor, Fusarium sp, Cyathus stercoreus* and *Ceriporiopsis subvermispora* (Guo *et al.* 2019, Tsegaye *et al.* 2019, Poszytek *et al.* 2016). These microorganisms produce an array of oxidative enzymes like, laccase, LiP, MnP, versatile peroxidase and dye decolorizing peroxidase that act on lignin through radical-based oxidation (Zhuo *et al.* 2018). Biological pretreatment requires no prior reductive preparation and does not affect the hemicellulosic portion. Degradation can take three to four weeks depending on the pH and nitrogenous supplementation of culture medium (Wagner *et al.* 2018).

1.3.5. Saccharification and fermentation

Most strategies for sugar release focus on the hydrolysis of cellulosic portions and are dependent on agrowaste characteristics, such as their water percentage, bulk weight per volume, degree of robustness and caloricity (Aboytes-Ojeda *et al.* 2016). This is performed by two types of enzymes: core enzymes which include cellulolytic enzymes, and the accessory enzymes which include xylanolytic enzymes. An enzyme mix consisting of both is necessary for the efficient degradation of biomass (Thite and Nerurkar 2019). Enzymatic hydrolysis is advantageous as it allows for a decrease in media viscosity due to its influence on concentration, particle size and bond formation between particles (van der Zwan *et al.* 2020).

Saccharification and fermentation were originally performed separately due to the different temperature requirements of each step (Olguin-Maciel *et al.* 2020). Sequential hydrolysis and fermentation (SHF) involved a primary dosage of cellulolytic enzymes and risked nutritional reduction of medium along with generation of furfurals and related microbial growth inhibitors. The obtained hydrolysate was heat-sterilized and the purified saccharides were introduced to the fermentative microorganisms in a separate bioreactor (Dong *et al.* 2016). SHF is disadvantageous as there is feedback inhibition by end products such as glucose and cellobiose (Du *et al.* 2014).

Simultaneous saccharification and fermentation (SSF) has several advantages over SHF. SSF allows for high (>15 %) solid loading which has been proven useful for ethanol production. High solid loading regulates hydration which in turn enhances distillation (Du *et al.* 2014). It is able to prevent feedback inhibition due to the rapid consumption of monosaccharides by fermenting microorganisms and maintains ideal medium conditions such as pH. It is usually carried out around 37 °C or around 55 °C (Barahona *et al.* 2020). SSF has been found to increase the rate of saccharification by (15-30) % (Choudhary *et al.* 2016). However, fermentative microorganisms used in SSF cannot be separated from the saccharification mix and hence SSF lacks recyclability. Simultaneous Saccharification, Filtration and Fermentation (SSFF) is a modification of SSF where a cross-flow membrane located between bioreactors, could filter the hydrolysate before fermentation while returning unreacted residues for hydrolysis (Nejad *et al.* 2016).

Consolidated bio-processing (CBP) is another upgraded version of SSF where a single organism (wild or recombinant) performs enzyme production, saccharification and fermentation in the same bioreactor (Harahap *et al.* 2020, Choudhary *et al.* 2016). CBP requires the modification of cellulolytic strains to improve fermentation and that of fermentative strains to improve cellulolysis. Apart from these, basic qualities such as utilization of a range of substrates, high yield, reduced inhibitor formation, pH and thermostability are also essential (Gomes *et al.* 2021). Saccharification at increased temperatures has certain advantages such as enhanced LCB disintegration, enhanced solubility and therefore reduced medium viscosity, reduced contamination. Therefore a thermotolerant enzyme would additionally allow for swift reaction rates, lesser reaction times and reduced quantities of enzymes needed for saccharification (Han *et al.* 2020).

1.3.6. Role of thermophiles in saccharification

Thermophiles have previously been found in paper sludge, hot springs, biorefinery effluents, acid mine sediment, biogas processors, effluvium, volcano debris, oceanic vents, oil rigs, soil and compost (Mehta *et al.* 2016, Kushkevych *et al.* 2019, Piazza *et al.* 2020). They grow optimally and reproduce in thermal conditions but may survive at ambient temperatures. They are classified as moderately thermophilic (45-70 °C), extremely thermophilic (70-80 °C) and hyperthermophilic (80-125 °C) (Zeldes *et al.* 2015, Canganella and Wiegel 2011). These microorganisms usually display attributes of biopolymer degradation, and the production of biocatalysts functioning optimally at elevated temperatures (Piazza *et al.* 2020).

Thermophiles are a source of industrially applicable heat-stable enzymes (Pandey *et al.* 2015). The transition temperature (T_m) and half life ($t_{1/2}$) of thermophilic cellulases exceeds that of mesophilic cellulases due to which they can tolerate extremes of pH, temperature, organic solvents, lengthy processing periods and still be reusable. Thermophilic bacteria are a source of many efficient CAZymes (Bohme *et al*. 2020, Espliego *et al*. 2018, Margaryan *et al.* 2018). Apart from structural integrity and long half lives, thermophilic bacterial enzymes also possess low feedback inhibition and high specific activity due to lowered medium consistency, increased substrate diffusion and availability at elevated temperatures thereby giving higher reaction velocities and higher product recovery (Olguin-Maciel *et al.* 2020, Gomes *et al.* 2016). Thermophilic hydrolases are produced by the genera *Bacillus, Geobacillus, Anoxybacillus, Acidothermus, Thermotoga* and *Thermoanaerobacterium* (Bibra *et al.* 2018). A number of studies have proven the capability of thermophilic bacteria to hydrolyze a variety of untreated and pretreated lignocellulosic substrates (**Table 1.3**). These include untreated rice and wheat straws, SCB, poplar biomass, algae, beet vinasse, corn and cassava, alkali pretreated pine needles and oat spelt xylan, and acid-steam exploded sweet sorghum bagasse and groundnut shells. Apart from these, there were also some synthetic substrates such as untreated Avicel, xylose and cellobiose. Thermophilic bacteria were able to saccharify these substrates at a pH range of (4.5-9) and temperatures (50-70) ºC. Yet most commercially available cellulases are of fungal origin (Bhatt and Bhatt 2019). Since the enzymes of thermophilic fungi such as *Thermomyces duponti, Thermomyces lanuginosus, Thermoascus crustaceous, Aspergillus fumigatus* and *Rhizomucor pusillus* have low product yield, give better saccharification only when combined with other enzymes and are difficult to recycle, bioprospecting of thermophilic bacterial enzymes could be a suitable alternative (Xia *et al.* 2016, Rodrigues *et al.* 2015, Sandona *et al.*2019).

Thermophilic bacterial enzymes could be relevant due to their inherent ability to provide greater regulatory powers over industrial processes (Zeldes *et al.* 2015). The direct replacement of the current enzymes in biorefineries by thermophilic enzymes would be problematic as the entire operational set-up would need to be reorganized. Adjustments of the existing machinery towards elevated temperatures would be laborious and expensive and would most likely involve a complete makeover. This would prove tedious for the industrialists involved. Instead thermophilic enzymes can be implemented at two different stages of the process where high temperatures occur. One stage is post-pretreatment cooling of hot slurry. The time gap caused by this cooling step could be used for a pre-saccharification. However due to too much physicochemical variations, design of a specific enzyme mix for this step has not been possible. The other stage is during the saccharification itself. Further utilization of thermophilic fermentors for one pot CBP would lead to process simplification. Thermophilic enzymes can be directly added to the commercial enzyme mix to increase saccharification efficiency and for simultaneous product recovery (Bhalla *et al.* 2013, Arora *et al.* 2015, Donato *et al.* 2019, Zuliani *et al.* 2021). Also utilization of thermophilic enzymes in commercial fermentation could reduce cooling expenses for the fermentation tank (Zeldes *et al.* 2015).

Table 1.3. Thermophilic bacteria used in saccharification

Name	pH, Temperature	Pretreatment	Saccharified substrates	Reference
Bacillus licheniformis	6, 50°C	None	Wheat straw, RS, Sugarcane bagasse	Aftab *et al.* 2018
Bacillus cereus	9, 50°C	None	Poplar	Tabssum *et al.* 2018
Bacillus subtilis K-18	7, 50°C	3% NaOH	Pine needles	Irfan *et al.* 2017
Anoxybacillus flavithermus TWXYL3	7, 65°C	0.4% NaOH	Oat spelt xylan	Ellis and Magnuson 2012
Aeribacillus pallidus	8, 50°C	None	Beet vinasse	Harirchi *et al.* 2020
Aeribacillus pallidus BTPS-2	7, 70°C	None	Algal biomass	Timilsina *et al.* 2020
Brevibacillus sp MT5	4.8, 60°C	None	Sugarcane bagasse	Khan *et al.* 2020
Paenibacillus alvei	5, 50°C	8% NaOH	Date palm leaves	Mostafa *et al.* 2020
Geobacillus stearothermophilus RFS2	6, 50°C	0.9125% HCl+steam explosion	Groundnut shell	Sharma *et al.* 2013
Clostridium thermobutyricum	5, 50°C	0.5% phosphoric acid+steam explosion	Sweet sorghum bagasse	Wang *et al.* 2015b
Ruminiclostridium thermocellum M3	4.5, 55°C	None	Poplar sawdust, corncobs, RS, cornstalks	Sheng *et al.* 2016
Thermoanaerobacterium sp R63	5, 65°C	None	Cassava pulp	Harnvoravongchai *et al.* 2020
Caloramator boliviensis	6.5, 60°C	None	Xylose, cellobiose	Crespo *et al.* 2012
Caldicellulosiruptor bescii	7.8, 70°C	Milling	Timothy grass	Yokoyama *et al.* 2014
Caldicellulosiruptor	7, 68°C	None	Avicel	Mladenovska *et*

lactoaceticus 6A	*al.* 1995

Thermophiles have undergone numerous modifications in their cell membranes, genomes, transcriptomes and proteomes to adjust to thermal conditions.

1.3.6.1. Adaptations of thermophilic cell membrane

Cell membranes of thermophilic bacteria have higher percentages of saturated lipids such as *iso*-fatty acids (Koga 2012). Long chain lipids increase the frequency of van der Waals interactions in the cell membrane which reduce its porosity and prevent loss of ions. Sphingolipids and ornithine lipids present have their amino and fatty head groups in close proximity to each other which allows for hydrogen bonding and further stabilization (Sollich *et al.* 2017). In particular, branched chain fatty acids (BCFA) are important in maintaining the cell membrane in a semi-fluid state. Thermophilic bacteria have a greater *iso*-BCFA to *anteiso*-BCFA ratio. Membrane-spanning mono, di, tetraethers and tetraesters are formed by conjoining adjacent *iso*-BCFA and become more prevalent with an increase in growth temperature. Other cell membrane components that impart thermotolerance include glycolipids, *cis*-mono unsaturated fatty acids, cyclopropyl compounds, polar carotenoids, diabolic acid (Siliakus *et al.* 2017) amino pentane tetrol, N-acetylhexosamine (Sollich *et al.* 2017).

1.3.6.2. Genomic adaptations

Thermophilic DNA prefers the A-DNA structure as opposed to the usual B-DNA due to the presence of GC stretches (Paz *et al.* 2004). GC base pairs having an additional hydrogen linkage as compared to AT base pairs require more energy to be broken (Nakashima *et al.* 2013). There is an abundance of electrostatic interactions, supercoils and positively charged proteins interacting with the DNA thereby stabilizing it (Jegousse *et al.* 2017). Rapid nucleotide repair systems prevent the formation of 8-oxoguanine, deamination of cytosine and strand breaks, provide resistance to mutation and a selective advantage for growth under thermophilic conditions (Hu *et al.* 2022a).

1.3.6.3. Transcriptomic adaptations

Guanine and cytosine having higher tendency towards hydrogen bond formation, create double bonds within the RNA giving it a semi-folded configuration (Paz *et al.* 2004). Higher GC content is found in the double stranded regions of thermophilic RNA (Meyer 2021) and strengthens tRNA, rRNA and ncRNA whereas mRNA is strengthened by high purine (GA) content. 16S rRNA is more compactly folded at elevated temperatures (Jegousse *et al.* 2017). Also uracil content of 16S rRNA decreases under growth at elevated temperatures (Khachane *et al.* 2005). Modification of a uridine (U54) to 2-thioribothymidine raises the melting temperature of thermophilic tRNA by 3 °C by causing puckering of the C3’ endo sugar and reverse the Hoogsteen pairing with the adenine at position 58 (A58) (Lorenz *et al.* 2017).

1.3.6.4. Proteomic adaptations

Thermophiles contain a higher percentage of amino acids, such as alanine (Ala), Glu, proline (Pro), arginine (Arg) and lysine (Lys) and a reduced percentage of His, Asn, glutamine (Gln), threonine (Thr), Ser, Cys, Asp and Gly (Meruelo *et al.* 2012, Wang *et al.* 2015c, Gao *et al.* 2021). Promotion of intermolecular interactions, due to an abundance of Arg and Glu and replacement of valine (Val), Thr and iso-leucine (Ile) by Leu and Ala, leads to the stabilization of secondary structures such as α-helices. In addition, incorporation of Pro reduces the unfolding entropy. There is high amount of hydrophobicity concentrated in the protein interior, limited number of coils and pockets, a reduction in surface area to volume ratio leading to reduced configurational alterations (Yakimov *et al.* 2016, Escuder-Rodriguez *et al.* 2018). Fewer reactive heavy side chains with longer side chains fitted into the protein core maximizes their interactions with each other. Ion pairing in thermophilic proteins with <4 Å distance between the ions compared to (6-8) Å distance in mesophiles, and linking of internal amino acids with surface residues and side chains with the main amino acid framework by almost 49 % of the hydrogen linkages present, leads to a compact protein structure (Finch and Kim 2018). Apart from extensive ion pairing and hydrogen bonding, there are various other factors contributing to protein stability. There are van der Waals interactions, glycosylation of residues, salt bridge formations, small bond lengths, broader Donor-Hydrogen-Acceptor angles, presence of Tyr, Trp and Phe, cation-Π linkages between Arg, Lys, Trp and Tyr (Pandey *et al.* 2016, Lorenz *et al.* 2017, Ahmed *et al.* 2022). Disulfide bonds are a type of covalent bond that fastens Cys and prevent molecular flexibility (Saelensminde *et al.* 2009). They act dually as in decreasing the random unwinding of intact protein chains while preventing further unwounding of partially opened protein chains (Niu *et al.* 2016).

1.3.7. Lignocellulolytic consortium and co-culture

Natural lignocellulose degradation is performed collaboratively by a variety of microorganisms including fungi, anaerobes and actinobacteria, which are the native microorganisms from lignocellulosic environment (Yan *et al.* 2012). They take part in biopolymer deconstitution by producing an expanse of biocatalysts. Particularly, hydrolytic enzymes sourced from bacteria have been noted to be useful in this aspect (Kumar and Chandra 2020).

Compost-related samples can be a propitious source of enzymes which assist in lignocellulosic material reprocessing, mainly due to the large scale decomposition of organic refuse by the resident microbes (Wu *et al.* 2022a). Vermicomposting is the controlled mesophilic bio-oxidative degradation of organic waste (agricultural, sewage, industrial) by the joint action of earthworms and microbes into vermicompost (Saha *et al.* 2022). Earthworms influence soil properties through the enhancement of nitrogen assimilation, hydration, aeration, soil enrichment through the assemblage of organic and inorganic materials and the production of glycoproteins which activate local microbes. Earthworms particularly facilitate the growth of rapidly proliferating microbes, such as Flavobacterium, Actinobacteria, Firmicutes and γ-Proteobacteria and shape microbial communities (Medina-Sauza *et al.* 2019).

The composition of consortial microbial communities is influenced by pH, temperature, and nutritional factors (Hu *et al.* 2022b). The scarcity of easily assimilable nutrients, abundance of structurally complex substrates and the availability of terminal electron acceptors, emphasizes the growth of a few basic or keystone microorganisms. This determines their microbial growth and overall taxonomic diversity in favour of superior depolymerizers (Banerjee *et al.* 2016, Chukwuma *et al.* 2021). Further differentiation is brought about by enrichment culturing. Different LCB are depolymerized by different microorganisms and shape the microbial communities accordingly (Carlos *et al.* 2018).

However, there are certain disadvantages to consortia. Most bacteria are difficult to culture due to their specific needs. Some species grow only during scarcity of nutrients and die off in enriched media. Their differential growth rates indicate that they would have been eliminated from the consortium, unless helper species were present to synthesize growth stimulators (Vartoukian *et al.* 2010). Intermicrobial competition is difficult to do away with due to release of certain molecules into the culture media during microbial growth, differential bacterial growth rates and lack of control over microbial dynamics (Duncker *et al.* 2021). The main hindrance of consortia being used for saccharification of LCB is that post-production the released sugars would partially be transferred towards the metabolic processes of other microorganisms (Jimenez *et al.* 2015).

Synthetic consortia, being made of simpler microbial populations but with lignocellulolytic capability nearly similar to that of enrichment cultures, offer a higher degree of control over microbial dynamics and bioprocessing. Simpler combinations of LCB-degrading microorganisms can be obtained by using reductive screening which enables intermicrobial interactions to occur without alterations (Kang *et al.* 2020). There are a lesser number of genes involved in a co-culture system than in a consortium as it depends on the number of biochemical pathways involved (Goers *et al.* 2014). This makes it easier to utilize synthetic consortia or microbial co-cultures for the exploration of the foundational mechanisms, both interspecific (different species) and intraspecific (different strains of the same microorganism) interactions with respect to lignocellulose depolymerization (Rosero-Chasoy *et al.* 2021). Simultaneous saccharification and co-fermentation has been achieved earlier on steam exploded SCB, by co-cultivating a cellulolytic *Bacillus cereus* GBPS9 and a xylanolytic *Bacillus thuriengensis.* However co-cultures are not exclusively made from bacteria, they may also be made between fungi (Ire *et al.* 2016, Hernandez *et al.* 2018).

1.3.8. Bioprospecting for genes encoding cellulase

Common strategies for bioprospecting of cellulolytic genes include polymerase chain reaction (PCR) based amplification of the gene of interest through primer designing. Another approach is the cloning of fragments from a metagenomic library into a suitable host and functional screening of the recombinant colonies so produced (Pandey 2018). Purified genomic DNA is cloned into different vectors based on fragment size, plasmids (2-10 kb), fosmids and cosmids (20-40 kb), Bacterial Artificial Chromosomes (100-200 kb) (Li *et al.* 2009). Genes involved in related metabolic pathways are usually closely packed in 20-40 kb nucleotide segments which are preferably cloned into fosmids and cosmids to avoid genetic loss (Pabbathi *et al.* 2021). Proper selection of host strains is important to avoid drawbacks such as lack of post-translational modifications, incomplete introduction of inserts while cloning, irregular promoter binding, discrepancy in codon incorporation, improper RNA formation and protein folding leading to insufficient expression and errors during library screening (low screening sensitivity) (Li *et al.* 2009, Berini *et al.* 2017). Of all the environmental sample-derived bacterial genes discovered till date, only 7 % of DNA with high GC %, such as that of thermophiles, has been properly expressed in *E.coli*. Other hosts that have been considered for the enhanced expression of lignocellulolytic enzymes include *Bacillus* sp, *Agrobacterium* sp, *Rhodococcus* sp, *Pseudomonas putida* and *Streptomyces* sp (Berini *et al.* 2017).

An interest in genes encoding cellulases dates back to 1986 when a *cex* gene (1452 aa), encoding exoglucanases, was sequenced from *Cellulomonas fimi*. Its structure consisted of a 41 aa long signal peptide preceeding the initiation codon and a semi-overlapping tetra-repeat sequence which can self-modify into secondary structures, succeeding the stop codon (O'Neill *et al.* 1986). A large number of relevant bacterial genes that encode cellulases, originate from the families Bacillaceae and Clostridiaceae. In thermophilic bacteria, most cellulases are encoded by the *cel* gene (**Table 1.4**). This gene does not have a conserved structure. Degeneracy in base sequences enables glucanases from different CAZyme families with varied substrate specificities, to be encoded (Cornet *et al.* 1983, Beguin *et al.* 1985). Among the many variants of this gene (*celA, celB, celC, celK, celL), celA* and *celB* are the most prevalent and mostly encode endoglucanase. Although β-glucosidases are rarely encoded by *cel, cel* genes encoding them have been found across GH1 (*Gluc1C, bglp, unbgl1A, bglB, bgl3*). Despite the overwhelming presence of endoglucanases and β-glucosidases, few genes encoding exoglucanases have been reported to date (CelK, CpCel48, Cel9A) (**Table 1.4**) (Che and Men 2019).

The main difficulty in bioprospecting for cellulases is that only (0.1-1) % of consortial microbes can be isolated through standard laboratory procedures. Therefore, non-culture based methodologies, such as metagenomics, metatranscriptomics and metaproteomics offer great promise as to the bioprospecting of prokaryotes for industrially applicable hydrolases (Berini *et al.* 2017). Of these, metagenomics is the direct analysis of the total genomic DNA that has been extracted and subcloned from an ecological system. It has been earlier performed on samples from forest soils, wood lice gut, termite gut and bamboo rat gut (Damon *et al.* 2012, Bredon *et al.* 2018, Calusinska *et al.* 2020, Bai *et al.* 2021). Metagenomics can be subdivided into structural metagenomics or taxonomic profiling which seeks to identify foundational specimens involved in the stabilization of environmental communities and evolutionary processes, biodiversity, intercellular relationships, and functional profiling which evaluates the genes encoding lignocellulolytic proteins and their specific roles (characterization). Functional profiling of a consortium is a valuable tool for the bio-prospecting of industrially relevant biocatalysts (Chettri *et al.* 2020, Prayogo *et al.* 2020).

While metagenomics focuses on biodiversity and gene pools, metatranscriptomics focuses on gene expression patterns and their regulation in environmental microbial communities

(Mukherjee and Reddy 2019). Metatranscriptomics is defined as PCR-based RNA sequencing and it has earlier elucidated the microbial dynamics at various levels of wood ash-supplemented soil samples (Bang-Andreasen *et al.* 2020). It offers an additional advantage in that it identifies the contributing members of the consortium through the tracking of gene expression. The general outline of metagenomics and metatranscriptomics serially include next generation sequencing (NGS) of highly purified nucleic acid (DNA or RNA) for library generation, quality control (QC) where low quality or adapter segments are trimmed, grouping (contig binning) of generated reads into operational taxonomic units based on high similarity (>97 %) (*de novo*), direct alignment to available sequences from database (closed-reference) and *de novo* alignment of remaining unmatched sequences (open-reference) and comparison between codon-dependent and -independent functional profiling (Shakya *et al.* 2019, Li *et al.* 2018a). Contig binning is defined as the designation of nucleotide fragments to genomes of particular species based on specific signals such as conserved sequences that persist throughout the fragment. Binning of DNA strands into contigs is both laborious and rewarding as in recognition of completed open reading frames (ORFs) and higher resolution for closely regulated gene clusters such as operons. Contig binning reduces the error that could have been caused by short segments and its efficiency depends on read coverage, nucleotide arrangement, GC percentage and their degrees of similarity with related proteins (Chen *et al.* 2021).

A major hindrance to the efficiency of these two omic-based strategies is their dependency on the proportions of bacteria and eukaryotic cells and on pre-enrichment (addition of a cellulosic substrate to the growth medium). While the growth of specific degraders is encouraged, evaluation is retarded due to changes in the latter (Batista-Garcia *et al.* 2016). Although metagenomic and metatranscriptomic analysis has helped elucidate microbial gene expression patterns, it still does not reveal much about the biochemical mechanisms involved. The location, transport, synthetic patterns, structural and functional annotation, post-translational modifications and inter-protein interactions is given by metaproteomics (Talavera-Caro *et al.* 2020). It allows for the mass analysis of intracellular, secretomal and newly translated proteins from their respective mRNA. Proteins are separated by liquid chromatography based on differences in their molecular weight and polarity. However, newly translated proteins would not avail of post-translational modifications. Currently the preferred methodology for metaproteomics is Liquid Chromatography-Mass Spectrophotometry (LC-MS) and tandem mass spectrometry (MS-MS) of trypsinated, purified peptide or protein extract. Comparison of calculated and obtained mass spectra would identify the protein (Edwards *et al.* 2017, Heyer *et al.* 2019). The main limitations of metaproteomics include larger size of library as compared to metagenome or metatranscriptome leading to greater dependence on computers and bio-informatic analysis, misplaced protein identity due to similar peptide sequences in closely related proteins and dependence on existing databases which would not contain information on novel proteins (Heyer *et al.* 2017). A combination of metatranscriptomics with proteomic mass spectrophotometry has previously identified numerous cellulolytic and hemicellulolytic enzymes and CBM, along with a sizeable amount of unclassified proteins. Shotgun sequencing of a wheat straw-degrading microbial community gave a detailed report of microbe-environment interactions with a connection between biodiversity, co-metabolism and relevant metasecretomic factors (Alessi *et al*. 2018).

Genetic factors and their interregulation among the consortial members need more exploration. Full consortial potential has not yet been utilized due to the lack of information about the various synergistic interactions, growth regulators and metabolic mechanisms causing an unavailability of rational design (Che and Men 2019).

Table 1.4. Cellulase encoding genes of thermophilic bacteria

Gene	**Number of amino acids**	**Functional pH and temperature of encoded enzyme**	**GH family**	**Source organism**	**Reference**
(Exoglucanase)					
celK	895aa	6, 65°C	-	*Clostridium thermocellum*	Kataeva *et al.* 1999
CpCel48	919aa	6, 50°C	GH48	*Clostridium phytofermentans*	Zhang *et al.* 2010
cel9A	800aa	7, 60°C	GH9	*Cellulomonas bogoriensis* 69B4	Li *et al.* 2020
celMytB	631aa	6.5, 55°C	GH5	*Saccharophagus* sp Myt-1	Sakatoku *et al.* 2014
(Endoglucanase)					
celA	368aa	5, 65°C	GH5	*Geobacillus* sp 70PC53	Ng *et al.* 2009
celB	959aa	4, 80°C	GH51	*Alicyclobacillus acidocaldarius*	Eckert and Schneider 2003
celA	-	(5.5-6.5), 75°C	-	*Clostridium thermocellum*	Schwarz *et al.* 1986
cenC	-	6, 70°C	GH9	*Clostridium thermocellum*	ul Haq *et al.* 2015
cel10	310aa	5, 50°C	GH8	*Klebsiella pneumoniae*	Attigani *et al.* 2016
celA	-	7, 60°C	-	*Bacillus lautus* PL236	Hansen *et al.* 1992
E1	-	5.5, 75°C	GH5	*Acidothermus cellulolyticus*	Chung *et al.* 2015
AcCel12B	-	4.5, 75°C	GH12	*Acidothermus cellulolyticus* 11B	Wang *et al.* 2015a
cel5A	619aa	5, 70°C	GH5	*Cellulomonas bogoriensis* 69B4	Li *et al.* 2020
cel9A	800aa	7, 60°C	GH9	*Cellulomonas bogoriensis* 69B4	Li *et al.* 2020
celL	535aa11	5, 50°C	GH6	*Cellulosimicrobium funkei* HY-13	Kim *et al.* 2016
celC	368aa	6.5, 50°C	GH8	*Salmonella typhimurium* UR	Yoo *et al.* 2004
cel5A	-	4, 50°C	-	*Eubacterium cellulosolvens*	Park *et al.* 2018b

celA4	715aa	3.4, 60°C	-	*Alicyclobacillus* sp A4	Bai *et al.* 2010
(β-glucosidase)					
Gluc1C	448aa	6, 50°C	GH1	*Paenibacillus* sp MTCC 5639	Adlakha *et al.* 2012
bglp	-	7, 60°C	GH1	*Anoxybacillus flavithermus* subsp *yunnanenis* E13	Liu *et al.* 2017
celB	-	5, 75°C	-	*Pyrococcus furiosus*	Bohmer *et al.* 2012
unbgl1A	469aa	6, 50°C	GH1	Soil metagenome	Lu *et al.* 2013
bglA	453aa	5.5, 70°C	-	*Caldicellulosiruptor saccharolyticus* DSM 8903	Hong *et al.* 2009
bglB	473aa	6.2, 60°C	GH1	*Microbispora bispora*	Wright *et al.* 1992
bgl3	479aa	6.5, 50°C	GH1	*Streptomyces* sp QM-B814	Perez-Pons *et al.* 1994