UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

MILENA MOREIRA VACILOTTO

Biochemical and biophysical characterization of two glucuronoxylanases aiming the production of xylooligosaccharides

São Carlos
2024

MILENA MOREIRA VACILOTTO

Biochemical and biophysical characterization of two glucuronoxylanases aiming the production of xylooligosaccharides

Dissertation presented to the Graduate Program in Physics at the Instituto de Física de São Carlos, Universidade de São Paulo to obtain the degree of Master of Science.
Concentration area: Applied Physics
Option: Biomolecular Physics
Advisor: Prof. Dr. Igor Polikarpov

Original version

São Carlos
2024

Vacilotto, Milena Moreira
Biochemical and biophysical characterization of two glucuronoxylanases aiming the production of xylooligosaccharides / Milena Moreira Vacilotto; advisor Igor Polikarpov -- São Carlos 2024.
78 p.

Dissertation (Master's degree - Graduate Program in Biomolecular Physics) -- Instituto de Física de São Carlos, Universidade de São Paulo - Brasil , 2024.

1. Xylanases. 2. GH30. 3. Xylooligosaccharides. 4. Crystallographic structure. I. Polikarpov, Igor, advisor. II. Title.

FOLHA DE APROVAÇÃO

Milena Moreira Vacilotto

Dissertação apresentada ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Mestra em Ciências. Área de Concentração: Física Biomolecular.

Aprovado (a) em: 26/02/2024

Comissão Julgadora

Dr(a).: Igor Polikarpov

Instituição: (IFSC/USP)

Dr(a).: Mario de Oliveira Neto

Instituição: (UNESP/Botucatu)

Dr(a).: Valdeir Arantes

Instituição: (EEL/USP)

ACKNOWLEDGMENTS

The master degree is a big step in my life and career. Although it was a long process, with ups and downs, finishing this project proved to myself that I'm capable of achieving my own goals. Since so many people helped me in these processes, I feel happy to include these acknowledgments.

To my younger sister Michaeli, who stood by my side and believed in me even when I didn't. I know I can always count on you for everything. To all of the picture of the cats you took, all of the food you made me, and for buying the ticket for the Taylor Swifts show.

To my best friend Caio, for being on my side even in my darkest days. To every time you went to the lab with me because you knew I didn't want to be alone, and to all of the food and Disney movies we shared together.

To Vanessa, who taught me how to be a scientist, and have been accompanying me since the beginning of my scientific career. To every time you helped me see a light at the end of the tunnel.

To my older sister Milan, for going to the show with me, and for the amazing conversations we had.

To my roommate, Bianca, who lived with me for seven years with and never stopped cheering for my successes. Also, for every time I took your food because I didn't want to go to the supermarket.

To professor Igor Polikarpov, for giving me the opportunity to learn to my heart's content. Also, for the all the guidance and support.

To Evandro, who helped me with the proteins crystallographic structure, and was so patient throughout the process.

To my lab colleagues, Paula, Ana, Anelyze, Pedro, Raissa, Lívia, Angélica, Lorgio, Marcelo, Rafael, Paula, Maria who made my days better and helped me every time I needed.
To my cats, Panqueca and Linguiça, who gave me emotional support.

To my parents, Carlos and Helena, who always supported my passion for studies, even if that meant to send me away from them.

To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Finance code 88887.601517/2021-00 and Fundação de Apoio à Física e à Química (FAFQ) – Finance code 1698 for financially supporting me and my research.

To the Instituto de Física de São Carlos and the Universidade de São Paulo, for all the amazing team and infrastructure, that allowed me to pursuit my dream.

To everyone that was directly or indirectly involved in this project.

My most sincere thanks!

“Of all the years that we stood there on the sidelines

Wishing for right now”

Taylor Swift

ABSTRACT

VACILOTTO, M. V. **Biochemical and biophysical characterization of two glucuronoxylanases aiming the production of xylooligosaccharides.** 2024. 78 p. Dissertation (Master in Science) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2024.

Production of value-added compounds and sustainable materials from agro-industrial residues is essential for better waste management and building of circular economy, especially considering the anthropogenic effects in the global warming and natural resources depletion. This includes the valorization of the hemicellulosic fraction of plant biomass, aiming to produce prebiotic oligosaccharides, widely explored in food and feed industries. In the present work, we conducted biochemical and biophysical characterization of two prokaryotic xylanases of family 30_8 from *Bacillus pumilus* and *Ruminococcus champanellensis*, and assessed their applicability for xylooligosaccharides production using alkaline pretreated corn cob and eucalyptus sawdust collected from a local market and a sawmill shop in Araraquara, respectively. Mass spectrometry and high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) analysis revealed that *Rc*Xyn30A liberates mainly long monoglucuronylated xylooligosaccharides and proved to be highly inefficient in the cleavage of X4, X5 and X6, whereas *Bp*Xyn30A produces both linear and branched oligosaccharides. Crystallographic structure of *Bp*Xyn30A and *Rc*Xyn30A catalytic domain were solved and refined to 2.16 Å and 1.37Å resolution, respectively. Structural analysis of the enzymes binding cleft showed a conserved set of amino acids interacting with glucuronic acid substitution in the subsite -2b by several hydrogen bonds and ionic interactions, a characteristic shared between true glucunoxylanases. Furthermore, *Rc*Xyn30A has a larger β5-α5 loop as compared to other GH30 xylanases, which might be crucial for creating an additional aglycone subsite (+3). Finally, *B. pumilus* xylanase obtained higher conversion yields from pretreated biomasses than *Rc*Xyn30A, although the latter presents a specific activity against glucuronoxylan 9 times higher than the former.

Keywords: Xylanases. GH30. Xylooligosaccharides. Crystallographic structure.

RESUMO

VACILOTTO, M. V. **Caracterização bioquímica e biofísica de duas glucuronoxilanases visando a produção de xilooligossacarídeos.** 2024. 78 p. Dissertação (Mestre em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos 2024.

A produção de compostos com valor agregado e materiais sustentáveis a partir de resíduos agroindustriais é essencial para uma melhor gestão de resíduos e a manutenção de uma economia circular, especialmente considerando os efeitos antropogênicos no aquecimento global e no esgotamento dos recursos naturais. Isto inclui a valorização da fração hemicelulósica da biomassa vegetal, visando a produção de oligossacarídeos prebióticos amplamente explorados nas indústrias de alimentos e rações animais. No presente trabalho, realizamos a caracterização bioquímica e biofísica de duas xilanases procarióticas da família 30_8 de *Bacillus pumilus* e *Ruminococcus champanellensis*, e avaliamos suas aplicabilidades na produção de xylooligossacarídeos utilizando sabugo de milho e serragem de eucalipto com pré-tratado alcalino, e coletados de um mercado local e de uma serraria em Araraquara, respectivamente. Análises por espectrometria de massas e cromatografia de troca iônica de alta performance com detecção por amperometria pulsada (HPAEC-PAD) revelaram que a *Rc*Xyn30A libera principalmente xilooligossacarídeos longos e monoglucuronilados, e provou ser altamente ineficiente na clivagem de X4, X5 e X6, enquanto que a *Bp*Xyn30A produz tanto oligossacarídeos lineares quanto ramificados. As estruturas cristalográficas da *Bp*Xyn30A e do domínio catalítico da *Rc*Xyn30A foram resolvidas e refinadas a uma resolução de 2,16 Å e 1,37Å, respectivamente. Análise estrutural dos sítios de ligação das enzimas mostrou um conjunto conservado de aminoácidos interagindo com a substituição de ácido glucurônico no subsítio -2b por meio de diversas ligações de hidrogênio e interações iônicas, uma característica compartilhada entre as “verdadeiras” glucunoxilanases. Além disso, a *Rc*Xyn30A possui uma alça β5-α5 maior do que as demais xilanases da família GH30, o que pode ser crucial para a criação de um subsítio aglicônico adicional (+3). Finalmente, a xilanase de *B. pumilus* obteve maiores rendimentos a partir das biomassas pré-tratadas do que a *Rc*Xyn30A, embora esta última apresente uma atividade específica em glucuronoxilano 9 vezes maior que a primeira.

Palavras-chave: Xilanases. GH30. Xilooligossacarídeos. Estrutura cristalográfica.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS, ACRONYMS AND SYMBOLS

ABF	Acetate/borate/phosphate buffer
AXOS	Arabinoxylooligosaccharides
BpXyn30A	*Bacillus pumilus* ATCC 7061 glucuronoxylanase
BV-BRC	Bacterial and viral bioinformatics resource center
CAZy	Carbohydrate-active enzymes database
CAZymes	Carbohydrate-active enzymes
CBM	Carbohydrate-binding modules
CC-Alk	Alkaline pretreated corn cob
CC-IN	*In natura* corn cob
CD-Search	Conserved domain search
DHB	2,5-dihydroxybenzoic acid
DNS	3,5-dinitrosalicylic acid
DP	Degree of polymerization
DTT	Dithiothreitol
E-Alk	Alkaline pretreated eucalyptus sawdust
EC	Enzyme commission
EFSA	European food safety authority
E-IN	*In natura* eucalyptus sawdust
ENSO	El niño-southern oscillation
FAO	Food and Agriculture Organization
FDA	Food and Drug Administration
GDP	Gross domestic product
GH	Glycoside hydrolases
GlcA	α-D-glucuronic acid
H_2SO_4	Sulfuric acid
HPAEC	High-performance anion exchange chromatography
HPLC	High-performance liquid chromatography
IEA	International energy agency
IPCC	Intergovernmental panel on climate change
IPTG	Isopropyl β-D-thiogalactopyanoside
k_{cat}	Turnover number
k_{cat}/K_M	Catalytic efficiency

K_M	Michaelis constant
KOH	Potassium hydroxide
LB	Luria-Bertani medium
LIC	Ligation-independent cloning
m/z	Mass-to-charge ratio
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MeGlcA	4-O-methyl-D-glucuronic acid
$MeGlcA_2X3$	Xylotriose with a MeGlcA in the second xylose unit
$MeGlcAX_2$	Xylobiose with a MeGlcA
MS	Mass spectrometry
Na_2CO_3	Sodium carbonate
NaOH	Sodium hydroxide
NRE	Non-reducing end
PCR	Polymerase chain reaction
PDB	Protein data bank
PMSF	Phenylmethylsulfonyl fluoride
RcXyn30A	*Ruminococcus champanellensis* 18P13 glucuronoxylanase
RE	Reducing end
RMSD	Root mean square deviation
SCFA	Short-chain fatty acid
SN2	Single displacement mechanism with a nucleophile
$t_{1/2}$	Half-life
T4 Pol	T4 polymerase
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
Trx	Thioredoxin
UXOS	Glucuronoxylooligosaccharides
X3	Xylotriose
X4	Xylotetraose
X5	Xylopentaose
X6	Xylohexaose
XOS	Xylooligosaccharides
Xyl*p*	Xylopiranosyl unit
λ	Decay constant

CONTENTS

1 Introduction

1.1 Biomass as an energy and by-products source

The impact of humankind in the global temperature reached approximately 1 °C above pre-industrial levels in 2017 and, with the current warming rate, an increase of 0.5 °C to that value is estimated by 2040, according to a 2022 report by the Intergovernmental Panel on Climate Change (IPCC).[1] Unsurprisingly, the announcement of the El Niño-Southern Oscillation (ENSO) phenomenon in June of 2023[2] was accompanied with expectations of warming intensification due to anthropogenic effects and its massive consequences around the world.

Currently, the world market relies heavily on fossil fuels as energy source, known to produce a variety of pollutants that contribute to the global warming. The transition to more sustainable energy sources and materials represents an important element in combating greenhouse gas emissions and, consequently, the climate change.[3] One of the worldwide climate goal is the decarbonization of economy, that can be accomplished by the replacement of non-renewable fossil-based products by renewable bio-based ones, in the so-called circular bioeconomy approach.[4]

In Brazil, agribusiness is one of the major economic sectors, responsible for approximately 27% of its gross domestic product (GDP) per capita in 2020.[5] A report by the International Energy Agency (IEA) showed that 45.4% of the country's energy supply was made up of renewables in 2019 (Figure 1), and almost 75% of the bioenergy came from solid biomass.[6] The sugarcane sector, for instance, is predominant due to sugar and first generation ethanol production, and about 121 million tons of waste (sugarcane bagasse) were generated in 2016 by the country alone.[7]

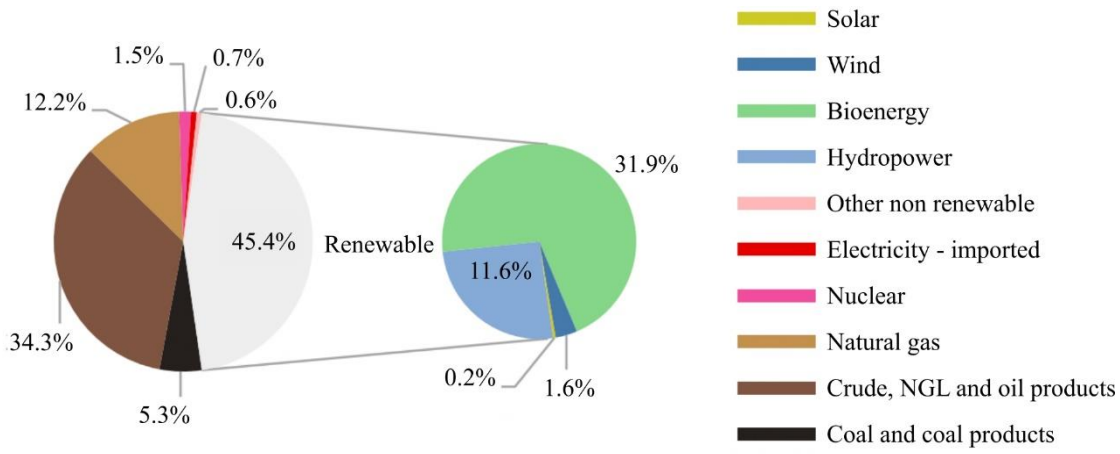

Figure 1 – Distribution of total energy supply in Brazil in 2019. The fraction in gray represents renewable sources, and percentages shown in the right graphic are relative to the total amount of energy.
Source: Adapted from COSTA.[6]

In these settings, the valorization approach of the biomass aims the maximization of energy output, which includes the use of sugar-rich feedstocks (sugarcane, corn, wheat) for the production of first-generation products, and conversion of their waste/residues into second-generation products. The upcycling of non-edible organic feedstocks (second generation biomass) to obtain value-added products, such as oligosaccharides[8], secondary metabolites[9] and biofuels[10], has become a topic of growing interest, because it promotes a sustainable waste management while aiming for energy efficiency.

1.1.1 Lignocellulosic biomass composition

Lignocellulosic biomass is defined as a non-edible (food or feed) plant or plant-based material characterized by its recalcitrant structure. It is mainly composed of polysaccharides, such as cellulose, hemicellulose, and pectin, as well as lignin, proteins, extractives and ashes, in proportions that greatly vary depending on the plant source, and all components associate with each other forming a complex structure.[11] Extractives consist of biopolymers without structural function that can be solubilized in water or neutral organic solvents (ethanol, cyclohexane), for example terpenoids, steroids, tannins and flavonoid compounds. The ashes fraction of the biomass is made up of minerals and other compounds, such as silicate, phosphate and magnesium, that produce solid residues after burning the organic material. Generally, wood/forestry biomass (sawdust, wood chips, etc) have low content of ashes compared to crop residues (corn cob, rice straw, etc).[12]

Cellulose is a homopolysaccharide of glucose residues connected by β-1,4 linkages with a degree of polymerization (DP) that can range from approximately 140 to more than 10000 units. Its individual linear chains arrange themselves in multiple layers that interact via hydrogen bonds, which creates a highly crystalline structure that confers the material its insolubility, stability and mechanical strength proprieties.[13] Pectin is a major component in primary cell walls of dicotyledons, but is only present in small amounts in monocotyledons and secondary cell walls of dicotyledons and, consequently, in the lignocellulosic biomass. It is composed of a water-soluble acid-sugar (ex. galacturonic acid) backbone chain linked by α-1,4 bonds that harbor neutral sugars as decorations. Pectin plays a role in cell adhesion and expansion, conferring flexibility to the cell wall.[14]

Hemicellulose is tightly bound to cellulose and, on the contrary of the latter, it is usually branched and contains pentoses (xylose, arabinose) and/or hexoses in its composition (mannose, galactose, glucose), besides compounds derived from hexoses, such as hexuronic acids (glucuronic acid, galacturonic acid) and deoxyhexoses (fucose, rhamnose), acetylation, methylation and ferulic acid esterification depending on the source. It provides stability and flexibility to the plant fibers.[13,15] In hardwoods (eucalyptus, beechwood, birchwood) and agricultural crops (corn cobs, sugarcane bagasse) xylan is the most abundant hemicellulose, and is composed of a backbone chain of β-1,4 linked xylose residues to which branching residues are often attached.[16]

In eucalyptus, for instance, the type of xylan present is the 4-O-methyl-D-gucuronoxylan or simply glucuronoxylan, that contains α-D-glucuronic acid (GlcA) or its methylated version (4-O-methyl-D-glucuronic acid; MeGlcA) connected to the xylopiranosyl (Xyl*p*) unit by α-1,2 linkages. Glucurono-arabinoxylans, are typically found in the grasses and cereals, and in addition to the GlcA and MeGlcA branches, the main chain also harbors arabinose residues attached to Xyl*p* in positions 2 and/or 3, which may contain esterification of phenolic acids such as ferulic acid and *p*-coumaric acid.[17]

Finally, lignin is an amorphous macromolecule made up of phenyl propane derivates (coniferyl, sinapyl and p-coumaryl alcohol), responsible for binding the cellulose and hemicellulose to the plant cell wall. It confers rigidness to the lignocellulosic matrix, enhances the plant cell wall hydrophobicity, and acts as a physical barrier that prevents access to the structural polysaccharides.[18] Lignin content tends to be observed in smaller quantities in grasses than hardwood (angiosperms), and softwood (gymnosperms) often presents the highest amounts of this compound.[12]

The same recalcitrant propriety that offers protection against plant degradation by microorganisms, insects and physical/chemical damage also arises as the greatest challenge in the cellulose and hemicellulose debranching and depolymerization for obtention of valuable resources from lignocellulosic biomass. Some strategies aim at the lignin removal from the lignocellulosic material in order to enhance the chemicals or enzymatic accessibility to their substrate. Since this polymer is only soluble in alkaline solutions, ionic liquids and specific solvents (organosolv pretreatment), those techniques are frequently explored for both forestry biomass and agricultural crop residues.[12]

In particular, the alkaline pretreatment allows the solubilization of the lignin and parts of the hemicellulose, especially for grass species such as corn cob. It also promotes disruption of biomass fibers, as it causes breakage of hydrogen that binds cellulose to hemicellulose and ester bonds, that binds hemicellulose to lignin. In addition, this pretreatment removes acetyl groups ramifications by esterification, which facilitates hemicellulose depolymerization by endo-acting enzymes. It is often carried out using chemicals such as sodium hydroxide (NaOH), sodium carbonate (Na_2CO_3) and potassium hydroxide (KOH).[17]

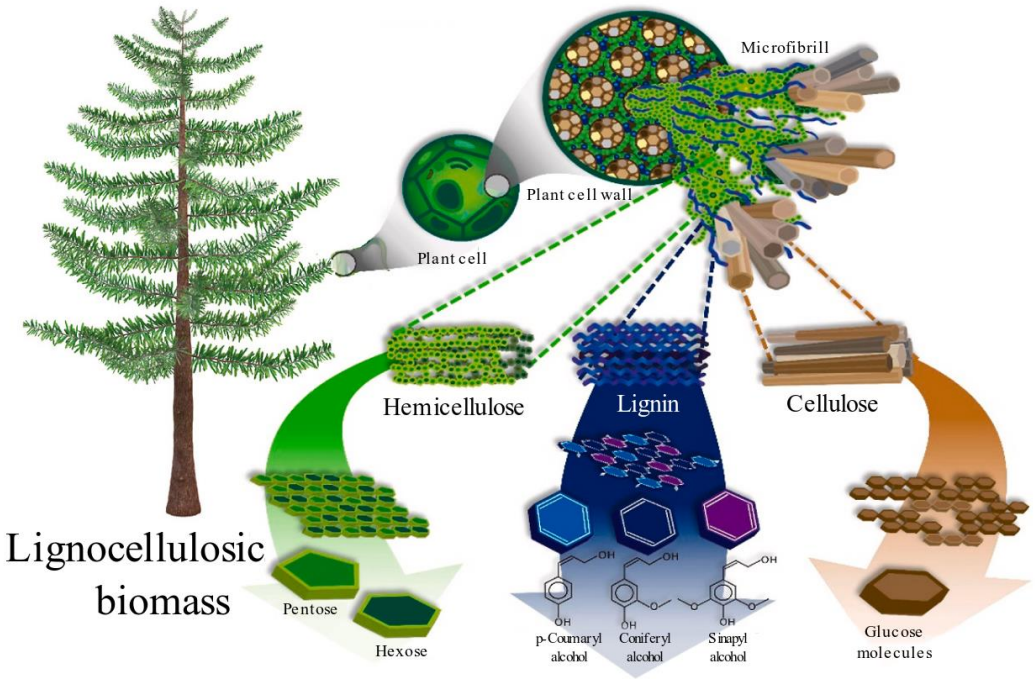

Figure 2 – Lignocellulosic biomass composition. Three of its main components include the polysaccharides cellulose and hemicellulose, and the amorphous polymer lignin.
Source: Adapted from MUJTABA *et al.*[11]

1.1.2 Enzymatic hydrolysis of the lignocellulosic biomass

The bioconversion of lignocellulosic biomass into products with aggregated value often relies on structural carbohydrates depolymerization by carbohydrate-active enzymes (CAZymes) due to a higher control of final products generation and its environmentally friendly approach. Glycoside hydrolases (GHs) are an expressive group among the CAZymes (http://www.cazy.org/Glycoside-Hydrolases.html), and they catalyze the hydrolysis of glycosidic bonds between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety.[19] They are classified in EC (Enzyme Commission) numbers based mainly on their substrate specificity and the chemical reaction they catalyze.[20]

GHs can be further categorized as endo or exo-enzymes, depending on whether the polysaccharide attack occurs somewhere within the polymer main chain or in one of its termini (reducing or non-reducing end). Substrate binding site topology reflects their regiospecificity: the former mode of action is characterized by an open cleft that allows random interactions with the polymer or tunnel-shaped site, common in processive enzymes, i.e., proteins that perform multiple rounds of catalysis before releasing the saccharide chain.[20]

In addition, glycoside hydrolases can either act via retaining or inverting mechanisms, i.e., whether the sugar is hydrolyzed in its equatorial or axial bond and if it retains or inverts its anomeric configuration.[20] Both modes of action are dictated by the spatial arrangement of catalytic residues (aspartate or glutamate) that operate as general acid/base or nucleophile. Retaining enzymes have catalytic groups separated by approximately 5 Å, which allows the formation of enzyme-substrate covalent intermediate, followed by double displacement mechanism with an intermediate (water, for most GHs). In contrast, inverting enzymes present a distance of approximately 10 Å between catalytic groups to accommodate a nucleophilic water between the sugar anomeric carbon and proteins catalytic base residue, and proceeds with a single displacement mechanism (SN2) (Figure 3).[19]

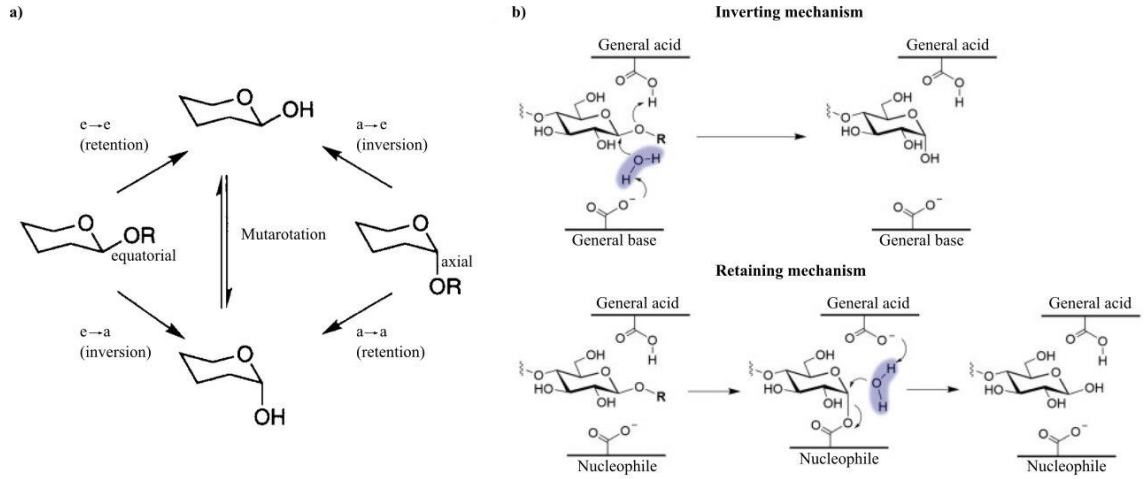

Figure 3 – Glycoside hydrolases retaining and inverting mechanisms. (a) Depiction of sugar hydrolysis in its equatorial and axial bonds. (b) Representation of reverting and retaining enzymes employing general acid/base residues and a general acid plus a nucleophile residue. Waters added during substrate cleavage are shaded in blue.
Source: (a) adapted from HENRISAT; DAVIES[20]; (b) adapted from PLANAS[19].

Glycoside hydrolases often display one or more domain structures without catalytic function called carbohydrate-binding modules (CBMs), that may attach to C- and/or N-terminus of the enzymes. CBMs are relatively small proteins, containing between 30 to 200 amino acids, and are classified within 100 families in CAZy database (http://www.cazy.org/Carbohydrate-Binding-Modules.html). Their characteristic of binding to substrate, especially insoluble ones, was shown to increase the proximity of enzyme to the polysaccharide, enhancing its catalytic activity.[21]

Enzymes capable of cleaving xylan are known as xylanases (EC 3.2.1.8), and are widely distributed among the GH families according to sequence, structure, molecular mechanism and substrate specificity.[22] More specifically, GH30_8 family comprises mainly prokaryotic endo-glucuronoxylanases (EC 3.2.1.136), characterized by their high specificity to xylans containing α-1,2 linked glucuronic acid (GlcA) or 4–O–methyl-D-glucuronic acid (MeGlcA) attached to the xylopiranosyl unit.[23]

1.1.3 Xylanolytic degrading bacteria

In nature, degradation of the plant cell wall is usually a cooperative process coordinated by members of a microbial community, that are typically found in environments that are rich in cellulosic materials, such as soil and rumen. For instance, bacteria play an important role in lignocellulosic biomass degradation in the rumen, but not all clades are equipped with sufficient enzymatic machinery suitable for this task. Some micro-organisms are capable of using long

and complexes polysaccharides as substrates, while others can only metabolize third-party degradation products.[24]

The human intestinal microbiota is also influenced by dietary glycans complexity, as shown elsewhere.[25] Rogowski and colleagues observed that only simple oligosaccharides liberated by *B. ovatus* growth on birchwood glucuronoxylan and wheat arabinoxylan can be used by *Bifidobacterium adolescentis*, whereas complex saccharides release from corn glucurono-arabinoxylan cannot be accessed by the bacteria.[25]

Members of *Bacteroidetes* and *Firmicutes* phyla tend to dominate the human gut, and they provide symbiotic benefits through the production of short-chain fatty acids (SCFAs) in the gastrointestinal tract of mammals by the fermentation of non-digestible carbohydrates; those molecules play a central metabolic role in the host's health.[26] Gram-positive *Firmicutes* might produce both free enzymes and cellulosomes during the plant cell wall digestion. Cellulosomes are a surface-associated system involved in the breakdown of complex substrates, like crystalline cellulose, through the aggregation of key enzymes for polysaccharides depolymerization (cellulases and hemicellulases). It is composed by a protein containing multiple cohesin domains called scaffoldin, onto which dockerin domains of enzymes attach by specific assembly.[24]

Both gram-positive *Bacillus pumilus* ATCC 7061 an *Ruminococcus champanellensis* 18P13 bacteria belong to *Firmicutes* phylum, and analysis of annotated genomes in Bacterial and Viral Bioinformatics Resource Center (BV-BRC)[27] reveled that whilst the former lacks cohesin domains, suggesting that it does not produce cellulosomes, the latter encodes for type I and II cohesin, responsible for formation of enzyme-cellulosome complex and the tethering the cellulosomes to the bacterial wall, respectively.[28]

Ruminococcus champanellensis is a gram-positive bacteria present in the human gut, and was first isolated from feces of a healthy individual.[30] The bacterium was shown to ferment oat spelt xylan and cellulose, but not its respective monosaccharides, appendant sugars (mannose, arabinose, galactose) or starch.[30] The other bacteria of the present study, *Bacillus pumilus*, is easily adaptable into different niches, allowing it to survive in the most diverse terrestrial and marine settings.[31]

In summary, we hypothesize that *R. champanellensis* glucuronoxylanase might enable the direct use of dietary glucuronoxylan by the bacteria through its association with cellulosomes, once *Rc*Xyn30A full sequence analysis demonstrates the presence of a dockerin domain[29], whereas *B. pumillus* xylanase likely act alone or in synergy with other enzymes during xylan degradation.

1.1.4 Emergent bio-based products from the hemicellulose

Among the emergent products and by-products from the hemicellulosic fraction of the biomass, xylitol, sorbitol, alcohol, furfural and oligosaccharides stand out.[16] Oligosaccharides have drawn special attention due to their versatile properties and wide range of applications in cosmetic, pharmaceutical and food and feed industries, specially related to their prebiotic properties, characterized by the promotion of beneficial bacteria growth in gut microbiota, such as *Bifidobacterium* sp. and *Lactobacillus* sp. [32-33]

Small xylooligosaccharides (mainly X2 and X3) produced via enzymatic hydrolysis of glucuronoxylan have been shown to act as carbon source for probiotic bacteria *B. adolescentis*, that consequently liberates SCFAs as acetic acid, propanoic acid and lactic acid.[34] It is yet to be determined whether the presence of ramifications in oligosaccharides strongly impacts the growth of classical probiotic bacteria, such as *Bifidobacterium* and *Lactobacillus*.

The global market of prebiotics is anticipated to reach USD 10.9 billion by 2027. Currently, xylooligosaccharides (XOS), products of xylan degradation, are classified as emerging prebiotics and are recognized as safe food by many organizations around the world (FDA, EFSA and FAO).[35] By definition, XOS are non-edible sugars made up of xylose residues with degree of polymerization (DP) from 2 to 10, which might harbor arabinose or MeGlcA decorations, resulting in arabinoxylooligosaccharides (AXOS) and glucuronoxylooligosaccharides (UXOS), respectively.[36-37]

Despite previous efforts to elucidate XOS bioactive properties, few advances have been made so far regarding the distinct properties of decorated oligosaccharides, perhaps because of the difficulty in obtaining and purifying these sugars.[37] Therefore, discovery of new enzymes with different substrate specificities is of a great importance for future studies on biotechnological applications aiming at production of xylooligosaccharides.

1.2 Physical techniques applied to structural biology

Structural biology is an area of study that employs a variety of physical techniques such as X-ray crystallography, nuclear magnetic resonance spectroscopy, mass spectrometry and small angle X-ray scattering to determine atomic level structure of macromolecules in order to assess their mechanistic function and unravel the cellular processes in which they are involved.[38] In the present work, we will briefly discuss the technics covered: mass spectrometry and X-ray crystallography. The former was applied in the characterization of the products

released by the xylanases, and the latter to obtain a 3D structure of the enzymes' catalytic domain.

1.2.1 Mass spectrometry

Mass spectrometry (MS) is an analytical technique that determines the mass-to-charge ratio (*m/z*) of ionized samples, i.e., the spectra are presented in terms of molecular mass in Daltons (Da) per unit of charge (78 g/mol would form an intact ion corresponding to 78 *m/z*). The main components of a MS equipment are an ion source, mass analyzer and detector. Regarding matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), samples are co-crystallized with a matrix capable of absorbing in the wavelength of the pulsed laser. Then, gas-phase ions are accelerated with a fixed potential into a vacuum tube until they reach a detector; since all the ions formed have the same kinetic energy, ions with higher *m/z* achieve lower velocities than their counterparts. The time taken to travel a fixed distance is used to determine the *m/z* ratio.[39]

Importantly, ions produced in MALDI in positive ion mode (analytes charged by protonation) are often in the form of adducts with H^+, K^+, Na^+, Li^+, or other cations. The matrix is crucial for oligosaccharides analysis, since carbohydrates present limited capacity to ionize due to absence of sites that can be protonated. 2,5-Dihydroxybenzoic acid (DHB) is an ionic liquid matrix commonly employed for carbohydrates detection, and it drives the ionization of analyte by ions transference. [40]

1.2.2 X-ray crystallography

Macromolecule crystallography is a technique that employs diffraction data collected from the passage of a beam of X-ray (such as synchrotron light) through a single crystal, aiming to determine the molecular structure.[41] Synchrotron radiation, for instance, is characterized by its high-brightness, which allows fast data collection and nanometer scale investigation, besides its broad spectrum (from infrared to X-ray).[42] Many steps are needed to obtain a crystallographic structure: target selection, cloning, expression and protein purification, crystallization, diffraction data collection, atoms positions determination, and structure analysis.[43]

Crystallization assays requires a pure concentrated protein close to its maximum solubility (usually between 5-15 mg/mL), and include an initial screening trial using

commercial crystallization kits that contain varying concentrations of different precipitant agents, salts and pH. The main objective is to achieve a supersaturation condition, so that the biomolecule can segregate from other molecules (precipitant agents such as salts, organic solvents and polymers), aggregate, nucleate (formation of a stable solid nucleus) and grow. Afterwards, promising conditions might be further screened to optimize crystal growth. The most used strategy for protein crystallization is the vapor diffusion, that consists in placing a sample drop containing a certain volume ratio of protein: precipitant solution in either a hanging or sitting position, in order for it to equilibrate with the mother-liquid placed in the reservoir.[43-44]

The crystalline structure resultant from the previous step will be made up of an arrangement of atoms in a regular and repetitive pattern; the smallest repeating component is called unit cell, and periodic arrangement of those unit cells is known as the crystal lattice. The former can be represented by its three lengths (a, b c) of the 3D structure and the angles between those axes (α, β, γ), whereas the latter is classified within the 14 Bravais lattices based on unit cell parameters and according to crystal symmetry, which results in the space group (represented as a set of symbols and numbers).[45]

Once the X-ray diffraction data was collected, one will encounter the "phases problem". Shortly, to reconstruct the electron density function from the diffraction pattern, both magnitude and phases of the X-ray must be known. However, while diffraction intensities are an output of the experiment, phases are always lost in the process. Although the phases can be obtained through direct methods, such as extensive calculations or data collection of crystals soaked with heavy-atoms (Hg, Pb, Pt, for example), when possible, the strategy known as molecular replacement is preferably used. It is based on the positioning of the atomic coordinates of the closest homologous with solved structure into the unknown unit cell to obtain the phases of the protein of interest. [46] With the development of programs with higher accurate *in silico* models' generation, as AlphaFold, molecular-replacement started to be performed using predicted structures.[47]

2 Objectives

2.1 Overall objectives

The recent demand for "green" products is related to the impacts of humankind on nature, mainly due to depletion of natural resources and high levels of atmospheric pollution that cause global warming. One of the approaches that has been explored a lot in recent years is the use of agro-industrial wastes for the production of value-added products, such as prebiotic oligosaccharides. In particular, xylooligosaccharides are considered emerging prebiotics, and although some studies have shown that linear XOS can be used as carbon source for probiotic bacteria, little progress has been made in order to determine whether decorations in the sugar backbone affects its prebiotic activity. Therefore, studies of enzymes that produce specific types of xylooligosaccharides, such as UXOS, are the first step towards the evaluation of those sugars' properties and possible biotechnological applications. In the present work, we aimed the biochemical and biophysical characterization of two novel GH30_8 xylanases from *Bacillus pumilus* (*Bp*Xyn30A) and *Ruminococcus champanellensis* (*Rc*Xyn30A) in order to assess their application in the production of xylooligosaccharides from two lignocellulosic biomasses collected from local stores. We opted for a milder pre-treatment of biomass using 1% (w/v) NaOH to disrupt plant cell fibers, and conducted initial enzymatic hydrolysis assays.

2.2 Specific objectives

1. Phylogenetic analysis of *Bp*Xyn30A and *Rc*Xyn30A;
2. Cloning, heterologous expression and purification of proteins;
3. Obtention of optimal pH and temperature of studied enzymes, evaluation of their activity against different commercial substrates (oligo and polysaccharides) and obtention of kinetic parameters with glucuronoxylan;
4. Analysis of xylanases thermal stability;
5. Eucalyptus sawdust alkaline pretreatment and chemical characterization;
6. Evaluation of xylanases activity on pretreated eucalyptus sawdust and corn cob;
7. Crystallization, data collection and structural analysis.

3 Material and Methods

3.1 Target selection and phylogenetic analysis

The amino acid sequences coding for CAZymes annotated in the NCBI's protein database (https://www.ncbi.nlm.nih.gov/protein) from microorganisms in the possession of the Molecular Biotechnology Group were submitted to the Conserved Domain Search (CD-Search)[29] and BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) for target selection for high-throughput cloning. Next, the genes were analyzed with the NCBI ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/) using the bacterial nucleotide option and SignalP 6.0[48] for signal peptide prediction. This step was accomplished by the researcher Dr. Marcelo Vizona Liberato (Campinas, São Paulo, Brazil). Two glucuronoxylanases, one from *Ruminococcus champanellensis* 18P13 = JCM 17042 (GenBank: BBEP01000003.1) and one from *Bacillus pumilus* ATCC 7061 (GenBank: NZ_ABRX01000001.1) were chosen for the present study.

In order to evaluate the phylogenetic relationship of our targets with other published glucuronoxylanases, we have built a phylogenetic tree with the amino acid sequences of *B. pumilus (Bp*Xyn30A) and *R. champanellensis* (*Rc*Xyn30A) enzymes and other 17 xylanases from GH30_8 family available on CAZy database (http://www.cazy.org), 14 of which were biochemically characterized and 6 had their crystallographic structures deposited in Protein Data Bank (PDB).[49] For this purpose, the GH domain was isolated from the sequences employing CD-Search.[29] Sequence alignment was performed using the ClustalW algorithm[50] in the Molecular Evolutionary Genetics Analysis software version 11 (MEGA11).[51] Then, the neighbor-joining method[52] was used to build the phylogenetic tree in MEGA11 and the Bootstrap test with 10000 replicates was employed to assess the tree reliability.

3.2 Gene cloning and transformation

The coding sequences of the glycosyl hydrolases from family 30_8 from *Bacillus pumilus* ATCC 7061 (WP_034619861.1) without the signal peptide (first 33 amino acids) and *Ruminococcus champanellensis* 18P13 = JCM 17042 (GenBank ID: WP_147645661.1) without the signal peptide (first 23 amino acids) and with a CBM9 (up to amino acid 603) were amplified from the bacterial genome by Polymerase Chain Reaction (PCR) as previously described[53], using the Ligation-Independent Cloning (LIC) technique, which does not require

the use of restriction enzymes (Figure 4). Primers were designed with regions (in bold) that anneal to pETTRXA-1a expression vector, as follows: *Bp*Xyn30A and *Rc*Xyn30A forward primers (5' – **CAGGGCGCCATG**GTCATGCCCGAATC – 3' and 5' – **CAGGGCGCCATG**GCAAGTGATGCGAATATTAATG – 3') and reverse primers (5' – **GACCCGACGCGGTTA**CACATCGATTGCGC – 3' and 5' – **GACCCGACGCGGTTA**GCGTTTGACCACAAAT – 3'), respectively. Briefly, 25 µL of reaction containing 25 ng genomic DNA, 0.5 µM of each primer, 0.2 mM dNTP Mix, 1 unit of Phusion polymerase and 1x Phusion polymerase buffer (both from New England Biolabs, Massachusetts, USA) were submitted to 3 cycles of: (1) 1x 98 °C for 1 min, (2) 30 x 98 °C for 10 s, 55 / 60 / 65 °C for 30 s, 72 °C for 45 s, (3) 1x 72 °C for 10 min.

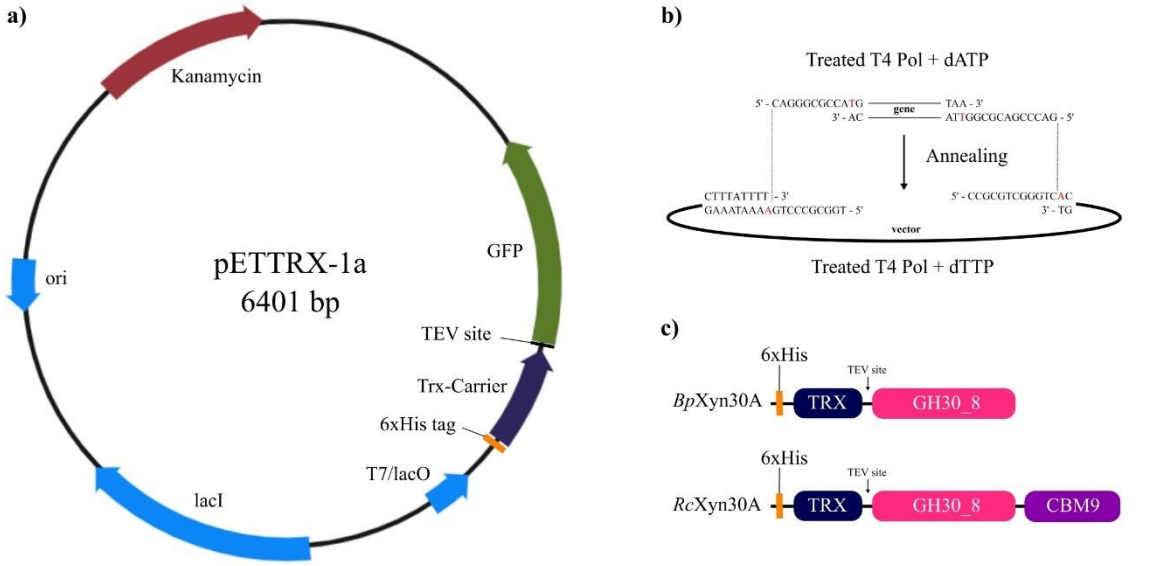

Figure 4 – (a) Features of pETTRX-1a expression vector. The vector has a resistance marker to kanamycin, a T7 promoter, a 6xHis tag to ease protein purification, a thioredoxin gene (Trx-Carrier) to enhance protein solubility, and a cleavage site for the Tobacco Etch Virus (TEV) protease. (b) Scheme of Ligation-Independent Cloning technique. Amplified genes and vector were treated with T4 polymerase (T4 Pol) and dATP and dTTP, respectively, to create single-stranded 5' overhangs. Nucleotide complementary to the dNTP added to the reaction is highlighted in red. (c) Modular structure of proteins involved in this study. Xylanases resultant from heterologous expression contain a 6xHis-Trx tag linked to their N-terminal, followed by a TEV cleavage site. In addition to a GH30_8 domain, *Rc*Xyn30A also contains a C-terminal CBM9.

Source: (a) By the author. (b) Adapted from CAMILO; POLIKARPOV[53].

PCR products were analyzed in 1% (w/v) agarose gel electrophoresis using 1x Safedye nucleic acid stain (Cellco, São Paulo, Brazil), followed by gel extraction with the Wizard® SV Gel and PCR Clean-Up System (Promega, Wisconsin, USA). In addition, the vector was linearized (primers forward: 5' – TGGCGCCCTGAAAATAAAG – 3' and primer reverse: 5' – CCGCGTCGGGTCAC– 3') using the same PCR steps described above, and reaction product

was treated with 20 units of *Dpn*I enzyme (New England Biolabs, Massachusetts, USA) overnight at 37 °C to remove template methylated DNA. DNA concentration was measured with the NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, USA) at 260 nm.

Subsequently, both vector and GH30_8 sequences were treated with T4 DNA polymerase (Fermentas, Massachusetts, USA) to form single-stranded 5' overhangs using the protocol described by CAMILO; POLIKARPOV[53]; shortly, 1.5 units T4 polymerase was incubated with 30 ng vector or 100 ng gene, 2.5 mM dTTP (for pETTRX-1a) or dATP (for the gene), 4 mM DTT and 1x T4 polymerase buffer at room temperature for 30 min, then inactivated at 75 °C for 20 min. The annealing of the T4-treated vector and insert was performed at room temperature for 10 min, in a proportion of 10:1 of treated insert: vector. 5 µL of the mixture was used to transform 50 µL *Escherichia coli* DH5α strain (Invitrogen, Massachusetts, USA) by thermal shock; the samples were incubated on ice for 20 min, submitted to heating at 42 °C for 2 min, cooled on ice for 2 min and kept at room temperature for 1 min. Then, bacteria were grown in 700 µL LB (Luria-Bertani) medium at 37 °C and 150 rpm for 90 min, and positive transformants were selected from colonies emerging in LB agar plates containing 50 µg/mL kanamycin.

Clones were further screened by colony PCR using the T7 Promoter Primer (Invitrogen, Massachusetts, USA) to confirm the presence of the insert in plasmids. The set up were as follows: single colonies were picked with a pipette tip under sterile conditions and swirled in a PCR 96-well plate containing 50 µL of a mixture of 0.1 µM of each T7 primer, 0.2 mM dNTP Mix, 1.25 units *Taq* DNA polymerase and 1x standard *Taq* reaction buffer (both from New England Biolabs, Massachusetts, USA) and submitted to 3 cycles of: (1) 1x 94 °C for 10 min, (2) 30x 94 °C for 30 s, 50 °C for 1 min, 72 °C for 2 min and 30 s, (3) 1x 72 °C for 5 min. Sequence analysis with the QIAGEN CLC Genomics Workbench software (version 22.0.1) showed that T7 amplified genes had an addition of 644 bp to the insert size.

Subsequently, plasmids were purified with Fast-n-Easy Plasmid Mini-Prep Kit (Cellco, São Paulo, Brazil), and 1 nM plasmid was transformed into competent *E. coli* Rosetta cells (DE3) (Invitrogen, Massachusetts, USA) in a similar procedure as described above, but employing kanamycin (50 µg/mL) and chloramphenicol (34 µg/mL) as resistance markers.

3.3 Protein heterologous expression and purification

E. coli Rosetta (DE3) transformed cells were grown in LB medium in the presence of 50 µg/mL kanamycin and 34 µg/mL chloramphenicol at 37 °C and 150 rpm. After the optical

density at 600 nm reached about 0.6, the protein expression was induced with 0.5 mM isopropyl β-D-thiogalactopyanoside (IPTG) for 16 h at 30 °C. The cell culture resultant from the previous step was centrifuged at 7878 g for 20 minutes at 4 °C and precipitated cells were gently resuspended in a lysis buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM dithiothreitol (DTT) and 1 mM phenylmethylsulfonyl fluoride (PMSF)). Next, the solution was incubated for 30 minutes with lysozyme (50 μg/mL) at 4 °C and submitted to ultrasound waves in six cycles of 30 s on and off with 40% of amplitude on ice using the F550 Sonic Dismembrator (Fisher Scientific, Hampton, USA), in order to lyse the cells.

Cell lysis products were centrifuged at 13000 g and 4 °C for 20 minutes and the supernatant was applied to a Ni-NTA Superflow resin (Qiagen, Hilden, Germany) equilibrated with buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl). The resin was washed with wash buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM imidazole) and GH30 xylanases were eluted using an imidazole gradient, consisting of four buffers containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl and imidazole pH 7.5 in concentrations 50 mM, 100 mM, 250 mM and 500 mM. The fractions containing the proteins were dialyzed to remove the traces of imidazole, and then incubated with 40 μg/mL of Tobacco Etch Virus (TEV) protease and 1 mM of DTT at 4 °C for 16 h to remove the 6xHis-thioredoxin tag. Subsequently, samples were submitted to a second purification step, consisting of an affinity chromatography on Ni-NTA resin, and cleaved protein TEV-free was eluted in 50 mM Tris-HCl pH 7.5 and 150 mM NaCl.

As a final purification step, *Rc*Xyn30A went through a gel filtration chromatography process at ÄKTA Purifier (GE Healthcare, Chicago, USA) using a HiLoad Superdex^{TM} 75 16/60 column (GE Healthcare, Chicago, USA) previously equilibrated with buffer (50 mM Tris-HCl pH 7.5 and 150 mM NaCl). Enzyme purity was evaluated with sodium dodecyl sulfate gel electrophoresis containing 15% polyacrylamide (SDS-PAGE 15%). Protein concentration was quantified using the NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, USA) at 280 nm using their theoretical mass (*Bp*Xyn30A = 44.4 kDa and *Rc*Xyn30A = 63.9 kDa) and molar extinction coefficient (*Bp*Xyn30A = 92.8 $M^{-1}.cm^{-1}$ and *Rc*Xyn30A = 78.2 $M^{-1}.cm^{-1}$).

3.4 Enzymatic assays

3,5-dinitrosalicylic acid (DNS) method[54] was used to perform enzymatic assays and quantification of total reducing sugars was carried out using a D-(+)-xylose standard. Experiments were conducted in triplicate with two controls (without enzyme), and except when

stated otherwise, prepared as follows: 0.5% (w/v) beechwood 4-O-methyl-glucuronoxylan, 180 nM *Bp*Xyn30A in 20 mM Tris-HCl buffer pH 7 or 15.6 nM *Rc*Xyn30A in 20 mM sodium phosphate buffer pH 6. Samples were incubated for 12 min at 55 °C or 7 min at 50 ºC for *Bp*Xyn30A or *Rc*Xyn30A, respectively, and reactions were stopped by adding DNS, followed by heating at 95 ºC for 5 min. The settings of enzyme concentration and time of reaction were obtained upon the fixation of either parameter and variation of the other one, and values were chosen in linear regions of products liberation.

Enzymatic activity was evaluated in 40 mM acetate/borate/phosphate (ABF) buffered solutions with pH ranges from 2 to 10 at a fixed temperature of 40 ºC. Subsequently, a temperature scan from 20 to 80 ºC was conducted under best pH conditions to assess the xylanases optimal temperatures. Next, the enzymes were preincubated at their optimal temperatures, 5 °C below that value and at 40 °C (for *Rc*Xyn30A) to evaluate their thermostability. Aliquots were removed from time to time and subjected to DNS assay. The graph of the residual activity as a function of time was linearized by applying the natural logarithm to the relative activity (%), and half-life ($t_{1/2}$) of xylanases was calculated as shown in equation 1, in which λ (decay constant) is the slope of the said curves.

$$t_{1/2} = \frac{\ln(2)}{\lambda} \quad (1)$$

The enzymes specific activity (U/mg) was obtained for substrates Avicel, carboxymethylcellulose (both from Sigma-Aldrich, St. Louis, USA), xyloglucan, arabinan, lichenan, rye arabinoxylan, β-glucan and beechwood 4-O-methyl-glucuronoxylan (all from Megazymes, Wicklow, Republic of Ireland) in order to know if *Bp*Xyn30A or *Rc*Xyn30A presented minor activities in addition to the endo-β-1,4-glucuronoxylanase. Since both enzymes showed activity only against glucuronoxylan, kinetic parameters were obtained by varying this substrate concentration from 0 to 12 g/L under the xylanases' optimum conditions and analyzed using the OriginLab software (version 8.0) with Michaelis Menten fitting.

3.5 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) technique[55-56] was used to investigate the effects of pH and salt on structural stability of the glucuronoxylanases. Experiments were carried out using 48 different citrate-glycine-phosphate buffers containing or not 150 mM of NaCl, and prepared as follows: 0.37 mg/mL enzyme, 50 mM buffered solutions with pHs from 1.2 to 10 and 1x diluted SYPRO Orange dye (5000x concentrated stock from Invitrogen,

Carlsbad, USA), in a final volume of 20 µL. An optical adhesive seal Microseal® 'B' was used to seal the 96-well thin-wall PCR plate (both from Bio-Rad, Hercules, USA). Reactions were conducted in a CFX96™ Real-Time PCR detection System (Bio-Rad, Hercules, USA), in which the temperature was varied from 25 and 90 °C with an increase of 1 °C every 30 s, and probe fluorescence intensity was measured at excitation/emission wavelengths of 490/530 nm. Melting temperatures were obtained through the derivate of curves using BioRad CFX Manager software.

3.6 Enzymatic cleavage pattern

*Bp*Xyn30A and *Rc*Xyn30A cleavage pattern was analyzed using 0.5% (w/v) beechwood 4-O-methyl-glucuronoxylan as substrate, and the necessity of glucuronic acid branches for enzymatic activity was evaluated using unbranched xylotetraose (X4), xylopentaose (X5), and xylohexaose (X6) (all from Megazymes, Wicklow, Republic of Ireland) at 0.05 mg/mL concentration. The samples were incubated with either *Bp*Xyn30A (180 nM for XOS and 1 µM for beechwood xylan) in 20 mM Tris-HCl buffer pH 7 or *Rc*Xyn30A (15.6 nM for XOS and 1 µM for beechwood xylan) in 20 sodium phosphate buffer pH 6 for 50 °C and 45 °C, respectively, and maintained at 1000 rpm for up to 24 h. Then, solutions were heated at 95 °C for 10 min, and reaction products were analyzed using the high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) system Dionex ICS-5000 equipped with the CarboPAC1 guard column (2 mm x 50 mm) and CarboPAC1 analytical column (2 mm x 250 mm) (all from Thermo Scientific, Waltham, USA). A mixture of xylose (Sigma-Aldrich, St. Louis, USA) and xylooligosaccharides with degree of polymerization from 2 (X2) to 6 (X6) (all from Megazymes, Wicklow, Republic of Ireland) were used as standards. As eluents, it was used 100 mM NaOH (buffer A) and 1 M sodium acetate with 100 mM NaOH (buffer B), and the running conditions were: 100% A for 5 min, 0-12% B for 15 min, 12-100% B for 5 min, 100-0% B for 2 min, 100% A for 8 min.

Soluble products released from beechwood 4-O-methyl-D-glucuronoxylan by 1 µM xylanases in 20 mM MES buffer pH 6 were further analyzed using matrix-assisted laser desorption/ionization with time-of-flight detection spectrometry (MALDI-TOF) using the Microflex LT MALDI-TOF (Bruker Daltonics, Massachusetts, EUA) instrument. Samples were prepared in triplicate in a proportion of 1:1 of reaction product and 2,5-Dihydroxybenzoic acid (DHB) (Sigma-Aldrich, St. Louis, USA) matrix at 20 mg/mL stock diluted in TA30 solvent (30:70 (v/v) Acetonitrile:TFA 0.1% in water). 1 µL of the mixture was applied to each spot of

the MSP 96 target polished steel BC (Bruker Daltonics, Massachusetts, EUA) and dried at room temperature. Spectra were acquired in an average of 6 laser shots within a range of 500 up to 2500 *m/z* using the linear positive-ion reflector mode and analyzed with flexAnalysis™ (Bruker Daltonics, Massachusetts, EUA) and OriginLab (version 2020) software.

3.7 Biomass pretreatment, characterization and enzymatic hydrolysis

Corn cobs (CC) used for assays were obtained from a local market, cut and milled with a knife mill to an approximate mash of 20, whereas eucalyptus sawdust was collected in a sawmill shop in Araraquara (São Paulo, Brazil). Biomass humidity was measured with a Moisture balance MOC - 120H (Shimadzu, Kyoto, Japan), and stored in plastic bags until use. The alkaline pretreatment was carried out with 1% (w/v) NaOH in water and 10% (w/v) biomass (dry weight) in an autoclave for 40 min at 121 °C. The solid was separated from the liquid by vacuum filtration, washed in running water until its pH reached close to 7, then again in deionized water, and left to dry in the incubator at 50 °C.

In natura and pretreated eucalyptus (E-IN and E-Alk, respectively) chemical characterization was conducted in triplicates as stated elsewhere[57], whereas CC-IN and CC-Alk compositions were determined in collaboration with the master's student Caio Cesar de Mello Capetti.[57] Briefly, extractives of untreated biomass kept in paper envelopes were removed with one cycle of 8-hour refluxing of 1:1 cyclohexane: ethanol using a Soxhlet extractor, followed by washing with deionized water in 7 cycles of 8-hour each. The envelopes were dried, and the mass difference was attributed to the proportion of extractives.

Soluble lignin and structural carbohydrates of E-IN and E-Alk were solubilized in a treatment with 72% H_2SO_4 for 7 min at 45 °C under constant stirring, in a proportion of 2 g biomass (dry weight) to 15 mL sulfuric acid, followed by incubation in an autoclave for 30 min at 121 °C with diluted acid (4% H_2SO_4).[57] The soluble fraction was analyzed by high-performance liquid chromatography (HPLC) system coupled with the column Aminex HPX-87H (300 x 7.8 mm) or Aminex HPX-87P (300 x 7.8 mm) (both from Bio-Rad, California, USA), in order to quantify acetic acid and monosaccharides, respectively, using calibration curves containing glucose, xylose, arabinose and acetic acid. The former column eluent was 5 mM sulfuric acid, while the latter was deionized water; both run conditions were 0.6 mL/min isocratic flow for 60 min. In addition, a mixture of 5% of the hydrolysate and 2% NaOH 6.5 M was read in a quartz cuvette at 280 nm in a spectrophotometer for obtention of the soluble lignin fraction.

The insoluble fraction was filtered using quantitative ashless filter paper (Whatman, Kent, United Kingdom) and washed with 1 L deionized water. The filter with insolubles was dried for 2 h at 105 °C and insoluble lignin was quantified through the weight of the sample. Finally, ashes content was evaluated by the mass difference of untreated biomass and the one burned in a muffle furnace (1 h at 200 °C, 1 h at 400 °C and 2 h at 800 °C).

Enzymatic hydrolysis of alkaline pretreated eucalyptus sawdust and corn cob was performed using 5% (w/v) biomass (dry weight) and 0.1 mg protein in 20 mM sodium phosphate buffer pH 6. Samples were prepared in triplicate with control reactions (absence of enzyme), and maintained at 40 °C and 1000 rpm in a Thermomixer C (Eppendorf, Hamburg, Germany). Aliquots were removed after 6 and 24 h, and reactions were stopped by heating at 95 °C for 10 min. Then, samples were centrifuged at 17000 g for 5 min and the supernatants were filtered with 0.22 μm CHROMAFIL Xtra PTFE-20/25 syringe filter (Macherey-nagel, Düren, Germany) prior to injection in HPAEC-PAD; oligosaccharides analysis was conducted as described in section 3.6. In addition, reaction products were quantified in terms of xylose equivalent using the DNS method.

3.8 Crystallographic structure determination and analysis

3.8.1 *Rc*Xyn30A partial digestion with papain and xylanases crystallization

Prior to crystallization assays, *Rc*Xyn30A full (GH + CBM9) was evaluated in the presence of papain from papaya latex (Sigma-Aldrich, St. Louis, USA) in 10, 20 and 30 min reactions at 25 °C containing proportions of 5/1, 25/1 and 50/1 of protein (mg/mL) / papain (mg/mL), 40 mM sodium phosphate buffer pH 6, 5 mM L-cysteine and 2 mM EDTA.[58] The linker between *Rc*Xyn30A catalytic domain and CBM9 was cleaved in 20 min reaction using the proportion of 25/1. After the incubation time, solution was applied to a HiLoad Superdex 75^{TM} 16/60 column equilibrated with 20 mM Tris-HCl pH 7.5 and 75 mM NaCl, and concentrated to 10 mg/mL stock. *Bp*Xyn30A contaminant free was prepared under the same buffer and concentration conditions as stated above.

Both catalytic domain crystals were obtained testing different commercial crystallization screen kits using the sitting-drop vapor-diffusion technique in an equal volume ratio of enzyme and mother liquor, and maintained in the Rock Imager 1000 (Formulatrix, Bedford, MA, USA) equipment at 20 °C. *Rc*Xyn30A crystals were detected under the crystallization condition 37.5% (v/v) precipitants agents (25% (v/v) MPD, 25% (w/v) PEG

1000 and 25% (w/v) PEG 3350), 0.1 M buffer solution (Tris-base / bicine) pH 8.5 and 0.09 M halogens, whereas for *Bp*Xyn30A crystals were observed under solution containing 20% (w/v) PEG 3350 and 0.18 M ammonium citrate. For the latter, the needle-shaped crystals formation was optimized using the hanging-drop method in 24 well crystallization plate by varying the precipitant agent from 15 to 27.5% and salt from 50 to 300 mM; the best condition proved to be 25% (w/v) PEG 3350 and 0.2 M ammonium citrate. Crystals were flash-frozen in liquid nitrogen using CryoLoops and stored in pucks for posterior data collection.

3.8.2 Data collection and structures determination

X-ray diffraction data were acquired under cryogenic conditions at the MANACA beamline of the Brazilian Synchrotron Light Laboratory (LNLS-Sirius, Campinas, Brazil) [42] using Pilatus 2M detector (Dectris, Baden, Switzerland). Data were indexed, integrated and scaled with XDSGUI software.[59] Xylanases structure was solved by molecular replacement in Phaser software[60] by using the phases from the one predicted using the AlphaFold algorithm.[61] Model structure building was performed with the Autobuild program[62] and refinement was carried out using PHENIX-refine[63], with manual modifications and addition of missing residues conducted in Coot.[64] Furthermore, protein structure quality was validated with MolProbity web service[65] and atomic coordinates were deposited in PDB. This stage of the work was carried out in partnership with the LNLS researcher Dr. Evandro Ares de Araujo (Campinas, São Paulo).

3.8.3 Structure and sequence analysis

*Bp*Xyn30A and *Rc*Xyn30A amino acid sequences were compared with experimentally determined structures of prokaryotic GH30 xylanases named: *Bs*Xyn30A (PDB entry 3KL0), *Ca*Xyn30A (PDB entry 5CXP), *Dc*Xyn30A (PDB entry 1NOF and 2Y24), *Pb*Xyn30A (PDB entry 4QAW), *Rp*Xyn30A (PDB entry 4FMV) and *At*Xyn30A (PDB entry 4CKQ). Sequence alignment was performed as described in section 3.1, and submitted to ESPript 3.0 [66] for structure-based sequence analysis using *Bp*Xyn30A as a query.

Interactions between the glucuronoxylanases active site and ligands were analyzed by superposition with *Dc*Xyn30A structure in complex with $MeGlcA_2X3$ (PDB entry 2Y24) and *At*Xyn30A structure in complex with xylobiose and xylotriose (PDB entry 5A6L), that share approximately 40% and 54% of identity to *R. champanellensis* xylanase, respectively, and

approximately 40% and 69% of identity to *B. pumilus* xylanase. GlcA decorated X6 was constructed with GLYCAM-Web (https://glycam.org/cb/) and superimposed with *Bp*Xyn30 and *Rc*Xyn30A structure in order to access the permissibility to decorations. Structural figures were generated with PyMOL Molecular Graphics System (Version 2.5.2 Schrödinger, LLC, New York, USA).

4 Results and discussion

4.1 Phylogenetic analysis

In order to infer the enzymes' functionality, we studied their evolutionary relationship with other prokaryotic GH30_30 sequences deposited in CAZy database. The phylogenetic tree subdivides itself into three major branches (Figure 5), and *Bp*Xyn30A belongs to the first branch, that contains mainly *Bacillus* sp. and *Paenibacillus* sp. enzymes. All biochemically characterized xylanases from this branch presented "true" GH30_8 activity, i.e., they were only active against glucuronic acid substituted xylan. The second and most heterogeneous branch harbors our *Rc*Xyn30A and contains enzymes faithful to their GH classification as well as enzymes with a promiscuous activity towards arabinoxylan. The third and final branch contains only *Acetivibrio thermocellus* ATCC 27405 xylanase, most likely due to its activity in a broad range of hemicelluloses (Table 1).

B. pumilus xylanase, like the rest of its clade, presented activity restricted to glucuronoxylan (as discussed in more detail in section 4.3). *Ruminococcus albus* xylanase, that is the closest related to *Rc*Xyn30A, has not been studied yet, and, therefore, we looked at the enzymes in the lower branches of this clade. *Ruminiclostridium papyrosolvens* DSM 2782 and *Clostridium acetobutylicum* ATCC 824 xylanases were described as promiscuous due to their activity on arabinoxylan, whereas the other enzymes exhibited activity only against glucuronoxylan (Table 1), similar to the one described in the present work. However, it is important to emphasize that since only a small number of proteins have been characterized, bootstrap values were too low to support strict evolutionary relationships, indicating that new GH30_8 xylanases must be studied to fulfill that gap.

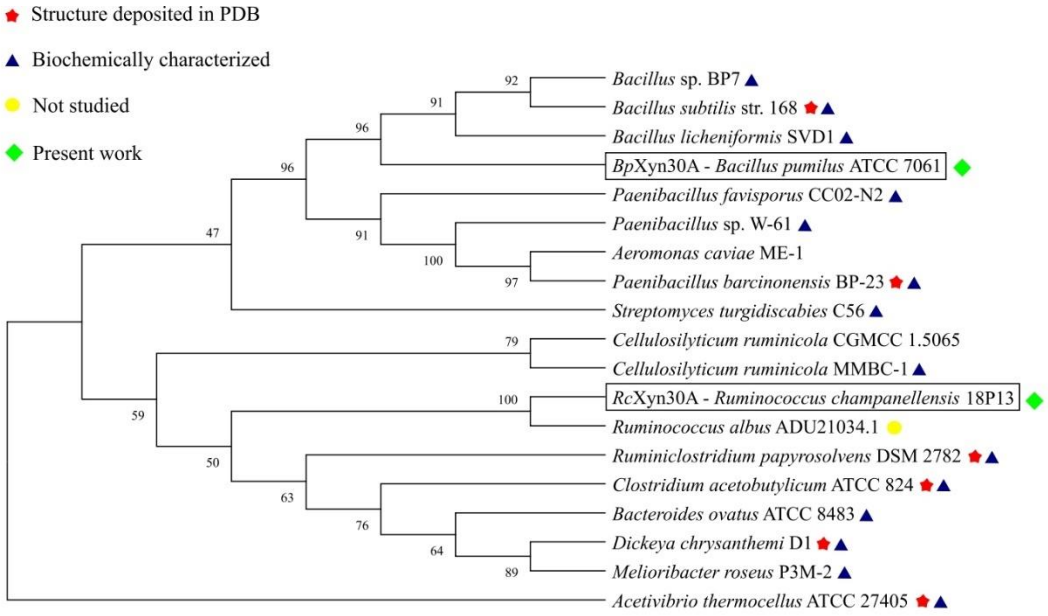

Figure 5 – Evolutionary relationship of GH30_8 xylanases. The study was performed using only the GH domain sequences of the proteins. Enzymes are represented by their organism of origin and information on the sequences used for this study can be found in Table 1. Red stars indicate xylanases with structures deposited in PDB; blue triangles indicate biochemically characterized proteins; yellow circle indicates a protein that has not been published yet; green diamonds indicate the target proteins of the present study.

Source: By the author.

Table 1 – Information regarding the xylanases used for the sequence alignment and construction of the phylogenetic tree. A survey of the substrates in which the enzymes were tested and presented a detectable activity was carried out, and is present below with their corresponding reference. Furthermore, xylanases' names were chosen based on their organism of origin and year of publication (suffix A and B). NP: not published.

Organism / Name	**Other domains**	**Genbank ID**	**PDB**	**Substrate with detected activity**	**Reference**
Bacillus pumilus ATCC 7061 / *Bp*Xyn30A	-	WP_034619861.1	NP	Glucuronoxylan, XOS (DP ≥ 3)	Present work
Ruminococcus champanellensis JCM 17042 / *Rc*Xyn30A	CBM9	WP_147645661.1	8VG9	Glucuronoxylan	Present work
Aeromonas caviae ME-1 / *Ac*Xyn30A	CBM35	AAB63573.1	-	NP	[67]
Bacillus licheniformis SVD1 / *Bl*Xyn30A	-	BAL45490.1	-	Glucuronoxylan, XOS (DP ≥ 3)	[68]
Bacillus sp. BP7 / *Bs*Xyn30B	-	ADM15019.1	-	Glucuronoxylan, XOS (DP ≥ 3)	[69]
Bacillus subtilis str. 168 / *Bs*Xyn30A	-	CAA97612.1	3KL0	Glucuronoxylan, XOS (DP ≥ 5)	[70,71]

(continued)

(continuation)

Bacteroides ovatus ATCC 8483 / *Ba*Xyn30A	BACON	ALJ48333.1	-	Glucuronoxylan, not tested for XOS	[25]
Cellulosilyticum ruminicola CGMCC 1.5065 / *Cr*Xyn30A	Ricin CBM2	ACZ98597.1	-	NP	[72]
Cellulosilyticum ruminicola MMBC-1 / *Cr*Xyn30B	CBM13 CBM2	UZE89494.1	-	Glucuronoxylan	[73]
Clostridium acetobutylicum ATCC 824 / *Ca*Xyn30A	Ricin	AAK76864.1	5CXP	Glucuronoxylan, arabinoxylan, XOS (DP ≥ 4)	[74]
Dickeya chrysanthemi D1 / *Dc*Xyn30A	-	AAB53151.1	1NOF	Glucuronoxylan, XOS (DP ≥ 5)	[75,76]
Melioribacter roseus P3M-2 / *Mr*Xyn30A	CBM9	AFN75320.1	-	Glucuronoxylan	[77]
Paenibacillus barcinonensis BP-23 / *Pb*Xyn30A	CBM35	AEY82463.1	4QAW	Glucuronoxylan	[78]
Paenibacillus favisporus CC02-N2 / *Pf*Xyn30A	-	AHA38215.1	-	Glucuronoxylan	[79]
Paenibacillus sp. W-61 / *Ps*Xyn30A	CBM6	BAA13641.1	-	Glucuronoxylan	[80]
Ruminiclostridium papyrosolvens DSM 2782 / *Rp*Xyn30A	CBM6 Dockerin I	EGD48159.1	4FMV	Glucuronoxylan, arabinoxylan, XOS (DP ≥ 4)	[81]
Streptomyces turgidiscabies C56 / *St*Xyn30A	-	WP_006381841.1	-	Glucuronoxylan	[82]
*Acetivibrio thermocellus ATCC 27405 / At*Xyn30A	CBM6 Dockerin I	ABN54208.1	4CKQ	Glucuronoxylan, arabinoxylan, arabinan, xyloglucan, glucomannan	[83,84]
Ruminococcus albus	CBM9	ADU21034.1	-	NP	NP

Source: By the author.

4.2 Cloning, heterologous expression and purification of *Bp*Xyn30A and *Rc*Xyn30A

B. pumilus and *R. champanellensis* xylanases were successfully cloned into pETTRX-1a/LIC expressing vector, and the resulting proteins were expressed in the soluble fraction in their fused form containing a 6xHis-thioredoxin tag, with approximated molecular masses of 58.7 and 78.2 kDa, respectively (Figure 6). Three purification steps were required to obtain *Rc*Xyn30A in its nearly pure form, including two nickel affinity chromatography and a molecular size exclusion chromatography, whereas *Bp*Xyn30A presented as a single band after two purification steps. After removal of the affinity tag, both xylanases migrated in the SDS-PAGE gel with molecular mass compatible to that deduced from their amino acid sequence with ProtParam[85] (44.4 kDa for *Bp*Xyn30A and 63.9 kDa for *Rc*Xyn30A).

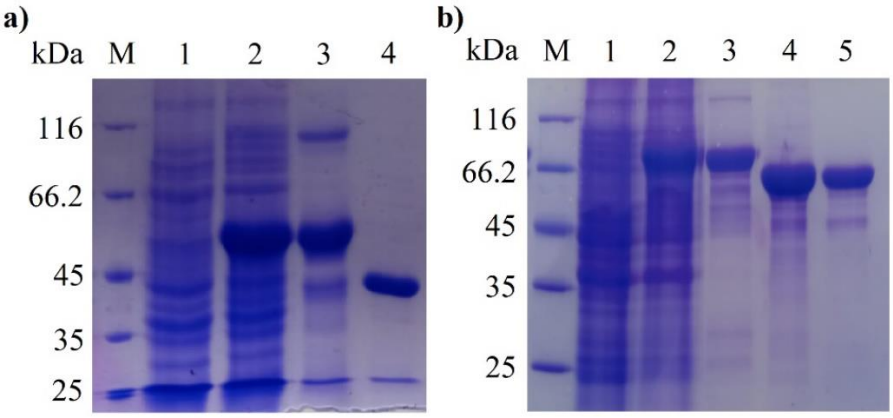

Figure 6 – SDS-PAGE gels of a) *Bp*Xyn30A and b) *Rc*Xyn30A expression and purification steps. M: low molecular weight marker; 1: cell culture before induction with IPGT; 2: cell culture after 16 h of protein expression; 3: *Bp*Xyn30A (58.7 kDa) *Rc*Xyn30A (78.2 kDa) after the first purification using nickel affinity chromatography; 4: non-fused xylanases (44.4 kDa and 63.9 kDa for *Bp*Xyn30A and *Rc*Xyn30A, respectively) after second purification using Ni^{2+} affinity chromatography; 5: *Rc*Xyn30A after the final purification step consisting of a size exclusion chromatography.
Source: By the author.

4.3 Enzymatic assays

4.3.1 Analysis of optimal conditions for enzymatic activity and thermostability

Biochemical analyses were carried out to identify the optimal conditions for enzymatic assays. *Bp*Xyn30 exhibited higher activity in pH 7 (Figure 7a), but retained almost 55, 90 and 84% of its activity in pHs 5, 6 and 8, respectively. As for the optimum temperature, this xylanase has its highest enzymatic activity between 50 and 60 °C (Figure 7b). On the other hand, *Rc*Xyn30A showed the best performance in pH 6 and at temperatures from 50 to 55 °C (Figures 7d and 7e), although it maintained above 60% of its activity in pHs 5 and 7, and

approximately 40% of activity in pH 8. These results are in accordance with the DSF assays (Figures 7c and 7f), since *B. pumilus* and *R. champanellensis* xylanases presented the highest melting temperatures (from 60 to 64 °C for the former and from 50 to 55 °C for the latter) for pHs between 5 and 8. According to the BRENDA database (https://www.brenda-enzymes.org/index.php)[86], prokaryotic GH30 xylanases described in literature present a broad range of optimum reaction conditions, that varies from pHs 5-10 and 30-70 °C, but most of them work in milder conditions, such as pH 6 and 40 °C (Table 2).

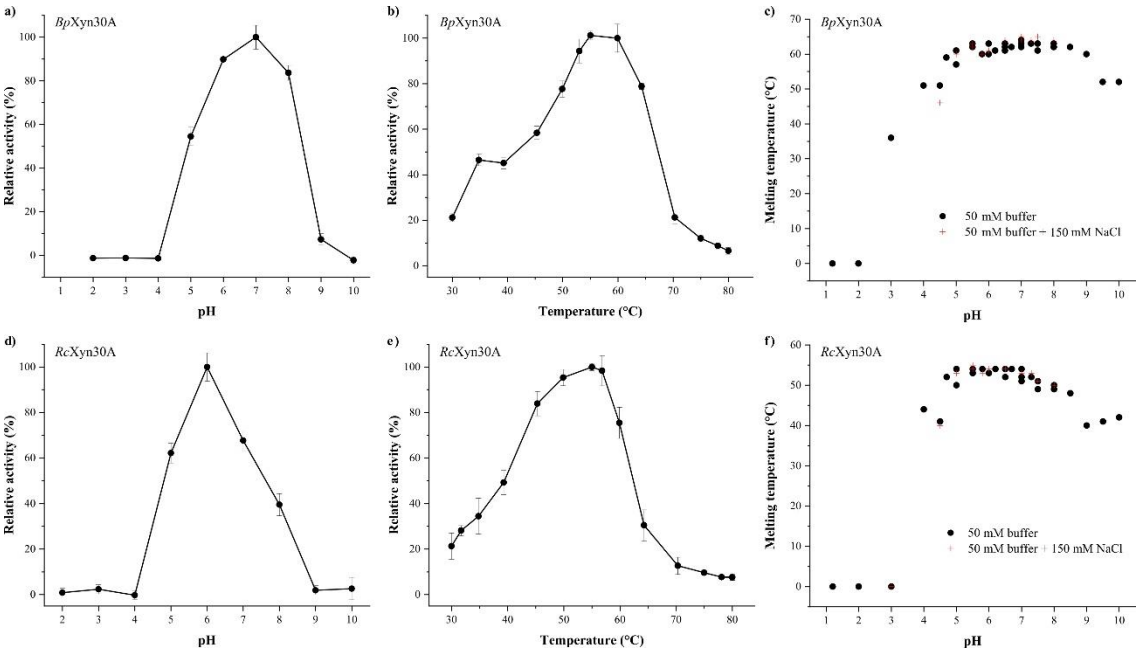

Figure 7 – Effects of (a; d) pH and (b; c) temperature on the activity of *Bp*Xyn30A (upper panels) and *Rc*Xyn30A (lower panels). Optimal pH was assessed in 40 mM ABF buffer at 40 °C, and temperature conditions were determined using 20 mM Tris-HCl buffer pH 7 and 20 mM sodium phosphate buffer pH 6 for *Bp*Xyn30A and *Rc*Xyn30A, respectively. (c; f) Xylanases structural stability was evaluated in a set of 50 mM citrate-glycine-phosphate buffers with pHs from 1.2 to 10 using DSF assay.
Source: By the author.

Next, we decided to determine the xylanases thermostability in different temperatures (Figure 8) by measuring the proteins activity against glucuronoxylan after their incubation at said conditions. *Bp*Xyn30A is highly stable at 50 °C, exhibiting a half-life of 36 ± 2 h, but times longer than 3.5 h at 55 °C resulted in fast enzyme inactivation ($t_{1/2}$ = 3.6 ± 0.1 h) (Figure 8a). Additionally, *Rc*Xyn30A is significantly more stable at 40 than 45 °C, considering that its half-life is 45 ± 6 h for the former and 2.2 ± 0.2 h for the latter (Figure 8b). Since both enzymes tolerate well milder temperatures, we chose to perform the enzymatic degradation of pretreated biomass at 40 °C. The high stability at said temperatures is an interesting feature in terms of biotechnological applications, as it allows the use of the enzymes for longer periods of time.

However, since *Bp*Xyn30A and *Rc*Xyn30A displayed $t_{1/2}$ of 3.6 ± 0.1 h and 0.35 ± 0.02 h (or approximately 21 min) at their optimal temperatures, respectively, and considering that incubation times for DNS assays were shorter than xylanases half-life in reaction conditions, we decided to maintain those temperatures for additional enzymatic characterization assays.

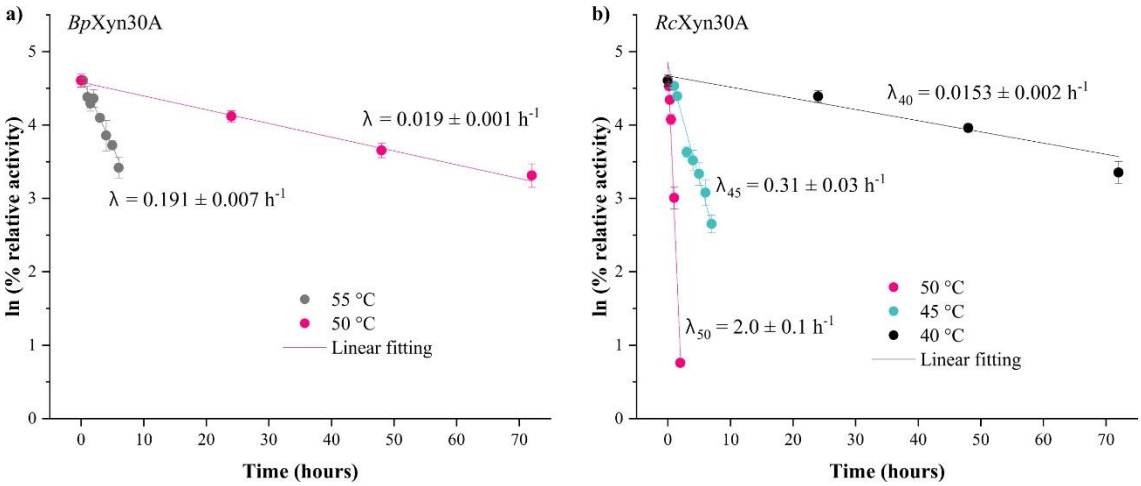

Figure 8 – Natural logarithm of (a) *Bp*Xyn30A and (b) *Rc*Xyn30A residual activity (in percentage of relative activity) as a function of time. Enzymes were preincubated in their optimal pH and under different temperature settings. Aliquots were removed over time and submitted to DNS assay.
Source: By the author.

4.3.2 Substrate specificity and kinetic parameters

Among all eight tested substrates, *B. pumilus* and *R. champanellensis* xylanases were only able to use beechwood 4-O-methyl-glucuronoxylan as substrate, with specific activities obtained using DNS method[54] equal to 125.6 ± 5.4 U/mg and 13.5 ± 0.7 U/mg, respectively. The absence of activity against rye arabinoxylan, composed of ~38% of arabinose ramifications on positions 2 and/or 3 of the backbone xylopyranose residues emphasizes the dependency of *Bp*Xyn30A and *Rc*Xyn30A on the xylan decoration, a major characteristic shared between GH30_8 xylanases.[23]

Compared to previously calculated values for xylanases from family 30, *Rc*Xyn30A showed the highest specific activity towards glucuronoxylan (Table 2), with exception of the *Bacillus subtilis* str. 168[70] and *Paenibacillus favisporus* CC02-N2[79] enzymes. For instance, the closest bacterial GH30 proteins phylogenetic related to the one described in this work, i.e., *Ruminiclostridium papyrosolvens* DSM 2782[81] and *Clostridium acetobutylicum* ATCC 824[74] xylanases, present specific activities almost 115-times and 1.4-times smaller than *Rc*Xyn30A, respectively. In contrast, *Bp*Xyn30A exhibited one of the lowest specific activities among those

of its phylogenetic branch, staying just above the xylanase from *B. licheniformis* SVD1 (Table 2).

Most characterized prokaryotic glucuronoxylanases have been studied in the presence of CBMs, and the vast majority of them belong to the families 6, 9 and 35 (Table 1), known to bind xylan. In a study published by Valenzuela *et al.* (2012)[78], *P. barcinonensis* GH30_8 xylanase linked to a CBM35 domain presented an identical substrate specificity to the catalytic domain alone, which diverges from what was observed with *Rc*Xyn30A, i.e., *R. champanellensis* GH30_8 domain alone exhibited a specific activity of 100.3 ± 5.4 U/mg. We hypothesize that the increase in the *Rc*Xyn30A full (GH + CBM9) activity is probably related to protein proximity to the substrate due to anchoring by the CBM.

Curiously, *R. champanellensis* CBM9 display no significant similarity to the two CBM9 structures deposited in PDB (PDB entries 7NWN and 1I82 from *Caldicellulosiruptor acetigenus* I77R1B and *Thermotoga maritima* MSB8, respectively), whereas it shares a percentage identity of approximately 50% with *Acetivibrio thermocellus* CBM22, according to a BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) search against PDB. This CBM22 binds to oat spelt xylan[87], an arabinoglucuronoxylan substituted mainly with arabinose and 4-O-methylglucuronic acid residues.[88]

Table 2 – Biochemical properties of published GH30_8 xylanases deposited in CAZy database. pH and temperature conditions described below are not necessarily the enzyme's optimal ones. Furthermore, the xylanases' names were chosen based on their organism of origin and year of publication (suffix A and B). NP: not published. * Experimental errors have not been reported.

Organism / Name	**Substrate**	**Specific activity (U/mg)**	**pH**	**Temperature (°C)**	**Reference**
Bacillus pumilus / *Bp*Xyn30A	Beechwood xylan	13.5 ± 0.7	7	55	Present work
R. champanellensis / *Rc*Xyn30A full	Beechwood xylan	125.6 ± 5.4	6	50	Present work
R. champanellensis / *Rc*Xyn30A GH	Beechwood xylan	100.3 ± 5.4	6	50	Present work
Aeromonas caviae ME-1 / *Ac*Xyn30A	NP	NP	NP	NP	[67]
Bacillus licheniformis SVD1 / *Bl*Xyn30A	Glucuronoxylan	7.87 ± 0.69	6	37	[68]
Bacillus sp. BP7 / *Bs*Xyn30B	Birchwood xylan	34 ± *	6	55	[69]
Bacillus subtilis str. 168 / *Bs*Xyn30A	Sweetgum xylan	130 ± 5	6	65	[70]
Bacteroides ovatus ATCC 8483 / *Ba*Xyn30A	Birchwood xylan	NP	7	37	[25]

(continued)

(continuation)

Cellulosilyticum ruminicola CGMCC 1.5065 / *Cr*Xyn30A	NP	NP	NP	NP	[72]
Cellulosilyticum ruminicola MMBC-1 / *Cr*Xyn30B	Glucuronoxylan	70.7 ± 0.5	6	50	[73]
Clostridium acetobutylicum ATCC 824 / *Ca*Xyn30A	Beechwood xylan	90.9 ± 7.5	4	40	[74]
Dickeya chrysanthemi D1 / *Dc*Xyn30A	4-O-Methyl-D-glucuronoxylan	18.1 ± 0.7	6	40	[89]
Melioribacter roseus P3M-2 / *Mr*Xyn30A	Beechwood xylan	55.3 ± *	6.5	80	[77]
Paenibacillus barcinonensis BP-23 / *Pb*Xyn30A	Beechwood xylan	30.3 ± 1.5	6.5	50	[78]
Paenibacillus favisporus CC02-N2 / *Pf*Xyn30A	Beechwood xylan	244.0 ± 4.1	6.5	50	[79]
Paenibacillus sp. W-61 / *Ps*Xyn30A	Birchwood xylan	NP	5	50	[80]
Ruminiclostridium papyrosolvens DSM 2782 / *Rp*Xyn30A	Sweetgum xylan	1.1 ± 0.1	4.5	30	[81]
Streptomyces turgidiscabies C56 / *St*Xyn30A	Beechwood xylan	NP	6.5	55	[82]
Acetivibrio thermocellus ATCC 27405 / *At*Xyn30A	Beechwood xylan	34.0 ± 0.8	6	70	[83]

Source: By the author.

Afterwards, kinetic parameters of both glucuronoxylanases were determined using beechwood glucuronoxylan as substrate (Table 3). Kinetic parameters of other xylanases from family GH30_8, all used in the construction of the phylogenetic tree (Figure 5), are also given in Table 3. *Bp*Xyn30A binds more effectively to beechwood glucuronoxylan than *Rc*Xyn30A, as observed in K_M values. This might be related to the fact that the *R. champanellensis* xylanase displays higher affinity towards the glucuronic acid decoration, which is a limiting factor for the binding of the enzyme/substrate complex; while *Bp*Xyn30A likely binds to other regions of this hemicellulose, including the linear ones (see discussion in sections 4.3.3 and 4.5). On the flip side, the catalytic efficiency (k_{cat}/K_M) of *Rc*Xyn30A is almost 3 times higher than *Bp*Xyn30A, showing that the former enzyme is more specific for glucuronoxylan than the latter.

Compared to the literature (Table 3), both values are within the range observed for GH30_8 xylanases.

Table 3 – Kinetic parameters of *Bp*Xyn30A, *Rc*Xyn30A and others xylanases from family GH30_8.
* Experimental errors have not been reported.

Organism / xylanase	**Substrate**	**K_M (mg.mL^{-1})**	**k_{cat} (s^{-1})**	**k_{cat}/K_M (mL.mg^{-1}.s^{-1})**	**Reference**
Bacillus pumilus / *Bp*Xyn30A	Beechwood xylan	4.8 ± 0.3	26.7 ± 0.5	5.5 ± 0.4	Present work
R. champanellensis / *Rc*Xyn30A full	Beechwood xylan	7.6 ± 0.9	122.5 ± 7.6	16.1 ± 3.0	Present work
Cellulosilyticum ruminicola MMBC-1 / *Cr*Xyn30B	Glucuronoxylan	0.7 ± 0.0	11.8 ± 0.4	16.9 ± 0.8	[73]
Melioribacter roseus P3M-2 / *Mr*Xyn30A	Birchwood xylan	10.5 ± 3.0	104.4 ± *	9.9 ± *	[77]
Bacillus subtilis str. 168 / *Bs*Xyn30A	Sweetgum xylan	1.63 ± *	2.64 ± *	1.62 ± *	[70]
Clostridium acetobutylicum ATCC 824 / *Ca*Xyn30A	Beechwood xylan	3.12 ± 0.15	86 ± 8	27.6 ± 3.9	[74]
Dickeya chrysanthemi D1 / *Dc*Xyn30A	Glucuronoxylan	1.64 ± 0.42	34.1 ± 1.6	20.8 ± 5.4	[76]
Acetivibrio thermocellus ATCC 27405 / *At*Xyn30A	Beechwood xylan	2.2 ± *	2008.3 ± *	912.9 ± *	[83]
Paenibacillus favisporus CC02-N2 / *Pf*Xyn30A	Beechwood xylan	4 ± *	222 ± *	55 ± *	[79]
Paenibacillus barcinonensis BP-23 / *Pb*Xyn30A	Beechwood xylan	14.72 *	25.17 ± *	1.71 ± *	[78]

Source: By the author.

4.3.3 Enzymatic cleavage pattern

The xylanases were tested against linear X4, X5 and X6 (Figure 10), and the enzymatic cleavage pattern was determined using beechwood glucuronoxylan as substrate (Figure 9). Samples analyzed by HPAEC-PAD showed that whilst *Bp*Xyn30A released X2, X3 e X4 as main products, as well as minor quantities of X5 and X6 (Figure 9a), *Rc*Xyn30A generated practically no linear XOS with DP 2-6 during xylan degradation (Figure 9c). That data

combined to the absence of xylotetraose, xylopentaose and xylohexaose hydrolysis (Figure 10b, 10d, 10f) by *Rc*Xyn30A after 24 h of reaction suggests that this glucuronoxylanase is not able or highly inefficient in the cleavage of unbranched xylooligosaccharides. In contrast, *B. pumilus* GH30 was able to convert xylotetraose into X2 (Figure 10a), xylopentaose into X2 and X3 (Figure 10c) and xylohexaose into X2, X3 and X4 after longer periods of incubation. However, chromatograms indicate that longer oligosaccharides tend to be more favorable for enzymatic hydrolysis by *Bp*Xyn30A, as can be observed by comparing the size of initial (0 min) and final (24 h) peaks of each substrate used.

A previous study with *Dc*Xyn30A demonstrated that this bona-fide GH30_8 xylanase from *D. chrysanthemi* exhibits a specific activity on X5 at least 600 times smaller than in $MeGlcA_3X4$[76], which shows that glucuronic acid plays an important role in substrate recognition by glucuronoxylan-specifics xylanases, such as *Rc*Xyn30A. Additionally, all *Bacillus* sp. enzymes from family 30 that have been characterized presented some degree of activity against linear XOS. This led us to hypothesize that *Bp*Xyn30A recognizes glucuronic acid substitutions like the rest of the prokaryotic glucuronoxylanases' family, but in the absence of cleavage sites near decorations, it can also hydrolyze linear regions of xylan.

To gain information regarding the products released by the enzymes when incubated with glucuronoxylan, reaction samples were also analyzed by MALDI-TOF mass spectrometry (Figure 9). In accordance with previous HPAEC-PAD results, linear xylooligosaccharides were detected in the *Bp*Xyn30A spectrum, along with MeGlcA branched XOS (Figure 9b). Curiously, this xylanase's products were smaller than those observed for *Rc*Xyn30A, which highlights *Bp*Xyn30A activity against unbranched oligosaccharides. Furthermore, *Rc*Xyn30A liberated long branched oligosaccharides, containing a MeGlcA decoration on the last but one xylose moiety towards the reducing end (Figure 9d), as previously demonstrated by Vršanská *et al.* (2007)[90] employing a *D. chrysanthemi* GH30_8 xylanase. The smallest observed product liberated by *Rc*Xyn30A was MeGlcAX5, indicating that the xylanase has more affinity for longer substrates.

Vršanská *et al.* (2007) also explored the relationship between the MeGlcA content in polysaccharides and maximum amount of products that can be generated by *Dc*Xyn30A[90]. This was accomplished by treating the oligosaccharides produced by the enzyme with an α-glucuronidase, and quantifying equivalents of xylobiose and glucuronic acid. The authors reached the conclusion that most products were monoglucuronylated, although doubly decorated products were also observed (ratio of 1.21 for MeGlcA / X2 content from beechwood xylan), which was associated with the uneven branch arrangement in the substrate. Only

monoglucuronylated sugars were observed in MALDI-TOF spectra of *Bp*Xyn30A and *Rc*Xyn30A reaction samples (Figure 9), which indicates a prevalence of more spaced decorations in the used substrate.

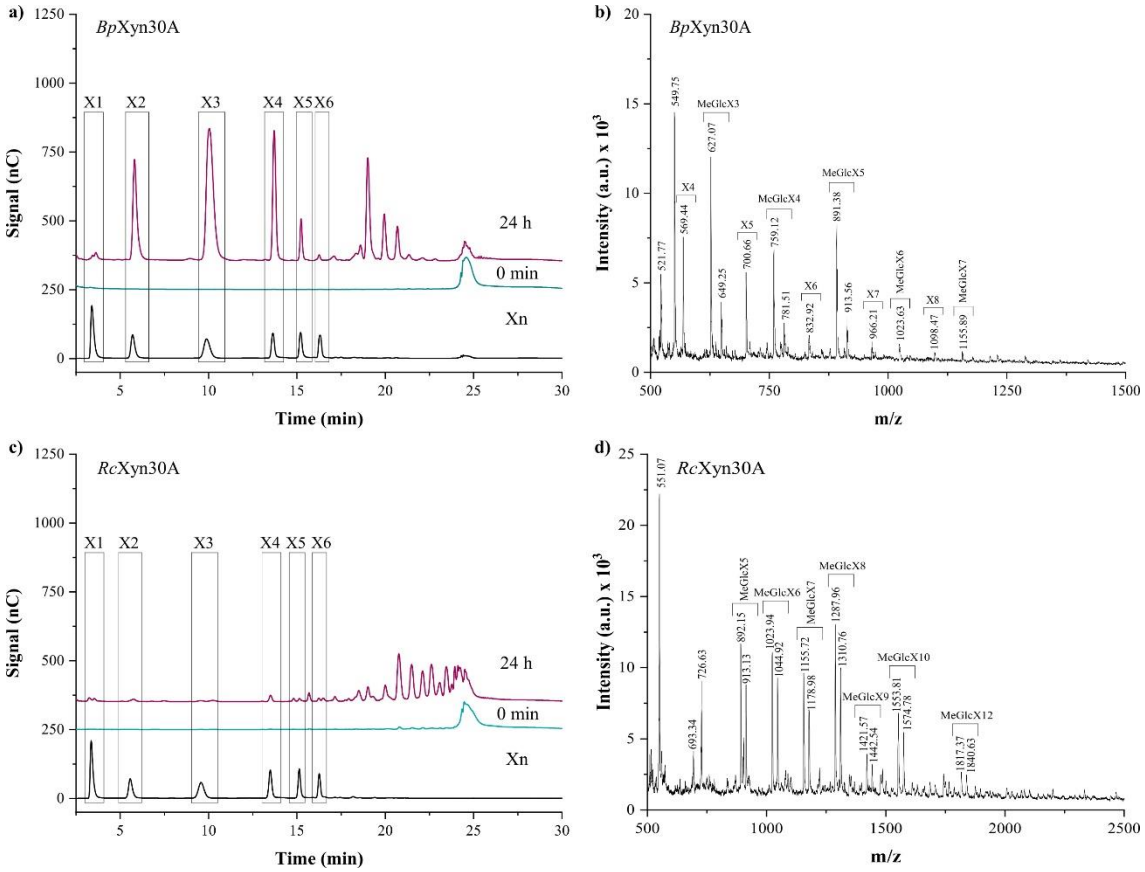

Figure 9 – Enzymatic cleavage pattern of (a; b) *B. pumilus* (upper panels) and (c; d) *R. champanellensis* lower panels) xylanases using 0.5% (w/v) beechwood 4-O-methyl-glucuronoxylan as substrate. Reactions were carried out with 1 µM enzyme and in the protein's optimal conditions of pH and temperature. Samples were analyzed by (a; c) HPAEC-PAD and (b; d) MALDI-TOF. The peaks within MALDI-TOF spectrum were assigned based on the masses of products with Na^+ or 2 Na^+ adducts. Marked standards represent, Xn: xylose and xylooligosaccharides (DP 2-6), X3: xylotetraose, X4: xylotetraose, X5: xylopentaose, X6: xylohexaose, MeGlcAXn: Xn with MeGlcA ramifications on the penultimate xylose moiety towards the reducing end.

Source: By the author.

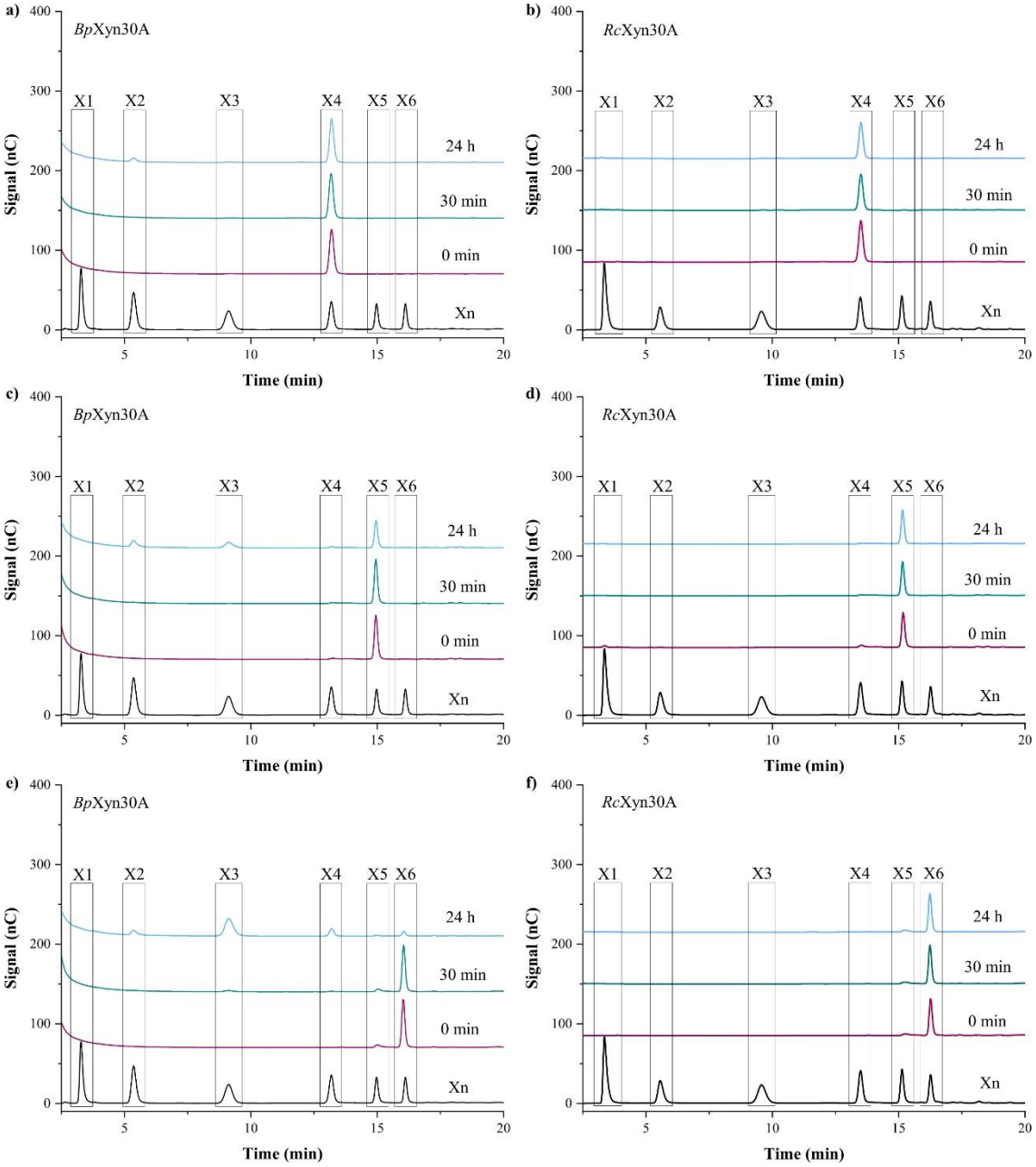

Figure 10 – Enzymatic cleavage pattern of *Bp*Xyn30A (right panels) and *Rc*Xyn30A (left panels) using 0.05 mg/mL of (a; b) xylotetraose, (c; d) xylopentaose or (e; f) xylohexaose. Reactions were carried out with either 180 nM *Bp*Xyn30A in 20 mM Tris-HCl buffer pH 7 or 15.6 nM *Rc*Xyn30A in 20 mM sodium phosphate buffer pH 6, and maintained at 50 °C and 45 °C, respectively. Samples were analyzed by HPAEC-PAD. Marked standards represent, Xn: xylose and xylooligosaccharides (DP 2-6), X3: xylotetraose, X4: xylotetraose, X5: xylopentaose, X6: xylohexaose.
Source: By the author.

4.4 Enzymatic hydrolysis of alkaline pretreated biomass

Composition of both *in natura* and treated biomasses with 1% (w/v) NaOH are given in Figure 11a. The xylan of E-IN and CC-IN contain 52% and 11% of acetylation, respectively, and the latter is further decorated with ~7% arabinofuranosyl units. Alkaline pretreatment

enriched the fractions of cellulose and hemicellulose, and given that xylan percentage in CC-Alk is 3.5-times higher than E-Alk, the former appears to be more promising for enzymatic hydrolysis, despite the presence of arabinose residues that could limit the substrate cleavage sites and enzyme accessibility.

The chemical composition of untreated eucalyptus sawdust (E-IN) is in line with what was described in literature using eucalyptus woodchips[91], *E. grandis*[92-93], *E. grandis* x *E. urophylla*[93], *E. grandis* plus *E. urophylla* mixture (1:1).[94] Values range from: 39 – 44% glucans, 14 – 27% lignin, 9 – 36% hemicellulose, 0.2 – 7% ashes, 3 – 27% extractives, ~3% acetyl groups. The highest values for lignin were observed with *E. grandis* sawdust and the mixture of *E. grandis* and *E. urophylla*, whereas for extractives were with *E. grandis* and *E. grandis* x *E. urophylla* bark. Eucalyptus woodchips presented the highest fraction of xylan and *E. grandis* bark obtained the smallest portion of said polysaccharide.

The *Eucalyptus* genus is widely used for reforestation in Brazil since the 1980s, due to proprieties such as short rotation (5 to 7 years) and differentiated use of the wood. Species such as *E. urophylla, E. grandis, E. globulus* and *E. urophylla* x *E. grandis* are among the most used to produce paper, cellulose and solids.[95] That said, it is possible that the eucalyptus sawdust used contained a mixture of different parts of *E. grandis*, *E. urophylla* and *E. grandis* x *E. urophylla*.

Regarding pretreated biomass, while the alkaline pretreatment on corn cobs resulted in delignification, with a reduction of approximately 12% in lignin dry mass, practically no effect was detected on sawdust eucalyptus, besides the deacetylation of xylan. The same pattern was observed with *E. grandis* x *E. urophylla* woodchips treated with 4% (w/v) NaOH at 90 °C for 2 h: although the pretreatment only partially reduced the content of lignin (less than 3%), it resulted in the opening of the cell wall structure and enhancement of fibers surface area.[96] We attribute the delignification difference to the amount of lignin in the agro- and wood residues.

Enzymatic hydrolysis of CC-Alk and E-Alk showed that *B. pumilus* and *R. champanellensis* glucuronoxylanases were able recognize the xylan fraction of both biomasses, which indicates the presence of glucuronic acid residues attached to the main xylose chain (Figure 11). Surprisingly, reducing sugar quantification revealed that even though *Rc*Xyn30A presents a much higher specific activity towards glucuronoxylan than *Bp*Xyn30A, the latter liberated 8- and 10-fold more soluble products when incubated with CC-Alk for 6 h and 24 h (Figure 11b). That difference might be explained by *Rc*Xyn30A affinity for GlcA branches, as previously discussed in section 4.3.3, by the inhibition of the enzyme by its own hydrolysis products or even by negative adsorption of *Rc*Xyn30A to the substrate, for example.

DNS assay did not detect sugars released by *Rc*Xyn30A from E-Alk, probably due to the low xylan content in the samples and the limited xylose detection by the method. Nonetheless, HPAEC analysis indicated that substrate hydrolysis took place, and all the chromatograms are consistent with those observed for beechwood xylan degradation by the glucuronoxylanases (Figures 11c and 11d). In summary, XOS yield from CC-Alk was significantly higher than from E-Alk, as initially expected, demonstrating that the percentage of xylan in the biomass is more important than the presence of only one type of branch in the substrate. Also, products formation by both xylanases revealed a time-dependent feature, which leaves room for improvements of XOS yield. Furthermore, higher conversion yields of pretreated biomass by *Bp*Xyn30A emphasizes its standalone activity in the bacteria context, whereas *Rc*Xyn30A likely depends on a set of enzymes to depolymerize the substrate. Finally, although the total conversion yields of enzymes alone were relatively low for biotechnological applications, it is yet to be seen if longer times and/or addition of xylanases from other families (GH10 and GH11, for example) could boost the biomass degradation.

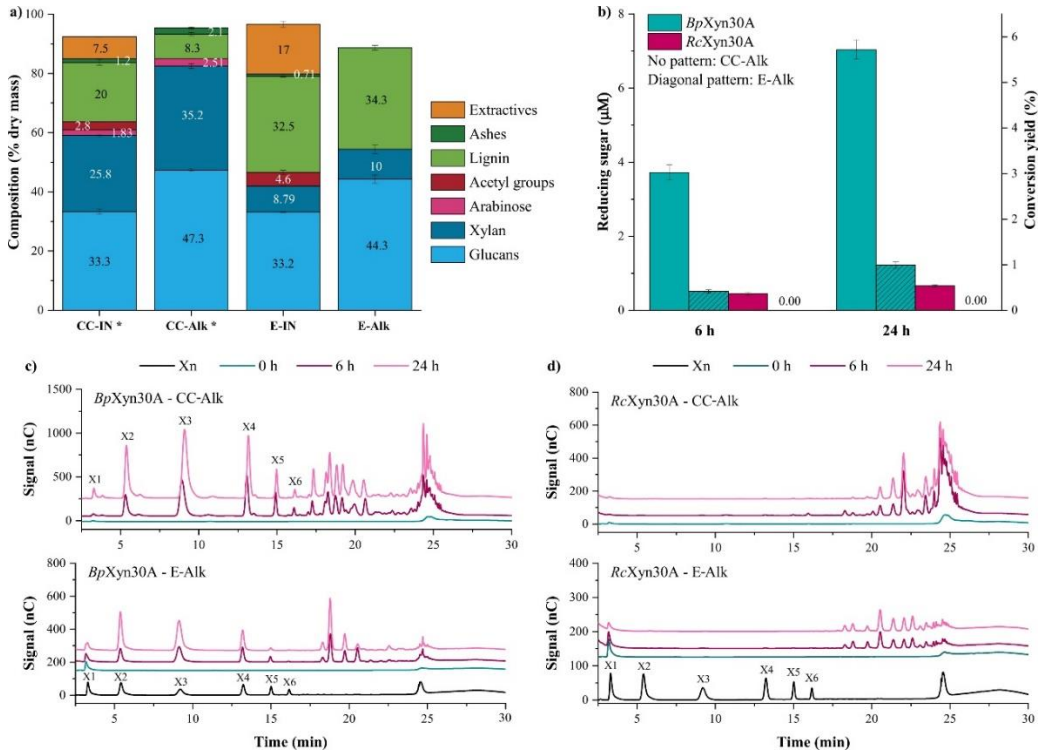

Figure 11 – Enzymatic hydrolysis of 1% (w/v) NaOH pretreated biomass. (a) Corn cob (CC) and eucalyptus sawdust (E) composition in % of dry mass before (CC-IN and E-IN) and after alkali pretreatment (CC-Alk and E-Alk). (b) Conversion yield of CC-Alk and E-Alk by *Bp*Xyn30A and *Rc*Xyn30A in terms of reducing sugar. DNS method did not detect product formation in reactions containing *Rc*Xyn30A and E-Alk. HPAEC-PAD chromatograms of (c) *Bp*Xyn30A and (d) *Rc*Xyn30A after 6 and 24 h of incubation with both pretreated biomasses. All Experiments were conducted in triplicate, and enzymatic hydrolysis was carried out with 5% (w/v) biomass (dry weight) and 0.1 mg protein in 20 mM phosphate buffer pH 6 at 40 °C. * CC-IN and CC-Alk characterization data were obtained previously.[57]

Source: By the author.

4.5 Amino acid sequence analysis and crystallographic structure determination

4.5.1 CBM cleavage with papain and X-ray crystallography

*Rc*Xyn30A construct presents two putative modules: a GH30 subfamily 8 domain, ranging from residues 14-410 (without considering the signal peptide), and a carbohydrate binding module (CBM) from family 9, which comprises the C-terminal residues 421-567. In order to crystalize *R. champanellensis* xylanase catalytic domain, the protein was partially digested with papain, and the resultant polypeptide after a one-step purification through size exclusion chromatography migrated as a single band in 15% SDS-PAGE electrophoresis with approximate molecular mass of 44 kDa (Figure 12), which is close to the theoretically predicted molecular mass of GH30_8 domain. The crystallographic structure confirmed the successful cleavage of the linker and integrity of GH domain. Data collection and refinement parameters of both glucuronoxylanases are given in Table 4.

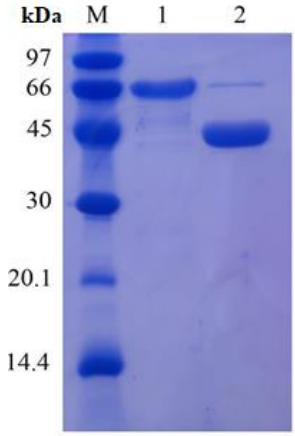

Figure 12 – *R. champanellensis* xylanase partial cleavage with papain. M: low weight molecular marker; 1: full-length *Rc*Xyn30A (GH and CBM9 domains); 2: GH3_80 domain of *Rc*Xyn30A after linker cleavage with papain and purification using size exclusion chromatography.
Source: By the author.

Table 4 – Data collection and refinement statistics. Values in parentheses refer to highest-resolution shell. $R_{merge} = \frac{\sum_{hkl}\sum_{i=1}^{n}|I_i(hkl)-\bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n}I_i(hkl)}$, $R_{meas} = \frac{\sum_{hkl}\sqrt{n/n-1}\sum_{i=1}^{n}|I_i(hkl)-\bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n}I_i(hkl)}$, $R_{pim} = \frac{\sum_{hkl}\sqrt{1/n-1}\sum_{i=1}^{n}|I_i(hkl)-\bar{I}(hkl)|}{\sum_{hkl}\sum_{i=1}^{n}I_i(hkl)}$, CC* $= \sqrt{\frac{2CC_{1/2}}{1+CC_{1/2}}}$, $R_{work,free} = \frac{\sum_{hkl}|F_{obs}(hkl)-F_{calc}(hkl)|}{\sum_{hkl}|F_{obs}(hkl)|}$, where I_i and $\bar{I}$ are the i[th] and mean measurement of intensity reflection, and $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitudes, respectively. A set of 5% randomly selected data excluded from the refinement was used for R_{free} calculation.

Data collection	***Bp*Xyn30A**	***Rc*Xyn30A**
Beamline	MANACA (LNLS-Sirius)	MANACA (LNLS-Sirius)
Wavelength (Å)	0.977	0.977
Resolution range (Å)	48.61 – 2.16 (2.237 – 2.16)	41.46 – 1.37 (1.419 – 1.37)
Space group	P 21 21 21	P 12_11
Unit cell		
a, *b*, *c* (Å)	59.084, 101.484, 158.689	56.62, 55.019, 63.329
α, β, γ (°)	90, 90, 90	90, 95.157, 90
Total reflections	102542 (9745)	536923 (46727)
Unique reflections	51300 (4885)	81419 (8035)
Multiplicity	2.0 (2.0)	6.6 (5.8)
Completeness (%)	98.54 (95.63)	99.87 (99.10)
Mean $I/\sigma_{(I)}$	8.58 (1.27)	11.53 (1.92)
Wilson B-factor (Å^2)	35.05	11.41
R_{merge}	0.06205 (0.5742)	0.1003 (0.8854)
R_{meas}	0.08775 (0.8121)	0.109 (0.8861)
R_{pim}	0.06205 (0.5742)	0.04215 (0.3613)
$CC_{1/2}$	0.996 (0.489)	0.998 (0.667)
CC*	0.999 (0.81)	0.999 (0.894)
Molecules per asymmetric unit	2	1
Refinement	***Bp*Xyn30A**	***Rc*Xyn30A**
N° of protein residues	778	398
N° of reflections used on refinement	51296 (4885)	81415 (8035)
N° of reflections used for R_{free}	2567 (244)	4071 (401)
R_{work}	0.1786 (0.2912)	0.1399 (0.2177)
R_{free}	0.2106 (0.3176)	0.1639 (0.2645)
CC (work)	0.971 (0.688)	0.973 (0.867)
CC (free)	0.962 (0.676)	0.969 (0.773)
RMSD bonds (Å)	0.002	0.006
RMSD angles (°)	0.53	0.91
Ramachandran favored (%)	97.03	96.97
Ramachandran allowed (%)	2.71	2.78
Ramachandran outliers (%)	0.76	0.25
Rotamer outliers (%)	0.76	1.46
Clash score	3.35	0.32
Average *B*-factor (Å^2)	38.15	14.24
PDB code	Not published yet	8VG9

Source: By the author

B. pumilus and *R. champanellensis* xylanases monomers consist of a $(\beta/\alpha)_8$-barrel domain, that belongs to Clan A and catalyze the double-displacement reaction[20], and a β-sheet domain (Figures 13a and 13b) tightly associated with the $(\beta/\alpha)_8$ catalytic core by N- and C-terminal linkers (amino acids 16 and 295-297 for *Bp*Xyn30A; amino acids 27 and 314-316 for *Rc*Xyn30A). Aside from the catalytic domain, St. John and colleagues[71] observed that the β-domain of *Bs*Xyn30A could bind to MeGlcAX2, and its role as a putative CBM was suggested by the authors. Structure-based sequence alignment (Figure 13c) revealed that about 18 amino acids compose the enzymes substrate-binding cleft, with two conserved catalytic residues (Glu138 and Glu148 in *Bp*Xyn30A; Glu148 and Glu247 in *Rc*Xyn30A). Compared to other enzymes from the same family, *R. champanellensis* xylanase distinguishes itself by presenting a larger L5 loop (colored in orange) (Figure 13d), that we propose to act generating additional binding site for the substrate (subsite +3). In addition, *Bp*Xyn30A and *Rc*Xyn30A have a small α-helix in their loop 1 (colored in red) (Figures 13a and 13b), that coincidentally harbors a very conserved amino acid among the GH30_8 family (Trp84 and Trp37, respectively), and has also been observed in *Bs*Xyn30A (PDB entry 3KL0) and *Pb*Xyn30A (PDB entry 4QAW).

Of all eight loops around the substrate binding site, only loop two (L2, colored in green) does not interact with the ligand (Figures 13a and 13b). Curiously, two of the GH30_8 xylanases with activity on both arabinose and glucuronic acid substituted xylan (i.e., *Ca*Xyn30A and *Rp*Xyn30A) have larger β2-α2 loops (Figure 13c). St. John and colleagues (2014) noticed that L2 loop in *Rp*Xyn30A provides additional intramolecular contacts with adjacent secondary structures and suggested its role as supporting the β-structure of L3 loop.[81]

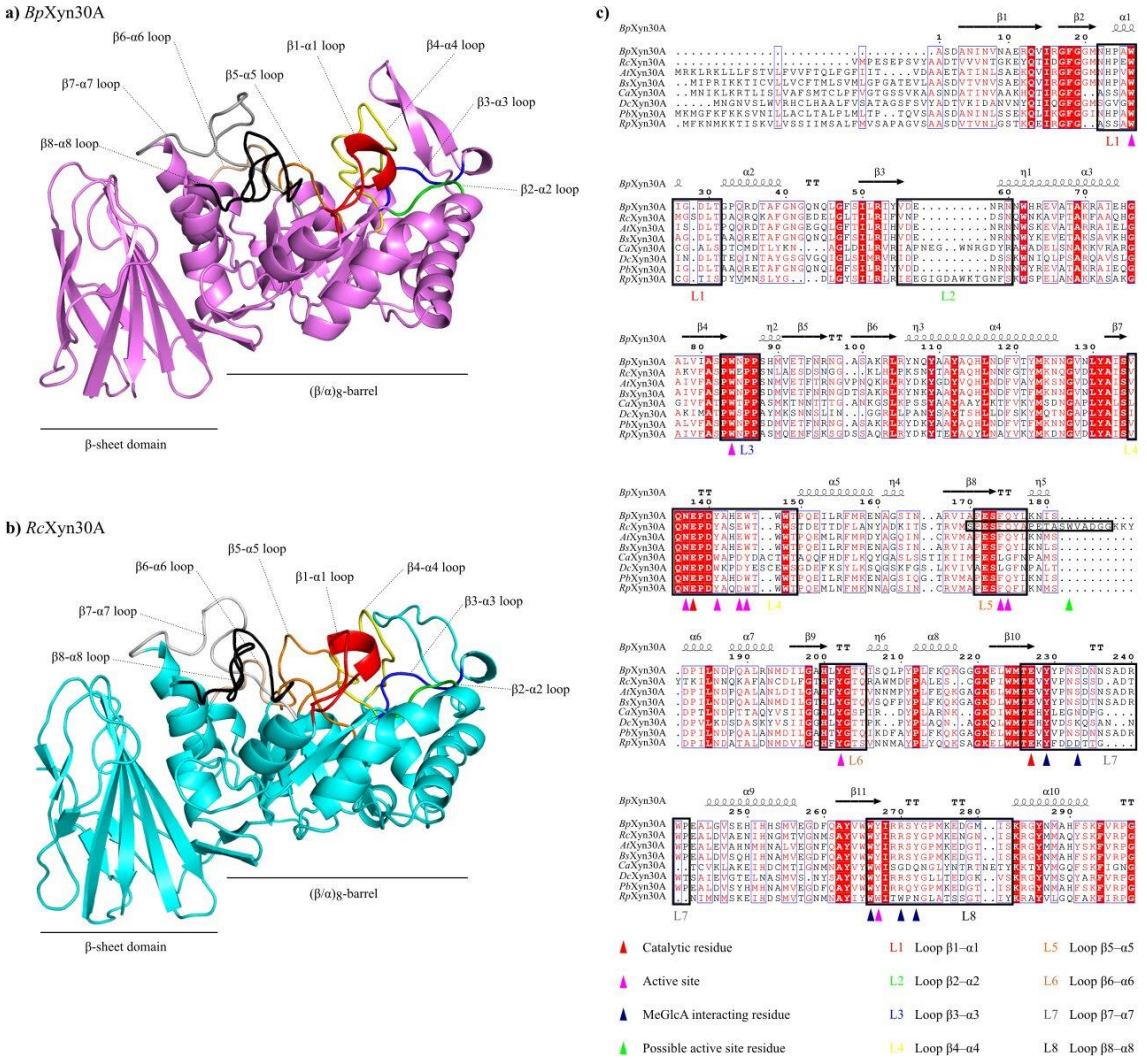

Figure 13 – X-ray crystallography of glucuronoxylanases and structure-based sequence alignment. Overview of the (a) *Bp*Xyn30A and (b) *Rc*Xyn30A structure, revealing a conserved $(\beta/\alpha)_8$-barrel and β-sheet domain. Loops circumventing the active site are highlighted in red (L1, loop β1-α1), green (L2, loop β2-α2), blue (L3, loop β3-α3), yellow (L4, loop β4-α4), orange (L5, loop β5-α5), salmon (L6, loop β6-α6), grey (L7, loop β7-α7) and black (L8, loop β8-α8); with the exception of L2, remaining loops can interact directly with ligands. (c) Partial structure-based sequence alignment of *Bp*Xyn30A, *Rc*Xyn30A, *At*Xyn30A (PDB entry 4CKQ), *Bs*Xyn30A (PDB entry 3KL0), *Ca*Xyn30A (PDB entry 5CXP), *Dc*Xyn30A (PDB entry 1NOF and 2Y24), *Pb*Xyn30A (PDB entry 4QAW) and *Rp*Xyn30A (PDB entry 4FMV). Red, pink, blue and green triangles indicate residues with catalytic activity, that forms the active site, interact with MeGlcA and possibly form positive subsites, respectively.
Source: By the author.

Superposition of *Bp*Xyn30A and *Rc*Xyn30A with *Dc*Xyn30A in complex with $MeGlcA_2X3$ (PDB entry 2Y24) resulted in a RMSD of 0.792 Å and 0.901 Å, respectively. MeGlcA interacting subsite (Figure 14), often referred to as -2b subsite by some authors[71,74,97], harbors important amino acids for substrate recognition, including a positively charged arginine (Arg270 in *Bp*Xyn30A and Arg289 in *Rc*Xyn30A) that forms two ionic bonds between its guanidinium group and MeGlcA C6 carboxylate group.[98] Absence of this Arg was shown to imply a change in substrate specificity: *Ca*Xyn30A[74] and *Rp*Xyn30A[81], for instance, have

aspartate (Asp268 in *Ca*Xyn30A) and tryptophan (Trp265 in *Rp*Xyn30A) occupying nearly the same space as arginine, and in both cases was observed a higher specific activity towards arabinoxylan.

In addition to the salt bridge, MeGlcA moiety is predicted to interact to *Bp*Xyn30A and *Rc*Xyn30A via five hydrogen bonds: three with hydroxyl groups of Tyr residues (Tyr229 in the former and Tyr249 in the latter with MeGlcA O2; Tyr272 in the former and Tyr291 in the latter with MeGlcA O5 and O6), one with the amide nitrogen of Trp residue (Trp266 NE1 in the former and Trp285 NE1 in the latter with MeGlcA O5) and one with OH group of Ser residue (Ser233 OG in the former and Ser253 OG in the latter with MeGlcA O3) (Figures 14b and 14d). All in all, interactions between the amino acids that form -2b subsite and methyl glucuronic acid residue appears to be crucial for substrate recognition by both glucuronoxylanases of the present study, although in *Bp*Xyn30A the distances between amino acids and MeGlcA were larger than those of *Rc*Xyn30A.

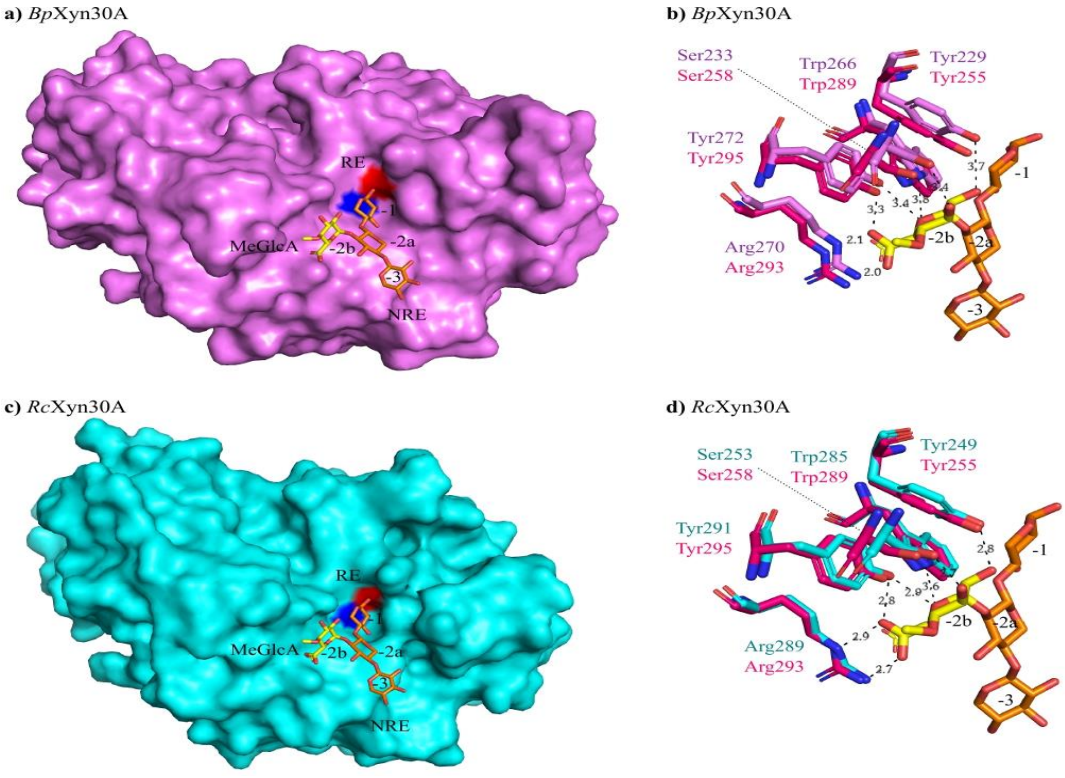

Figure 14 – Superposition of MeGlcA$_2$X3, obtained from *Dc*Xyn30A structure in complex with ligand (PDB entry 2Y24), with (a) *Bp*Xyn30A and (c) *Rc*Xyn30A catalytic center. Catalytic acid/base residues (Glu138 for the former and Glu148 for the latter) and nucleophile (Glu227 for the former and Glu247 for the latter) are colored in red and blue, respectively, and RE and NRE stands for reducing end and non-reducing ends. (b; d) Amino acids in the vicinity of MeGlcA in *Bp*Xyn30A (light purple), *Rc*Xyn30A (cyan) and *Dc*Xyn30A (pink) are represented as sticks; xylose and MeGlcA are colored in orange and yellow, respectively. Black dashed lines indicate interacting atoms between protein and ligand, and distances are given in Å.

Source: By the author.

4.5.2 Analysis of substrate-binding site

To better analyze the glucuroxylanases binding cleft, we performed two superpositions with the *Bp*Xyn30A and *Rc*Xyn30A structures: the first one with *Dc*Xyn30A structure in complex with $MeGlcA_2X3$ and an imidazole molecule (PDB entry 2Y24)[98], in order to study the amino acids interacting with ligands in the negative and +1 subsites, respectively, and secondly, with *At*Xyn30A-E225A structure in complex with xylobiose and xylotriose (PDB entry 5A6M).[84] The latter protein is a nucleophilic residue mutant of *Acetivibrio thermocellus* xylanase (*At*Xyn30A) that has a RMSD of 0.368 Å and 0.547 Å when superimposed with *Bp*Xyn30A and *Rc*Xyn30A, respectively, and the only GH30_8 xylanase with solved structure with xylooligosaccharides occupying positive subsites. Additional comparisons were conducted using *Bacillus subtilis* str. 168 protein (*Bs*Xyn30A) bond to $MeGlcAX_2$ (PDB entry 3KL5, RMSD of 0.294 Å with *Bp*Xyn30A and 0.546 Å with *Rc*Xyn30A)[71] (Figures 15 and 16).

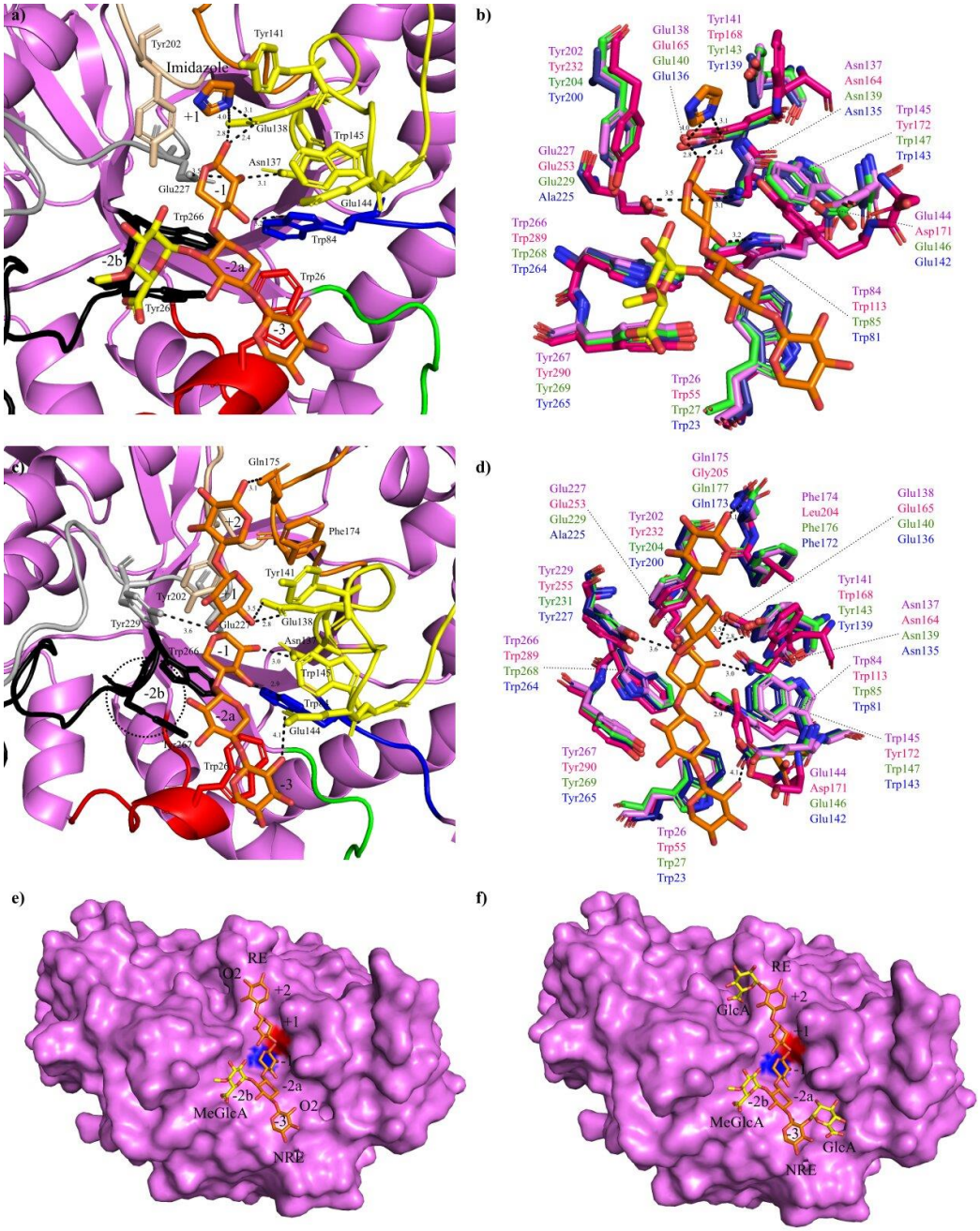

Figure 15 – Structural bases for *Bp*Xyn30A recognition of glucuronoxylan. Positioning and details of interaction of (a) $MeGlcA_2X3$ and imidazole, co-crystallized with *Dc*Xyn30A (PDB entry 2Y24), and (c) xylobiose and xylotriose, co-crystallized with *At*Xyn30A (PDB entry 5A6M), in *B. pumilus* xylanase catalytic site. Loops in the vicinity of ligands and its respective interacting amino acids are colored in red (L1, loop β1-α1), green (L2, loop β2-α2), blue (L3, loop β3-α3), yellow (L4, loop β4-α4), orange (L5, loop β5-α5), salmon (L6, loop β6-α6), grey (L7, loop β7-α7) and black (L8, loop β8-α8). Hydrogen bonds are marked with black dashed lines and distances are given in Å. Comparison of *Bp*Xyn30A amino acids (in light purple) involved in binding (b) $MeGlcA_2X3$ and imidazole and (d) X2 and X3 with the correspondent residues of *Dc*Xyn30A (in pink), *At*Xyn30A (in dark blue) and *Bs*Xyn30A (in green). Front view of the enzyme substrate binding site in the presence of (e) $MeGlcA_2X3$ (occupying glycone subsites) plus X2 (occupying aglycone subsites) and (f) $MeGlcA_{2,3}X3$ (occupying glycone subsites) plus $GlcA_2X2$ (occupying aglycone subsites). Ligands fitted within negative subsites were build based on $MeGlcA_2X3$, whereas those of positive subsites were computationally extrapolated from X2, co-crystallized with *At*Xyn30A. Catalytic acid/base residue (Glu138) and nucleophile (Glu227) are colored in red and blue, respectively, and RE and NRE stands for reducing end and non-reducing end.

Source: By the author

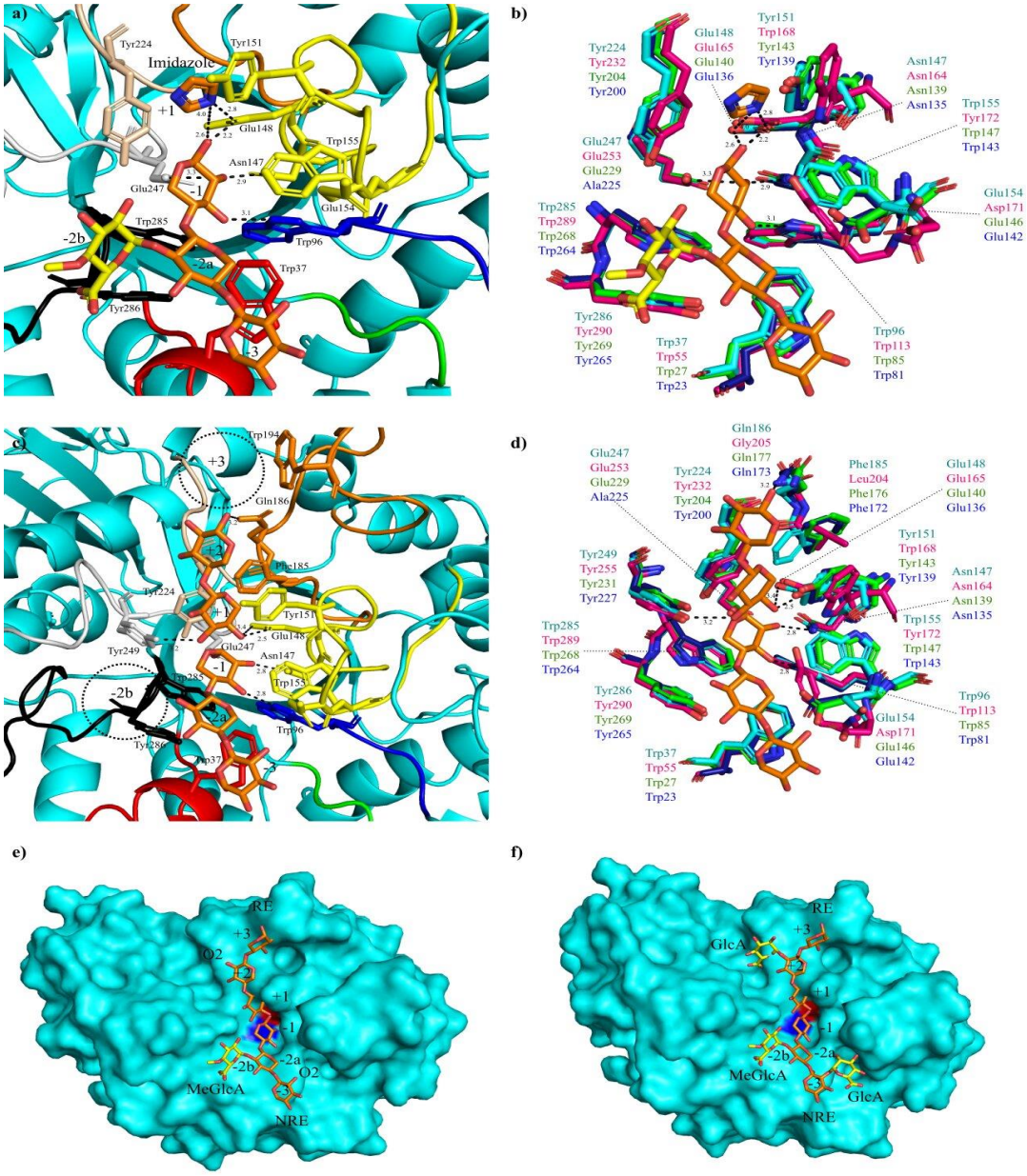

Figure 16 – Structural bases for *Rc*Xyn30A recognition of glucuronoxylan. Positioning and details of interaction of (a) MeGlcA$_2$X3 and imidazole, co-crystallized with *Dc*Xyn30A (PDB entry 2Y24), and (c) xylobiose and xylotriose, co-crystallized with *At*Xyn30A (PDB entry 5A6M), in *R. champanellensis* xylanase catalytic site. Loops in the vicinity of ligands and its respective interacting amino acids are colored in red (L1, loop β1-α1), green (L2, loop β2-α2), blue (L3, loop β3-α3), yellow (L4, loop β4-α4), orange (L5, loop β5-α5), salmon (L6, loop β6-α6), grey (L7, loop β7-α7) and black (L8, loop β8-α8). Hydrogen bonds are marked with black dashed lines and distances are given in Å. The aromatic ring of Trp194 (in loop 5) might be involved in binding a xylose moiety in subsite +3. Comparison of *Rc*Xyn30A amino acids (in cyan) involved in binding (b) MeGlcA$_2$X3 and imidazole and (d) X2 and X3 with the correspondent residues of *Dc*Xyn30A (in pink), *At*Xyn30A (in light purple) and *Bs*Xyn30A (in green). Front view of the enzyme substrate binding site in the presence of (e) MeGlcA$_2$X3 (occupying glycone subsites) plus X3 (occupying aglycone subsites) and (f) MeGlcA$_{2,3}$X3 (occupying glycone subsites) plus GlcA$_2$X3 (occupying aglycone subsites). Ligands fitted within negative subsites were build based on MeGlcA$_2$X3, whereas those of positive subsites were computationally extrapolated from X2, co-crystallized with *At*Xyn30A. Catalytic acid/base residue (Glu148) and nucleophile (Glu247) are colored in red and blue, respectively, and RE and NRE stands for reducing end and non-reducing end.

Source: By the author

With the exception of the extra stacking interaction in the +3 subsite of *Rc*Xyn30A, all amino acids involved in the enzyme binding cleft formation are the same for the glucuronoxylanases of this study. The xylopyranose unit in subsite -1 is expected to be interacting with several amino acids (Figures 15a, 15b, 16a and 16b). Five hydrogen bonds can take place, two of them between the O1 of xylose moiety and positions OE1 and OE2 of Glu138 in *Bp*Xyn30A and Glu148 in *Rc*Xyn30A, other two coordinating the O2 Xyl*p* by a glutamate OE2 (Glu227 in *Bp*Xyn30A and Glu247 in *Rc*Xyn30A) and an asparagine ND2 (Asn137 in *Bp*Xyn30A and Asn147 in *Rc*Xyn30A), and, finally, between the NE1 of a tryptophan residue (Trp84 in *Bp*Xyn30A and Trp96 in *Rc*Xyn30A) and the sugar O3 atom. Additionally, the xylopyranose unit engages in stacking interactions with the aromatic rings of Trp266 (in *Bp*Xyn30A) and Trp285 (in *Rc*Xyn30A) residues and is stabilized by hydrophobic interactions with Trp145 (in *Bp*Xyn30A) and Trp155 (in *Rc*Xyn30A) side chains (Trp147 in *Bs*Xyn30A and Trp143 in *At*Xyn30A). Tyr172 in *Dc*Xyn30A, that occupies the same position of Trp145 in *Bp*Xyn30A and Trp155 in *Rc*Xyn30A, is facing towards subsite -2a instead. At subsite -2a (Figures 15a, 15b, 16a and 16b), tryptophan and a tyrosine (Trp26 and Tyr267 in *Bp*Xyn30A; Trp37 and Tyr286 in *Rc*Xyn30A) side chains provide a nonspecific stacking with xylopyranose moiety, while the xylopyronose at subsite -3 is held in place only by a stacking interaction with Trp26 in *Bp*Xyn30A or Trp37 in *Rc*Xyn30A.

In the positive subsites (Figures 15c, 15 d, 16c and 16d), subsite +1 is predicted to hold the xylopyranose ring by two stacking interactions with opposite sided tyrosine residues (Tyr141 and Tyr202 in *Bp*Xyn30A; Tyr151 and Tyr224 in *Rc*Xyn30A), besides a hydrogen bond between OH of Tyr229 in *Bp*Xyn30A or Tyr249 in *Rc*Xyn30A, and carboxylate group of the enzyme's catalytic acid/base glutamate with atoms O3 and O4 of Xyl*p*, respectively. Subsite +2 interacts with the xylopyranose through hydrophobic interactions with a phenylalanine residue (Phe174 in *Bp*Xyn30A and Phe185 in *Rc*Xyn30A) and by a hydrogen bond of xylosyl O1 atom and Gln175 NE2 *Bp*Xyn30A or Gln186 NE2 in *Rc*Xyn30A. Lastly, the ultimate xylose unit in subsite +3 in *R. champanellensis* xylanase might be involved with stacking interactions with Trp194 aromatic side chain.

Finally, we analyzed which subsites of the xylanases active site, in addition to subsite -2 that requires a MeGlcA, could possibly accept GlcA decorations. For this purpose, we positioned a X3 and $MeGlcA_2X3$ in the enzymes aglycone and glycone sites, respectively. Coincidentally, the O2 atoms of xylopyronose moieties from subsites +2, +1 and -3 of *Bp*Xyn30A and *Rc*Xyn30A are facing outwards the protein surface (Figures 15e and 16e), suggesting the permissibility for decorations in these positions. Manual insertion of GlcA

(Figures 15f and 16f) led us to conclude that subsites +2, -2 and -3 of both glucuronoxylanases are able to accommodate glucuronic acid decorations, while the presence of such decoration in subsite +1 would result in steric hindrance with the appendant of subsite -2b.

Given that ligand interacting amino acids alone could not explain the difference of *B. pumilus* and *R. champanellensis* xylanases cleavage pattern, i.e., the release of linear xylooligosaccharides by the former and not by the latter, we decided to look around the binding cleft for possible answers. Structural comparison with other prokaryotic GH30 with crystallographic structure deposited in PDB (Figure 17a) revealed that some enzymes present a β-fold after the loop L3 (β3-α3) (demarcated with a rectangle), as is the case of *Bp*Xyn30A, *Bs*Xyn30A and *At*Xyn30A, whereas *Rc*Xyn30A and *Dc*Xyn30A have a loop instead (Figure 17b). In *Bs*Xyn30A, this structure was said to be stabilized by an arginine side chain in loop 3 through stacking interaction with a tryptophan in β4-α4 loop, and a hydrogen bond between the carbonyl of the main chain of a lysine residue and a main chain nitrogen of a tryptophan. Said interactions were also observed in *Bp*Xyn30A between Arg96, Trp147 and Lys102 (Figure 17c). In *Rc*Xyn30A, however, besides a hydrogen bond between N Lys111, that occupies the same position as Lys202 in *Bp*Xyn30A, and the oxygen of Ala152, no other major interactions were observed (Figure 17d). All in all, we believe that the rigidity provided by loop 3 to the loop 4 of *Bp*Xyn30A might be responsible for maintaining the binding site more opened, which would allow initial substrate recognition beyond glucuronic acid decorations in subsite -2b.

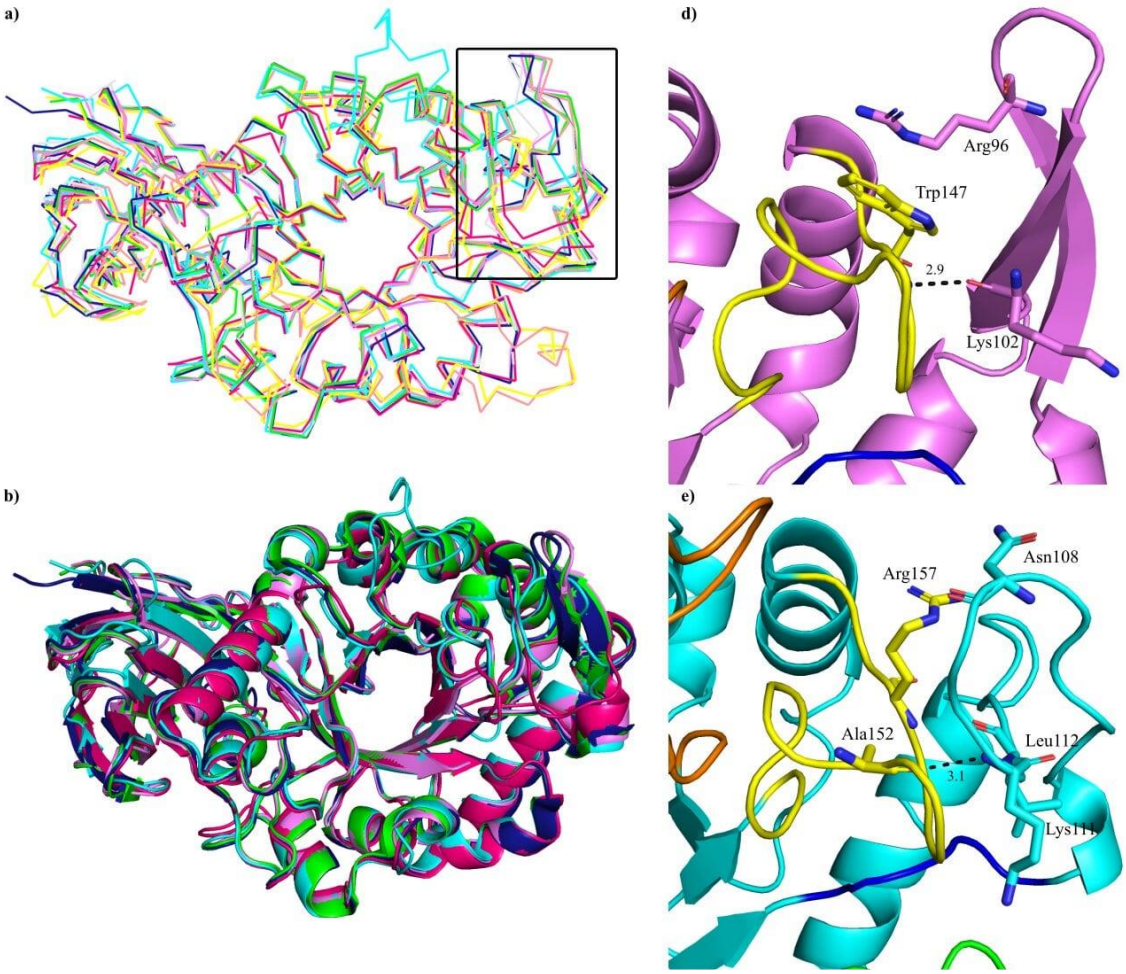

Figure 17 – Structural comparison of glucuronoxylanases from the present study with other prokaryotic GH30 with solved crystallography structure deposited in PDB. (a) Ribbon representation of the superposition of *Bp*Xyn30A (light purple), *Rc*Xyn30A (cyan), *Dc*Xyn30A (pink), *At*Xyn30A (dark blue), *Bs*Xyn30A (green), *Ca*Xyn30A (yellow), *Rp*Xyn30A (light pink), *Pb*Xyn30A (light grey). Region demarcated with a rectangle highlights the presence or absence of β-folding after loop β3-α3. (b) Cartoon representation of superposition of *Bp*Xyn30A (light purple), *Rc*Xyn30A (cyan), *Dc*Xyn30A (pink), *At*Xyn30A (dark blue) and *Bs*Xyn30A (green). Amino acids interactions between loop 4 (yellow, β4-α4) and region after loop 3 (blue, β3-α3) of (c) *Bp*Xyn30A and (d) *Rc*Xyn30A. Hydrogen bonds are marked with black dashed lines and distances are given in Å.
Source: By the author

5 Conclusion

Here, we described the biochemical characterization of *Bp*Xyn30A and *Rc*Xyn30A xylanases and their high-resolution crystallographic structure with 2.16 Å and 1.37Å of resolution, respectively, using synchrotron radiation. Both enzymes are active in mild conditions (neutral pH and temperature close to the human body), with specific activities against glucuronoxylan as 13.5 ± 0.7 U/mg for the former and 125.6 ± 5.4 U/mg for the latter. CBM9 domain removal resulted in a decrease of 20% of *Rc*Xyn30A activity, emphasizing its importance in the context of the enzyme's microorganism of origin.

MALDI-TOF and HPAEC-PAD analysis showed that while *Bp*Xyn30A is able to cleave unbranched oligosaccharides, *Rc*Xyn30 is highly inefficient at it. As a result, *B. pumilus* xylanase chromatograms contained both linear XOS and MeGlcA monosubstituted ones in the last but one xylose moiety towards the reducing end, whereas only branched xylooligosaccharides were detected in *R. champanellensis* cleavage pattern assays.

*Bp*Xyn30A and *Rc*Xyn30A, like other classical GH30_8 xylanases, harbor conserved amino acid residues in subsite -2b that strongly bind to MeGlcA, via ionic and hydrogen bonds. The decoration displays a crucial role in substrate recognition, which explains the lack of enzyme activity against arabinose substituted xylan. However, it does not elucidate the absence of linear xylooligosaccharides in *Rc*Xyn30A product profiles. We hypothesized that the presence of a β-fold after the loop L3 in *B. pumilus* xylanase acts stabilizing loop 4, thus limiting its movements, that is responsible for several stacking and hydrogen bonds interactions with substrate. Moreover, the presence of a larger β5-α5 loop in *Rc*Xyn30A forms a possible third aglycone subsite (subsite +3), which has not been previously observed for the enzymes from GH30 family.

Finally, enzymatic hydrolysis of alkaline pretreated biomass revealed a time-dependent feature, with soluble products consistent with those observed for commercial glucuronoxylan. At the same time, *Bp*Xyn30A achieved higher conversion yields than *Rc*Xyn30, probably due to the requirement of GlcA decorations by *R. champanellensis*. We conclude that both glucuronoxylanases have potential for the production of decorated XOS and/or linear ones (in *Bp*Xyn30A case) from the lignocellulosic biomass, especially considering the low temperature reaction condition and the enzyme's thermostability. Nonetheless, new experiments to evaluate biomass saccharification by *Bp*Xyn30A and *Rc*Xyn30A in combination with other GHs, such as xylanases from families 10 and 11, are important to further assess their biotechnological applications.

REFERENCES

1 IPCC. **Global Warming of 1.5°C:** IPCC special report on impacts of global warming of 1.5°C above pre-industrial levels in context of strengthening response to climate change, sustainable development, and efforts to eradicate poverty. Cambridge: Cambridge University Press, 2022. DOI: 10.1017/9781009157940.

2 ROUGH years ahead. **Nature Climate Change**, v. 13, n. 7, p. 589–589, 2023. DOI: 10.1038/s41558-023-01745-x.

3 ANTAR, M. *et al*. Biomass for a sustainable bioeconomy: an overview of world biomass production and utilization. **Renewable and Sustainable Energy Reviews**, v. 139, p. 110691, 2021. DOI: 10.1016/j.rser.2020.110691.

4 YRJÄLÄ, K.; RAMAKRISHNAN, M.; SALO, E. Agricultural waste streams as resource in circular economy for biochar production towards carbon neutrality. **Current Opinion in Environmental Science & Health**, v. 26, p. 100339, 2022. DOI: 10.1016/j.coesh.2022.100339.

5 MEDINA, G. S. The economics of agribusiness in developing countries: areas of opportunities for a new development paradigm in the soybean supply chain in Brazil. **Frontiers in Sustainable Food Systems**, v. 6, 2022. DOI: 10.3389/fsufs.2022.842338.

6 PELKMANS, L. (ed.). Implementation of Bioenergy in Brazil - 2021 update. 2021. Avaiable from: https://www.ieabioenergy.com/wp-content/uploads/2021/11/CountryReport2021_Brazil_final.pdf. Acessible at: 2 Jan. 2024.

7 BELLOTE, A. F. J. *et al*. **Biomass and its participation in the Brazilan energy matrix**. 2021. Avaiable from: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/222448/1/Biomass-and-its-participation-in.pdf. Accessible at: 2 jan. 2024.

8 NARISETTY, V. *et al.* Valorization of renewable resources to functional oligosaccharides: Recent trends and future prospective. **Bioresource Technology**, v. 346, p. 126590, 2022. DOI: 10.1016/j.biortech.2021.126590.

9 JIMENEZ-LOPEZ, C. *et al.* Agriculture waste valorisation as a source of antioxidant phenolic compounds within a circular and sustainable bioeconomy. **Food & Function**, v. 11, n. 6, p. 4853–4877, 2020. DOI: 10.1039/D0FO00937G.

10 MOODLEY, P.; TROIS, C. Lignocellulosic biorefineries: the path forward. *In:* RAY, R. C. (ed.). **Sustainable biofuels:** opportunities and challenges. India: Academic Press, 2021. p. 21–42. DOI: 10.1016/B978-0-12-820297-5.00010-4.

11 MUJTABA, M. *et al.* Lignocellulosic biomass from agricultural waste to the circular economy: a review with focus on biofuels, biocomposites and bioplastics. **Journal of Cleaner Production**, v. 402, p. 136815, 2023. DOI: 10.1016/j.jclepro.2023.136815.

12 OKOLIE, J. A. *et al*. Chemistry and specialty industrial applications of lignocellulosic biomass. **Waste and Biomass Valorization**. v. 12, n. 5, p. 2145–2169, 2021. DOI: 10.1007/s12649-020-01123-0.

13 ZHOU, X.; BROADBELT, L. J.; VINU, R. Mechanistic understanding of thermochemical conversion of polymers and lignocellulosic biomass. *In:* GEEM, K. M. V. (ed.). **Advances in chemical engineering**. Belgium: Elsevier, 2016. p. 95–198. DOI: 10.1016/bs.ache.2016.09.002.

14 XIAO, C.; ANDERSON, C. T. Roles of pectin in biomass yield and processing for biofuels. **Frontiers in Plant Science**, v. 4, 2013. DOI: 10.3389/fpls.2013.00067.

15 DA SILVA, A. E. *et al.* Xylan, a promising hemicellulose for pharmaceutical use. *In:* VERBEEK, C. (ed.) **Products and applications of biopolymers.** New Zealand: InTech, 2012. p. 61-84. DOI: 10.5772/33070.

16 HUANG, L.-Z. *et al.* Recent developments and applications of hemicellulose from wheat straw: a review. **Frontiers in Bioengineering and Biotechnology**, v. 9, 2021. DOI: 10.3389/fbioe.2021.690773.

17 SANTIBÁÑEZ, L. *et al.* Xylooligosaccharides from lignocellulosic biomass: a comprehensive review. **Carbohydrate Polymers**, v. 251, p. 117118, 2021. DOI: 10.1016/j.carbpol.2020.117118.

18 ZOGHLAMI, A.; PAËS, G. Lignocellulosic biomass: understanding recalcitrance and predicting hydrolysis. **Frontiers in Chemistry**, v. 7, 2019. DOI: 10.3389/fchem.2019.00874.

19 PLANAS, A. Glycoside hydrolases: mechanisms, specificities, and engineering. *In:* GOYAL, A.; SHARMA, K. (ed.). **Glycoside hydrolases**. India: Elsevier, 2023. p. 25–53. DOI: 10.1016/B978-0-323-91805-3.00011-3.

20 HENRISSAT, B.; DAVIES, G. Structural and sequence-based classification of glycoside hydrolases. **Current Opinion in Structural Biology**, v. 7, n. 5, p. 637–644, 1997. DOI: 10.1016/S0959-440X(97)80072-3.

21 SIDAR, A. *et al.* Carbohydrate binding modules: diversity of domain architecture in amylases and cellulases from filamentous microorganisms. **Frontiers in Bioengineering and Biotechnology**, v. 8, 2020. DOI: 10.3389/fbioe.2020.00871.

22 COLLINS, T.; GERDAY, C.; FELLER, G. Xylanases, xylanase families and extremophilic xylanases. **FEMS Microbiology Reviews**, v. 29, n. 1, p. 3-23, 2005. DOI: 10.1016/j.femsre.2004.06.005.

23 PUCHART, V.; ŠUCHOVÁ, K.; BIELY, P. Xylanases of glycoside hydrolase family 30 – An overview. **Biotechnology Advances**, v. 47, p. 107704, 2021. DOI: 10.1016/j.biotechadv.2021.107704.

24 GHARECHAHI, J. *et al.* Lignocellulose degradation by rumen bacterial communities: new insights from metagenome analyses. **Environmental Research**, v. 229, p. 115925, 2023. DOI: 10.1016/j.envres.2023.115925.

25 ROGOWSKI, A. *et al.* Glycan complexity dictates microbial resource allocation in the large intestine. **Nature Communications**, v. 6, n. 1, p. 7481, 2015. DOI: 10.1038/ncomms8481.

26 GRONDIN, J. M. *et al.* Polysaccharide utilization loci: fueling microbial communities. **Journal of Bacteriology**, v. 199, n. 15, 2017. DOI: 10.1128/JB.00860-16.

27 OLSON, R. D. *et al*. Introducing the bacterial and viral bioinformatics resource center (BV-BRC): a resource combining PATRIC, IRD and ViPR. **Nucleic Acids Research**, v. 51, n. D1, p. D678–D689, 2023. DOI: 10.1093/nar/gkac1003.

28 DUARTE, M. *et al.* A dual cohesin–dockerin complex binding mode in Bacteroides cellulosolvens contributes to the size and complexity of its cellulosome. **Journal of Biological Chemistry**, v. 296, p. 100552, 2021. DOI: 10.1016/j.jbc.2021.100552.

29 MARCHLER-BAUER, A.; BRYANT, S. H. CD-Search: protein domain annotations on the fly. **Nucleic Acids Research**, v. 32, Suppl_2, p. W327–W331, 2004. DOI: 10.1093/nar/gkh454.

30 CHASSARD, C. *et al.* Ruminococcus champanellensis sp. nov., a cellulose-degrading bacterium from human gut microbiota. **International Journal of Systematic and Evolutionary Microbiology**, v. 62, n. 1, p. 138–143, 2012. DOI: 10.1099/ijs.0.027375-0.

31 BRANQUINHO, R. *et al*. Phylogenetic and clonality analysis of Bacillus pumilus isolates uncovered a highly heterogeneous population of different closely related species and clones. **FEMS Microbiology Ecology**, v. 90, n. 3, p. 689–698, 2014. DOI: 10.1111/1574-6941.12426.

32 LYU, Y. *et al.* Dose-dependent effects of dietary xylooligosaccharides supplementation on microbiota, fermentation and metabolism in healthy adult cats. **Molecules**, v. 25, n. 21, p. 5030, 2020. DOI: 10.3390/molecules25215030

33 KHANGWAL, I. *et al*. Understanding the xylooligosaccharides utilization mechanism of lactobacillus brevis and bifidobacterium adolescentis: proteins involved and their conformational stabilities for effectual binding. **Molecular Biotechnology**, v. 64, n. 1, p. 75–89, 2022. DOI: 10.1007/s12033-021-00392-x.

34 VACILOTTO, M. M. *et al.* Paludibacter propionicigenes GH10 xylanase as a tool for enzymatic xylooligosaccharides production from heteroxylans. **Carbohydrate Polymers**, v. 275, p. 118684, 2022. DOI: 10.1016/j.carbpol.2021.118684.

35 MARKETS AND MARKETS. **Prebiotic ingredients market size, share and industry leaders**. 2022. Avaiable from: https://www.marketsandmarkets.com/ResearchInsight/prebiotics-ingredients-market.asp. Accessible at: 8 Oct. 2023.

36 BOUICHE, C. *et al.* Differential antioxidant activity of glucuronoxylooligosaccharides (UXOS) and arabinoxylooligosaccharides (AXOS) produced by two novel xylanases. **International Journal of Biological Macromolecules**, v. 155, p. 1075–1083, 2020. DOI: 10.1016/j.ijbiomac.2019.11.073.

37 CAPETTI, C. C. M. *et al.* Recent advances in the enzymatic production and applications of xylooligosaccharides. **World Journal of Microbiology and Biotechnology**, v. 37, n. 10, p. 169, 2021. DOI: 10.1007/s11274-021-03139-7.

38 CARUGO, O.; DJINOVIĆ-CARUGO, K. Structural biology: a golden era. **PLOS Biology**, v. 21, n. 6, p. e3002187, 2023. DOI: 10.1371/journal.pbio.3002187.

39 GLISH, G. L.; VACHET, R. W. The basics of mass spectrometry in the twenty-first century. **Nature Reviews Drug Discovery**, v. 2, n. 2, p. 140–150, 2003. DOI: 10.1038/nrd1011.

40 YAMAGAKI, T. *et al.* Mechanism for odd-electron anion generation of dihydroxybenzoic acid isomers in matrix-assisted laser desorption/ionization mass spectrometry with density functional theory calculations. **Rapid Communications in Mass Spectrometry**, v. 30, n. 24, p. 2650-2654, 2016. DOI: 10.1002/rcm.7761.

41 HUXFORD, T. X-Ray crystallography. *In:* MALOY, S.; HUGHES, K. (ed.). **Brenner's encyclopedia of genetics**. New York: Elsevier, 2013. p. 366–368. DOI: 10.1016/B978-0-12-374984-0.01657-0.

42 LIU, L. *et al.* The Sirius project. **Journal of Synchrotron Radiation**, v. 21, n. 5, p. 904–911, 2014. DOI: 10.1107/S1600577514011928.

43 CHAYEN, N. E.; SARIDAKIS, E. Protein crystallization: from purified protein to diffraction-quality crystal. **Nature Methods**, v. 5, p. 147–153, 2008. DOI: 10.1038/nmeth.f.203.

44 MCPHERSON, A.; GAVIRA, J. A. Introduction to protein crystallization. **Acta Crystallography Section F -** structural biology communications, v. 70, p. 2-20, 2014. DOI: 10.1107/S2053230X13033141.

45 AMEH, E. S. A review of basic crystallography and x-ray diffraction applications. **International Journal of Advanced Manufacturing Technology**, v. 105, p. 3289–3302, 2019. DOI: 10.1007/s00170-019-04508-1.

46 EVANS, P.; MCCOY, A. An introduction to molecular replacement. **Acta Crystallography Section F -** structural biology communications, v. 64, p. 1-10, 2008. DOI: 10.1107/S0907444907051554.

47 MCCOY, A. J., SAMMITO, M. D.; READ, R. J. Implications of AlphaFold 2 for crystallographic phasing by molecular replacement. **Acta Crystallography Section F -** structural biology communications, v. 78, p. 1-13, 2022. DOI: 10.1107/S2059798321012122.

48 TEUFEL, F. *et al*. SignalP 6.0 predicts all five types of signal peptides using protein language models. **Nature Biotechnology**, v. 40, n. 7, p. 1023–1025, 2022. DOI: 10.1038/s41587-021-01156-3.

49 BERMAN, H. M. *et al*. The protein data bank Helen. **Nucleic Acids Research**, v. 28, n. 1, p. 235–242, 2000. DOI: 10.1093/nar/28.1.235.

50 THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, n. 22, p. 4673–4680, 1994. DOI: 10.1093/nar/22.22.4673.

51 TAMURA, K.; STECHER, G.; KUMAR, S. MEGA11: molecular evolutionary genetics analysis version 11. **Molecular Biology and Evolution**, v. 38, n. 7, p. 3022–3027, 2021. DOI: 10.1093/molbev/msab120.

52 SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular biology and evolution**, 1987. DOI: 10.1093/oxfordjournals.molbev.a040454.

53 CAMILO, C. M.; POLIKARPOV, I. High-throughput cloning, expression and purification of glycoside hydrolases using Ligation-Independent Cloning (LIC). **Protein Expression and Purification**, v. 99, p. 35–42, 2014. DOI: 10.1016/j.pep.2014.03.008.

54 MILLER, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. **Analytical Chemistry**, v. 31, n. 3, p. 426–428, 1959. DOI: 10.1021/ac60147a030.

55 LO, M.-C. *et al.* Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. **Analytical Biochemistry**, v. 332, n. 1, p. 153–159, 2004. DOI: 10.1016/j.ab.2004.04.031.

56 PANTOLIANO, M. W. *et al.* High-density miniaturized thermal shift assays as a general strategy for drug discovery. **Journal of Biomolecular Screening**, v. 6, n. 6, p. 429–440, 2001. DOI: 10.1177/108705710100600609.

57 CAPETTI, C. C. M. *et al.* Enzymatic production of xylooligosaccharides from corn cobs: assessment of two different pretreatment strategies. **Carbohydrate Polymers**, v. 299, p. 120174, 2023. DOI: 10.1016/j.carbpol.2022.120174.

58 COLUSSI, F. **Caracterização bioquímica, biofísica e biofísica da Celobiohidrolase I de Trichoderma harzianum envolvida na hidrólise da biomassa lignocelulósica**. Supervisor: Igor Polikarpov. 2012. 191 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2012.

59 BREHM, W. *et al.* XDSGUI : a graphical user interface for XDS, SHELX and ARCIMBOLDO. **Journal of Applied Crystallography**, v. 56, n. 5, p. 1585–1594, 2023. DOI: 10.1107/S1600576723007057.

60 MCCOY, A. J. *et al;* Phaser crystallographic software. **Journal of Applied Crystallography**, v: 40, n. 4, p.658-674, 2007. DOI: 10.1107/S0021889807021206.

61 JUMPER, J. *et al.* Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, n. 7873, p. 583–589, 2021. DOI: 10.1038/s41586-021-03819-2.

62 TERWILLIGER, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. **Acta Crystallographica Section D** - biological crystallography, v. 64, n. 1, p. 61–69, 2008. DOI: 10.1107/S090744490705024X.

63 AFONINE, P. V. *et al*. Towards automated crystallographic structure refinement with phenix.refine. **Acta Crystallographica Section D** - biological crystallography, v. 68, n. 4, p. 352–367, 2012. DOI: 10.1107/S0907444912001308.

64 EMSLEY, P. *et al.* Features and development of Coot. **Acta Crystallographica Section D** - biological crystallography, v. 66, n. 4, p. 486–501, 2010. DOI: 10.1107/S0907444910007493.

65 CHEN, V. B. *et al.* MolProbity : all-atom structure validation for macromolecular crystallography. **Acta Crystallographica Section D** - biological crystallography, v. 66, n. 1, p. 12–21, 2010. DOI: 10.1107/S0907444909042073.

66 ROBERT, X.; GOUET, P. Deciphering key features in protein structures with the new ENDscript server. **Nucleic Acids Research**, v. 42, n. W1, p. W320–W324, 2014. DOI: 10.1093/nar/gku316.

67 SUZUKI, T. *et al*. Cloning and expression of a 58-kDa xylanase VI gene (xynD) of Aeromonas caviae ME-1 in Escherichia coli which is not categorized as a family F or family G xylanase. **Journal of Fermentation and Bioengineering**, v. 84, n. 1, p. 86–89, 1997. DOI: 10.1016/S0922-338X(97)82792-4.

68 SAKKA, M. *et al*. Characterization of Xyn30A and Axh43A of Bacillus licheniformis SVD1 identified by its genomic analysis. **Enzyme and Microbial Technology**, v. 51, n. 4, p. 193–199, 2012. DOI: 10.1016/j.enzmictec.2012.06.003.

69 GALLARDO, O. *et al*. Characterization of a Family GH5 Xylanase with activity on neutral oligosaccharides and evaluation as a pulp bleaching aid. **Applied and Environmental Microbiology**, v. 76, n. 18, p. 6290–6294, 2010. DOI: 10.1128/AEM.00871-10.

70 ST. JOHN, F. J.; RICE, J. D.; PRESTON, J. F. Characterization of XynC from Bacillus subtilis subsp. subtilis Strain 168 and analysis of its role in depolymerization of glucuronoxylan. **Journal of Bacteriology**, v. 188, n. 24, p. 8617–8626, 2006. DOI: 10.1128/JB.01283-06.

71 ST JOHN, F. J. *et al.* Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase. **Journal of Molecular Biology**, v. 407, n. 1, p. 92-109, 2011. DOI: 10.1016/j.jmb.2011.01.010.

72 CAI, S. *et al.* Cellulosilyticum ruminicola, a newly described rumen bacterium that possesses redundant fibrolytic-protein-encoding genes and degrades lignocellulose with multiple carbohydrate- borne fibrolytic enzymes. **Applied and Environmental Microbiology**, v. 76, n. 12, p. 3818–3824, 2010. DOI: 10.1128/AEM.03124-09.

73 LIU, J. *et al.* Functional identification of two novel carbohydrate-binding modules of glucuronoxylanase CrXyl30 and their contribution to the lignocellulose saccharification. **Biotechnology for Biofuels and Bioproducts**, v. 16, n. 1, p. 40, 2023. DOI: 10.1186/s13068-023-02290-7.

74 ST JOHN, F. J. *et al.* A plasmid borne, functionally novel glycoside hydrolase family 30 subfamily 8 endoxylanase from solventogenic clostridium. **Biochemical Journal**, v. 475, n. 9, p. 1533–1551, 2018. DOI: 10.1042/BCJ20180050.

75 LARSON, S. B. *et al*. First crystallographic structure of a xylanase from glycoside hydrolase family 5: implications for catalysis. **Biochemistry**, v. 42, n. 28, p. 8411–8422, 2003. DOI: 10.1021/bi034144c.

76 ŠUCHOVÁ, K. *et al.* Glucuronoxylan recognition by GH 30 xylanases: a study with enzyme and substrate variants. **Archives of Biochemistry and Biophysics**, v. 643, p. 42–49, 2018. DOI: 10.1016/j.abb.2018.02.014.

77 RAKITIN, A. L.; ERMAKOVA, A. Y.; RAVIN, N. V. Novel endoxylanases of the moderately thermophilic polysaccharide-degrading bacterium melioribacter roseus. **Journal of Microbiology and Biotechnology**, v. 25, n. 9, p. 1476–1484, 2015. DOI: 10.4014/jmb.1501.01061.

78 VALENZUELA, S. V.; DIAZ, P.; PASTOR, F. I. J. Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module. **Applied and Environmental Microbiology**, v. 78, n. 11, p. 3923–3931, 2012. DOI: 10.1128/AEM.07932-11.

79 PADILHA, I. Q. M. *et al.* A glucuronoxylan-specific xylanase from a new Paenibacillus favisporus strain isolated from tropical soil of Brazil. **International Microbiology:** the official journal of the Spanish society for microbiology, v. 17, n. 3, p. 175–184, 2014. DOI: 10.2436/20.1501.01.220.

80 DUNG, N. V. *et al.* Purification and properties of β-1,4-Xylanases 2 and 3 from aeromonas caviae W-61. **Bioscience, Biotechnology, and Biochemistry**, v. 57, n. 10, p. 1708–1712, 1993. DOI: 10.1271/bbb.57.1708.

81 ST JOHN, F. J. *et al.* A novel member of glycoside hydrolase family 30 subfamily 8 with altered substrate specificity. **Acta Crystallographica Section D -** biological crystallography, v. 70, n. 11, p. 2950–2958, 2014. DOI: 10.1107/S1399004714019531.

82 MAEHARA, T. *et al.* GH30 Glucuronoxylan-specific xylanase from streptomyces turgidiscabies C56. **Applied and Environmental Microbiology**, v. 84, n. 4, 2018. DOI: 10.1128/AEM.01850-17.

83 VERMA, A. K.; GOYAL, A. A novel member of family 30 glycoside hydrolase subfamily glucuronoxylan endo-β-1,4-xylanase (CtXynGH30) from Clostridium thermocellum orchestrates catalysis on arabinose decorated xylans. **Journal of Molecular Catalysis B:** enzymatic, v. 129, p. 6–14, 2016. DOI: 10.1016/j.molcatb.2016.04.001.

84 FREIRE, F. *et al.* Conservation in the mechanism of glucuronoxylan hydrolysis revealed by the structure of glucuronoxylan xylanohydrolase (Ct Xyn30A) from Clostridium thermocellum. **Acta Crystallographica Section D -** structural biology, v. 72, n. 11, p. 1162–1173, 2016. DOI: 10.1107/S2059798316014376.

85 GASTEIGER, E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Research**, v. 31, n. 13, p. 3784–3788, 2003. DOI: 10.1093/nar/gkg563

86 SCHOMBURG, I. *et al.* The BRENDA enzyme information system–from a database to an expert system. **Journal of Biotechnology**, 2017. DOI: 10.1016/j.jbiotec.2017.04.020.

87 XIE, H. *et al.* Clostridium thermocellum Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding. **Biochemistry**, v. 40, n. 31, p. 9167–9176, 2001. DOI: 10.1021/bi0106742 .

88 LITTUNEN, K. *et al.* Effect of xylan structure on reactivity in graft copolymerization and subsequent binding to cellulose. **Biomacromolecules**, v. 16, n. 4, p. 1102–1111, 2015. DOI: 10.1021/bm501732b.

89 HURLBERT, J. C.; PRESTON, J. F. Functional characterization of a novel xylanase from a corn strain of Erwinia chrysanthemi. **Journal of Bacteriology**, v. 183, n. 6, p. 2093–2100, 2001. DOI: 10.1128/JB.183.6.2093-2100.2001.

90 VRŠANSKÁ, M. *et al.* Mode of action of glycoside hydrolase family 5 glucuronoxylan xylanohydrolase from Erwinia chrysanthemi. **FEBS Journal**, v. 274, n. 7, p. 1666–1677, 2007. DOI: 10.1111/j.1742-4658.2007.05710.x.

91 LI, Z. *et al.* Effects of mild alkali pretreatment and hydrogen-donating solvent on hydrothermal liquefaction of eucalyptus woodchips. **Energy & Fuels**, v. 29, n. 11, p. 7335–7342, 2015. DOI: 10.1021/acs.energyfuels.5b01625.

92 CHEN, X. *et al.* Total utilization of lignin and carbohydrates in Eucalyptus grandis: an integrated biorefinery strategy towards phenolics, levulinic acid, and furfural. **Biotechnology for Biofuels**, v. 13, n. 1, p. 2, 2020. DOI: 10.1186/s13068-019-1644-z.

93 LIMA, M. A. *et al.* Effects of pretreatment on morphology, chemical composition and enzymatic digestibility of eucalyptus bark: a potentially valuable source of fermentable sugars for biofuel production – part 1. **Biotechnology for Biofuels**, v. 6, n. 1, p. 75, 2013. DOI: 10.1186/1754-6834-6-75.

94 MAFEI, T. D. T. *et al*. Extraction and characterization of hemicellulose from eucalyptus by-product: assessment of enzymatic hydrolysis to produce xylooligosaccharides. **Applied Biochemistry and Biotechnology**, v. 190, n. 1, p. 197–217, 2020. DOI: 10.1007/s12010-019-03076-0.

95 PALUDZYSZYN FILHO, E.; PACHECO, A. R.; DITTMAR, H.; CORDEIRO, C. A. Estratégias para o melhoramento de eucaliptos tropicais na EMBRAPA. Colombo: EMBRAPA, 2004. Available from: https://ainfo.cnptia.embrapa.br/digital/bitstream/CNPF-2009-09/36786/1/doc99.pdf. Accessible at: 15 Jan. 2024.

96 LI, H.-Y. *et al.* Evaluation of the two-step treatment with ionic liquids and alkali for enhancing enzymatic hydrolysis of Eucalyptus: chemical and anatomical changes. **Biotechnology for Biofuels**, v. 9, n. 1, p. 166, 2016. DOI: 10.1186/s13068-016-0578-y.

97 RHEE, M. S. *et al.* Engineering the Xylan utilization system in bacillus subtilis for production of acidic xylooligosaccharides. **Applied and Environmental Microbiology**, v. 80, n. 3, p. 917–927, 2014. DOI: 10.1128/AEM.03246-13.

98 URBÁNIKOVÁ, Ľ. *et al.* Structural basis for substrate recognition by Erwinia chrysanthemi GH30 glucuronoxylanase. **FEBS Journal**, v. 278, n. 12, p. 2105–2116, 2011. DOI: 10.1111/j.1742-4658.2011.08127.x.