UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

PROTEIN ENGINEERING OF XYLANASE XYN10C FROM *PAENIBACILLUS BARCINONENSIS*

André Nunes de Campos

MESTRADO EM MICROBIOLOGIA APLICADA

2011

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

PROTEIN ENGINEERING OF XYLANASE XYN10C FROM *PAENIBACILLUS BARCINONENSIS*

Dissertação orientada por Prof. Javier Francisco Pastor (UB)
e Prof. Manuela Carolino (FCUL)

André Nunes de Campos

MESTRADO EM MICROBIOLOGIA APLICADA

2011

PROTEIN ENGINEERING OF XYLANASE XYN10C FROM *PAENIBACILLUS BARCINONENSIS*

André Nunes de Campos

MASTER THESIS

2011

This thesis was fully performed at the Microbiology Department of the Faculty of Biology of Universitat de Barcelona under the direct supervision of Prof. Dr. Javier Francisco Pastor.

Prof. Dr. Manuela Carolino was the internal designated supervisor in the scope of the Master in Applied Microbiology of the Faculty of Sciences of the University of Lisbon.

Acknowledgments

I would like to sincerely thank to Prof. Dr. Javier Francisco Pastor, member of the Coordinator Commission of the Advanced Microbiology Master studies of the Faculty of Biology, University of Barcelona, for his guidance, support and motivation during the year of research I've performed in his associated laboratory. Without his interpretation skills, the results here presented would have been misunderstood or simply unnoticed.

Abstract

Keywords: *Paenibacillus barcinonensis*; endoxylanase; modular enzymes; carbohydrate binding modules; CBM9b; GH10; arabinoxylan.

Paenibacillus barcinonensis Xyn10C is a modular xylanase composed of an N-terminal tandem CBM22 module, a catalytic domain belonging to GH10 and a C-terminal tandem CBM9 module. The enzyme was previously cloned and a preliminary characterization was performed. In our work several truncated enzymes, containing different combinations of the modules of the xylanase have been constructed, purified and characterized. The activity assays here performed show that the complete enzyme (wild type Xyn10C) and a truncated variant containing only the GH10 catalytic domain are highly active on arabinoxylans, commonly found in cereal hemicelluloses, while show lower activity on glucuronoxylans, typical of hardwood species. Additionally, the diminished activity of the isolated GH10 domain on some arabinoxylans indicate that the CBM22 and CBM9 ancillary tandem domains play a role for the degradation these substrates, probably facilitating hydrolysis to the less accessible insoluble fraction of arabixoxylans. This is strongly supported by the fact that *P. barcinonensis* was isolated from a rice field. The microcrystalline binding capacity of Xyn10C was also investigated. The results here obtained indicate that the enzyme has binding capacity to insoluble cellulose and that this capacity is mediated through CBM9b, the external domain of the CBM9 tandem, which has been already reported for other enzymes with similar architecture. Together these results point Xyn10C as a potential candidate for application in agro-industrial processes that involve the degradation of straw and cereal biomass, a frequent by product of agriculture, such as bioethanol production, providing at the same time more insights on the biology of modular carbohydratases and on their contribution to xylan degradation in natural habitats.

Resumo

A parede celular de células vegetais é uma estrutura rígida que permite que organismos multicelulares mantenham a sua forma e que possam crescer verticalmente. Esta parede celular tem uma composição complexa e variada, não só entre organismos, mas também consoante a idade e até entre tecidos do mesmo organismo. Não obstante, o seu maior componente é a celulose. Este polisacárido é composto por monómeros de glucose unidos por ligações glicosidicas β (1-4), criando uma estrutura filamentar. Estes polímeros individuais podem agrupar-se criando fibras de celulose. O segundo maior componente destas paredes é a hemicelulose. A cadeia principal deste polissacárido é estruturalmente semelhante à celulose no sentido em que é constituída por monossacáridos conectados por ligações β (1-4) mas neste caso é composta por diferentes açúcares sendo que a glucose é pouco frequente. De facto, o polímero hemicelulósico mais frequente é o xilano que tem a sua cadeia principal constituída por monómeros de xilose unidos por ligações β (1-4). Tal como outras hemiceluloses e contrariamente à celulose, os polímeros de xilano apresentam frequentemente ramificações constituídas por mono ou dissacáridos. A natureza destas ramificações está na base da nomenclatura de xilanos. Arabinoxilanos são polímeros de xilano que possuem uma frequência importante de ramificações compostas por arabinose unidas a uma xilose por meio de uma ligação α (1-3). Por outro lado, glucuronoxilanos possuem um número importante de ramificações de ácido D-glucurónico ou 4-*O*-metilglucurónico unidos por uma ligação α (1-2) ou mais raramente α (1-3). As cadeias de xilano provenientes de plantas gimnospérmicas superiores, também conhecidas por 'softwoods', possuem os dois tipos de ramificações e são por isso denominados de glucuroarabinoxilanos. Nas plantas angiospérmicas superiores, ou 'hardwoods', os glucuronoxilanos compõem a maioria dos polímeros de xilano e inclusive, a cadeia principal de xilose encontra-se acetilada em 60 a 70% dos monómeros. Em gramíneas, e particularmente nos cereais, a predominância é de arabinoxilano.

Diversos microrganismos, incluído bactérias e fungos, têm a capacidade de degradar polímeros de xilano enzimaticamente. Devido à sua complexidade, a sua hidrólise total é mediada por diversos enzimas. As endoxilanases são enzimas que hidrolizam a cadeia principal de xilano em locais aleatórios resultando na sua divisão. Por outro lado, enzimas como arabinofuranosidases ou glucuronosidases são responsáveis pela remoção das ramificações. Para estes microrganismos, a degradação de polímeros de xilano permite não só utilizar os oligossacáridos resultantes para o seu metabolismo mas também aumentar a disponibilidade dos polímeros de celulose. Esta propriedade está na base da aplicação de xilanases de origem microbiana em processos industriais, nomeadamente no fabrico de papel onde a sua actividade facilita o branqueamento das pastas de papel o que se reverte numa diminuição da utilização de compostos de cloro, nocivos para o ambiente. Os enzimas xilanoliticos também podem ser utilizados para a clarificação de sumos e vinhos e melhorar a consistência de cerveja e as qualidades nutricionais de ração animal, nomeadamente de aves.

Paenibacillus barcinonensis é uma bactéria Gram-positiva que foi isolada devido ao seu forte poder de degradação de xilano. O seu sistema xilanolítico parece ser complexo e é composto por diferentes enzimas. Três xilanases distintas foram já clonadas e caracterizadas, incluindo a endoxilanase C (Xyn10C), que constitui o objecto de estudo do presente trabalho. A análise da sua sequência proteica revelou que se trata de uma xilanase modular que para além do domínio catalítico, pertencente à família 10 das hidrolases glicosídicas (GH10), possuí ainda outros dois módulos de união a hidratos de carbono (CBM), cada um deles repetido em tandem. Do lado N-terminal do domínio GH10 encontra-se um módulo em tandem da família 22 (CBM22) e do lado C-terminal existe um módulo em tandem pertencente à família 9 (CBM9). Estas famílias compreendem membros que apresentam união a xilano e a celulose, respectivamente. A subdivisão de módulos tandem CBM9 foi ainda proposta em subfamília a) e b) em que membros da subfamília a) representam o módulo adjacente ao domínio catalítico e apresentam mutações em resíduos cruciais para a ligação ao substrato ao passo que membros da subfamília b) são os módulos externos de um tandem ou os módulos CBM9 que se apresentam isolados. Estes são capazes de se unir a celulose e a conservação das suas sequências proteicas admite maior previsibilidade quanto a esta capacidade para membros não caracterizados da subfamília. Consequentemente, a estrutura geral de Xyn10C pode ser esquematizada da seguinte forma: CBM22-CBM22-GH10-CBM9a-CBM9b.

Num estudo prévio, os extractos celulares de uma estirpe *Escherichia coli* recombinante que expressava Xyn10C constutivamente foram utilizados para fazer uma caracterização da sua actividade enzimática. O substrato para o qual a enzima exibiu mais actividade foi glucuronoxilano proveniente de bétula seguido por arabinoxilano de aveia. Não foram detectados níveis importantes de actividade sobre celulose amorfa ou cristalina. A capacidade de união a celulose microcristalina, sob o nome comercial de Avicel, também foi demonstrada e foi sugerido que seria mediada pelo tandem CBM9a-b.

A avaliação das funções de cada domínio individualmente, mas também de combinações seleccionadas, é um passo necessário para a compreensão do papel que desempenham em Xyn10C. Inclusivamente, este tipo de informação pode permitir, através de engenharia genética, desenhar enzimas com funções e qualidades específicas para determinado processo industrial, como o branqueamento de pastas de papel. Assim, a clonagem e caracterização de diferentes variantes truncadas de Xyn10C é crucial para perceber o seu verdadeiro potencial de aplicação industrial, o que também contribui para o conhecimento geral sobre enzimas modulares, e em particular, xilanases.

Neste trabalho, foram clonadas, produzidas e purificadas diferentes variantes de Xyn10C. A proteína completa (sem péptido sinal) e o domínio catalítico GH10 foram produzidos em fusão com uma cauda de histidina na sua extremidade amínica (N) ao passo que o módulo tandem CBM9a-b e os módulos GH10-CBM9a-b foram produzidos com uma cauda de histidina na extremidade carboxílica (C). Esta cauda de histidina permitiu a purificação das proteínas de interesse através da afinidade que apresenta para iões de níquel (Ni^{2+}). O módulo externo do tandem da família 9,

CBM9b, foi ainda clonado em fusão com um domínio GST na sua extremidade amínica (N), permitindo a purificação da proteína fusão graças à sua afinidade para moléculas de glutationa.

A actividade enzimática do enzima completo e do seu domínio catalítico GH10 foi determinada para 5 xilanos de origem diferente: glucuronoxilando de bétula (birchwood xylan - BIX) e de faia (beechwood xylan - BEX) e arabinoxilano de aveia (oat spelt xylan - OSX), centeio (rye xylan - RYX) e trigo (wheat xylan - WHX). As duas enzimas apresentaram um perfil de actividades específicas muito semelhante, ambas com maior actividade sobre RYX e menor sobre BIX. O único substrato que provocou diferenças significativas na actividade destes enzimas foi o WHX, sendo que a actividade do domínio catalítico GH10 foi menor. As actividades foram também medidas sobre as fracções solúveis e insoluveis de BEX e OSX de forma a compreender melhor o papel dos domínios de ligação a hidratos de carbono. Ambas as enzimas apresentaram actividades semelhantes entre BEX solúvel e insolúvel ao passo que a actividade sobre OSX insolúvel foi bastante menor que a actividade sobre OSX solúvel. Esta diminuição foi significantemente maior para o domínio catalítico GH10. Juntos estes resultados sugerem que a incorporação dos domínios tandem CBM22 e CMB9 ao domínio catalítico melhoram a sua capacidade de hidrólise para alguns arabinoxilanos ao passo que parece não influenciar a actividade sobre glucuronoxilanos.

A proteína completa Xyn10C, o tandem CBM9a-b e o domínio CBM9b, enquanto unido ao domínio GST (GST-CBM9b), foram submetidos a ensaios de união a celulose microcristalina (Avicel). Através de métodos qualitativos para a actividade de celulose, nomeadamente tratamento zimográfico de géis de acrilamida, verificou-se que Xyn10C apresenta pouca afinidade para celulose. Resultados semelhantes foram obtidos para CBM9a-b e CBM9b indicando que a afinidade da proteína é mediada pelo domínio CBM9b e que esta afinidade é baixa nas condições testadas.

No geral, este enzima apresenta maior preferência para arabinoxilanos. Corroborantemente, os domínios auxiliares de união a hidratos de carbono que possui promovem a actividade sobre certos arabinoxilanos e, em alguns casos, este efeito pode ser devido ao aumento da hidrólise da fracção insolúvel do substrato. Parece, portanto, que Xyn10C é um enzima que reflecte uma adaptação do organismo para o seu habitat, uma vez que foi isolado de um arrozal e o conteúdo de xilano da planta de arroz é maioritariamente arabinoxilano.

Table of Contents

Introduction

I.1. Plant cell wall

One of the main characteristics that distinguish most of the plant cells from the other type of cells occurring in nature is its extracellular matrix[47]. Plant cells have developed highly complex cell walls that allow multicellular organisms to develop hard tissues that build and hold its shape. Although it can be seen as a rigid structure, this matrix can be considered to be semi-permeable in some cases, in the sense that it allows the passage of small molecules like proteins and other nutrients, specially carbon dioxide and water[59].

The major component of the cell wall is cellulose, a structural polysaccharide made of glucose units linked together by β (1-4) glycosidic bonds. This polymer presents a linear conformation with protruding hydroxyl (OH) groups that allow the formation of microfibrils thanks to hydrogen bonds between two -OH from different chains, aligning them. Hemicelluloses are other polysaccharides that can be found in the plant cell wall. These compounds also present a β (1-4) bond between the pyranosyl residues but in this case the polymer is usually branched and the backbone structure is composed by other sugars apart from glucose such as xylose or mannose. Hemicelluloses interact non-covalently with cellulose fibers and are responsible for the cross-link between these and pectins, the third main type of polysaccharides. Pectins are a family of complex polysaccharides that contain α (1-4) linked D-galacturonic acid. The pectic polysaccharides can be classified into homogalacturonans, rhamnogalacturonans and substituted galacturonans according with their composition and function. Its solubilization is associated with fruit ripening[17].

Another chemical compounds that can be found in secondary wall is lignin, a major component on higher plants. Lignin is a hydrophobic amorphous polymer consisting of phenylpropane monomers arranged in extremely complex three dimensional structures. Its degree of polymerization is often hard to determine because of the irregularity of the polymer structure as this compound consists of random, connected substructures built with the monomer units[32].

In wood, lignin is found to be covalently connected to hemicelluloses, namely to the sugar ramifications of xylan compounds[29].

I.2. Hemicellulose

The term "hemicellulose" was introduced by Schuzle in 1982[53] to define the fractions of the plant material that could only be solubilized by aqueous alkali and not by hot water or chelating agents. These polysaccharides represent 20-40% of lignocellulosic biomass[40], the plant biomass composed of cellulose, hemicellulose and lignin that often is a by product of several industries. Hemicelluloses are found to be covalently linked to lignin and because they also interact with cellulose, they have a key role in maintaining wood integrity. This also creates a protective effect against enzymatic degradation due to the inaccessibility of the polymers which constitutes a major

drawback in the biotechnological usage of this raw material[60]. The cross-links between hemicelluloses and lignin that have been identified are diferulic acid bridges and ester linkage between lignin and glucuronic acid attached to xylan[19].

I.2.1. Hemicellulose composition

As opposed to cellulose that is considered to be of homogeneous constitution, hemicelluloses comprise several other different monomers that are not only hexoses (glucose, mannose, galactose) but also pentoses (xylose, arabinose) and other acids that can form short lateral branches on the polymer backbone. The composition of hemicellulose changes depending on the organism, its age and even the specification of the tissue. There are two main types of hemicellulose polysaccharides: glucomannans and glucuroxylans. In the former, the backbone is composed of β (1-4) linked D-glucopyranose and D-mannopyranose residues while in the latter is composed of β (1-4) linked D-xylopyranoses. These carbohydrates chains present several types of substituents like glucuronic acid or arabinofuranose that, according to its nature, can change drastically its proprieties. The designation of these polymers is also based on the frequency of the substituents.

Major differences in the hemicellulose polymers proportion are found between members of softwood (gymnosperms) and hardwood (angiosperms) plants. The main hemicelluloses present in softwood plants, such as conifers, are galactoglucomannans and arabinoglucuronoxylan while hardwoods, like members of the genus *Eucalyptus* that are largely used in the paper industry, have generally a greater proportion of glucuronoxylans[10]. Members of the family Poaceae, albeit being also angiosperms, differ significantly in the hemicellulose composition from other higher plants. They exhibit larger amounts of arabinoxylan and a unique hemicellulose compound, the cereal β-glucan, which is a linear homopolysaccharide composed of glucose units like cellulose but the β (1-4) glicosydic linked chain is frequently interrupted by β (1-3) linkages, creating bends in the linear polymer[39]. This applies to cereals like rice, wheat, rye, corn and barley that are on the basis of human and animal alimentation and are therefore largely cultivated.

I.3. Xylan

Xylan polysaccharides are considered to be the second most abundant polysaccharide on the planet since they represent the majority of hemicelluloses[45]. Even hemicelluloses from softwood species that also exhibit significant amounts of other type of polysaccharides, namely glucomannans, have high contents of xylan compounds. Xylan is a heteropolysaccharide with a backbone composed of D-xylopyranose linked by β (1-4) glycosidic bonds that carries a variable number of neutral or uronic monosaccharide subunits or short oligosaccharide chains. Xylan derived from hardwood plants is mainly branched by α (1-2) or more rarely by α (1-3) glycosidic bonds with 4-*O*-methyl-D-glucuronic acid or D-glucuronic acid on about 10% of the xylose residues. In addition, it presents approximately 60-70% of the xylose residues esterified with acetic

acid at the hydroxyl group of carbon 2 and/or 3. On the other hand, softwood xylan is mainly arabinoglucuronoxylan which in addition to the 4-*O*-methyl-D-glucuronic and the D-glucuronic acids, also presents α (1-3) linked L-arabinofuranose substitutions. Acetyl groups are only rarely found on softwood xylan. Finally, cereal species of the family Poaceae have arabinoxylan as the most prominent xylan polysaccharide in which substitutions are almost entirely made of L-arabinofuranose linked by α (1-3) linkage branches[58].

I.3.1. Xylan degrading enzymes

Many microorganisms such as fungus and bacteria possess the ability to degrade xylan compounds through the use of enzymes. Interestingly, this xylan lysis capacity is not restricted to any specific type of these microbes as it can be found not only in mesophiles organisms but also in termophiles or alkiliphiles. Other organisms, for example isolates from the rumen of animals, are able to ferment xylan under anaerobic conditions[60]. Xylanase activity at these conditions is particularly interesting for industrial application where the fermentations or processes are developed at high temperatures or alkaline conditions, such as the paper pulp bleaching process[11].

It is very frequent for a xylan degrading microorganism to produce more than one xylanase when grown on xylan supplemented medium. One of the most impressive cases is that of the fungal plant pathogen *Phanerochaete chrysosporium* that produces over 30 different xylan hydrolyzing enzymes[22]. Additionally, some bacteria are able to secrete a larger protein complex composed by several subunits with xylanolytic activity. These complexes were denominated by xylanosomes as an analogy to cellulosomes. In fact, cellulosomes are reported to present xylanase subunits and so it seems that this type of protein complex acts as a discrete "tailor-made" multienzyme complex. Therefore, the xylanolytic activity can be essential to facilitate cellulose availability to other cellulases in the multienzyme since xylan and cellulose are closely associated in the plant cell wall matrix. Moreover, the complex often presents the ability to bind to cellulose which greatly enhances its effectiveness to degrade lignocellulosic tissues and, if the complex is cell-associated, it promotes the proximity of the cell[48]. Xylanosome modulation has been observed in response to different nitrogen sources in the bacteria *Streptomyces olivaceoviridis*[20] 2005 and different carbon sources in *Paenibacillus curdlanolyticus*[62].

The complete hydrolysis of xylan polysaccharides requires different enzymes, which reflects its complexity (**Fig. I.1**). Endoxylanases (EC 3.2.1.8) are the enzymes capable of hydrolyzing glycosidic bonds of the polysaccharide chain splitting it at random locations. β-xylosidases (EC 3.2.1.37) are the second type of enzymes that hydrolyze xylosidic bonds, cleavig xylobiose and xylooligomers to xylose. The occurrence of true exoxylanases that, similarly to activity of exocellulases on cellulose, would cleave processively sugar dimmers from chain ends seems not to occur in xylan degradation as side chain would interfere the process. The side groups present in xylan are liberated by debranching enzymes. L-arabinofuranosidases (EC 3.2.1.55) are capable of hydrolyzing the terminal, non reducing α-L-arabinofuranosyl groups of arabinoxylans and have

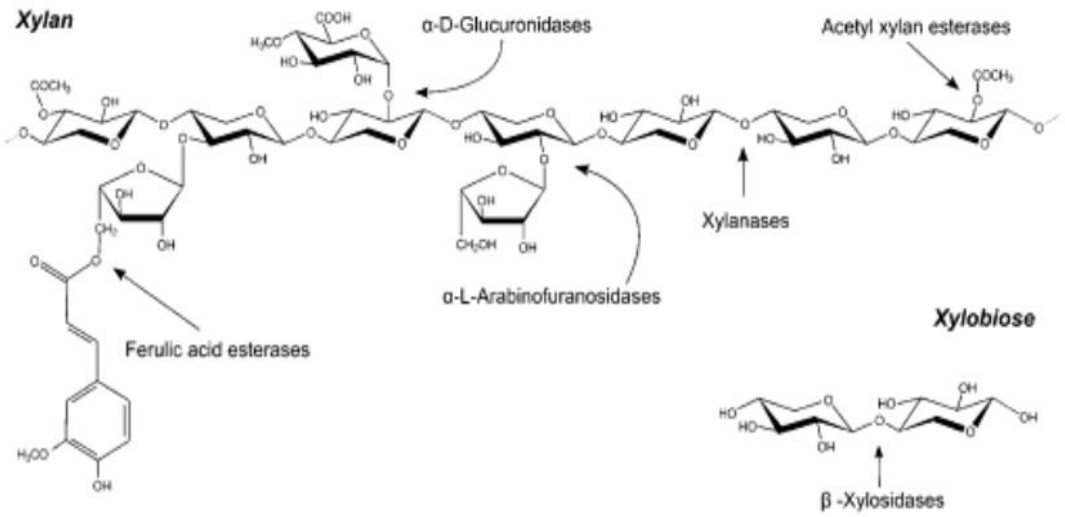

Figure I.1. Structure of xylan and the sites of attack by xylanolytic enzymes. The backbone of xylan chains is composed of β (1-4) linked xylopyranoses residues. This backbone can be variously substituted by side chains of arabinosyl, glucuronosyl, methylglucuronosyl, acetyl, feruloyl and p-coumaroyl residues. Hydrolysis of the xylan backbone is carried out by xylanases that hydrolyse internal linkages in xylan, and β-xylosidases that release xylose units from xylobiose and xylooligomers, while removal of xylan side cahains is catalysed by α-L-arabinofuranosidases, α-D-glucuronidases, acetyl xylan esterases, ferulic acid esterases and *p*-coumaric acid esterases. *Adapted from Pastor (2007)*[43].

been reported to be produced by many microorganisms including fungi, actinomycetes and other bacteria[49]. Xylan α (1-2) glucuronidases (EC 3.2.1.131) are responsible for the detachment of D-glucuronic and 4-*O*-methyl-D-glucuronic acids from the xylan main chain and have a very important role in wood degradation since the α (1-2) linkage between these and the xylose residues is very stable[33]. While the enzymes mentioned above are glycoside hydrolases, some carbohydrate esterases can also be involved in xylan degradation. Acetyl xylan esterases (EC 3.1.1.72) are capable of removing the acetyl groups from the xylose residues but other esterases like feruloyl esterases and *p*-coumaroyl esterases are thought to be responsible for the hydrolysis of the ester linkages involved in the covalent bond between xylan polysaccharides and lignin[46]. The combined usage of different xylanolytic enzymes, such as endoxylanases and debranching enzymes, produces a synergic effect in the xylan enzymatic degradation[18]. In this way, as the side chains are being detached, the main polymer becomes more vulnerable to endoxylanases and, at the same time, that endoxylanase activity allows debranching enzymes to act on substituents attached to easily accessible oligosaccharides rather than on the polymeric substrate.

I.4. Xylanases

I.4.1. Classification of xylanases

The first classification of microbial endoxylanases, based on its physicochemical proprieties like isoelectirc point and molecular mass, was able to distinguish two main groups[63]. Interestingly,

this classification was congruent with the one proposed years later for glycoside hydrolases that separated proteins into families according to its aminoacid sequences[30]. Family 10 glycoside hydrolases (GH10), formerly known as family F, are xylanases that present high molecular weight and low pI value. Family 11 (GH11), formerly referred to as family G, is composed of xylanases with low molecular weight and high pI values. These two families can also be differentiated by their substrate affinity. Contrarily to family 11, family 10 endoxylanases can attack glicosidic bonds next to a branch towards the non-reducing end. Members of family 10 require only two unsubstituted xylopyrosil residues while family 11 members require three consecutive xylopyranosyl residues. Moreover, some endoxylanses of family 10 have been show to have β-xylosidade activity, indicating that they may present several catalytic sites[3]. Recently several new xylanases have been classified in families 8, 5 and 30 of glycoside hydrolases families[16,26].

I.4.2. Structure of xylanases

The three-dimensional structures of members belonging to the two main families of xylanases, family 10 and 11, have been determined. Family 10 enzymes are generally large enzymes that present a $(\beta/\alpha)_8$ TIM-barrel architecture. The substrate binds to the shallow groove created on the bottom of this 'bowl'. On the other hand, GH11 family members are smaller sized well packed proteins with molecular organization mainly of b-sheet that fold into a motif known as the b jelly-roll fold. The active site is located at a cleft and is less accessible than the catalytic site of GH10 family members. The structural differences of these two families can be associated with their substrate preferences. Enzymes of GH10 family may show lower substrate specificity than members of family 11 because the catalytic site is located at a superficial position, allied to its possible greater flexibility due to its larger size[43].

I.4.3. Xylanases as modular enzymes

Xylanases, like cellulases and other carbohydratases, can be composed of a catalytic domain alone or be modular enzymes, presenting several other domains apart from the catalytic. These domains are able to fold and function independently and are connected by linker peptides that are flexible to some extent, as they don't show any secondary structure. The ancillary domains have been characterized to mediate the protein binding to carbohydrates such as cellulose or xylan, but have also been found to promote the binding to cellulosome scaffolding proteins. There is also some domains that appear conserved among several enzymes but which function is still unclear, such as the Fn3 domains or the SLH domains[48].

I.4.3.1. Carbohydrate binding modules

Several hydrolytic enzymes, and importantly glycoside hydrolases, have been shown to possess one or more carbohydrate binding modules (CBM). While these modules promote the polysaccharide hydrolysis by increasing the enzyme concentration at the point of attack, they are

not essential for the enzyme to function[8]. Initially they were designed by cellulose binding domains (CDB) and have been organized by types, I to XIII, based on their aminoacidic sequence. The discovery of other substrate affinities of these domains led to the renaming of the class and increase of the number of families[12].

Although members of the same family tend to bind to similar substrates, in some cases it is still hard to predict the functionalities based in the aminoacidic sequence alone. CelB9, a family 9 endoglucanase from *Paenibacillus barcinonensis*, the bacteria studied in the present work, is a modular enzyme and, besides the catalytic domain, it presents two family 3 carbohydrate binding module (CBM3) intercalated by a Fn3 domain at the C-terminal of the protein[14,44]. While the outermost CBM3, CBM3b, showed great affinity to microcrystalline cellulose (Avicel) and didn't contributed to the enzyme's activity, CBM3c revealed to play a critical role in the endoglucanase activity while it didn't display Avicel binding capacity[14]. This work clearly evidences the importance of these non catalytic domains in modular enzymes and how it is difficult to predict their function since two members of the same family appeared to have opposite functions.

The biotechnological application of carbohydrate binding domains is potentially vast. The family 9 carbohydrate binding domain of *Thermotoga maritima* Xyn10A, one of the most studied CBM, has been fused to other proteins and shown to be an effective tag for purification through cellulose binding[36]. Moreover, the binding of this module to *Ct*CDP, a cellodextrin phosphorylase from *Clostridium thermocellum*, have been shown to enhance the catalytic proprieties of the enzyme on regenerated amorphous cellulose[64].

I.5. Xylanase application

I.5.1. Xylanase application in paper industry

One of the most challenging setbacks in the paper industry is the separation of cellulose from the other components found in wood. Removal of residual lignin from Kraft pulp, the paper pulp produced by the kraft process, is a critical process to obtain pulp of very high brightness and brightness stability as its presence confers a characteristic brown colour. Usually, this is done through chemical treatment of the pulps using chlorine based chemicals. The effluent that are produced during the bleaching process, especially those following the chlorination stages, are the major contributors to waste water pollution from the pulp paper industry. The assistance of hemicelluloses, particularly xylanases, in this process often results in the reduction in chlorine consumption[61] and the application of xylanases together with safer bleach agents, such as oxygen and hydrogen peroxide has been extensively investigated creating projections of a totally chlorine-free pulp technology[54]. Due to its specific conditions, enzymes that are active at high temperatures and/or alkali conditions are particularly sought after for its potential application in such processes.

I.5.2. Xylanase application in other industries

The economically viable extraction of bioethanol, estimated to substitute 30% of the fuel used for transportation, and other bio-based chemicals from lignocellulosic biomass is challenged by the cost-effectiveness of lignocellulose fractionation techniques[35]. Xylanases can be used to disassemble this plant biomass since xylan is an important element to maintain the integrity of lignocellulosic biomass. This treatment allows the release of sugars that can be submitted to fermentation and biorefinery[27,35]. Therefore, research and development in this area is crucial for the economical success of lignocellulose biorefinaries.

Moreover, enzymatic xylan degradation has attracted much attention because of its practical applications in various agro-industrial processes, such as digestibility enhancement of animal feedstock, clarification of juices, and improvement in the consistency of beer[58].

I.6. Xylanase 10 C of *Paenibacillus barcinonensis*

I.6.1. *P. barcinonensis* characterization

Paenibacillus barcinonensis is a gram-positive, alkali tolerant, facultative anaerobic mesophillic bacterium. The bacterial species was characterized after the BP-23 strain, isolated from soil samples collected from a rice field in the delta of the Ebro river (Spain)[6]. The strain was allocated to the *Paenibacillus* genus by comparison of the 16S rRNA sequence though that data, along with phenotypic characterization, allowed the proposal for a new species[51]. BP-23 strain was selected among the colonies isolated from the sample for its high xylanolytic potential. Indeed, the strain possesses a multiple enzyme system for xylan degradation which allows its growth in a medium with xylan as the only carbon source. The most abundant xylanase produced by the strain when grown on xylans is xylanase A (Xyn10A), previously referred to as XynA. Its application in pulp bleaching was tested and it showed high effectiveness as a bleaching aid, reducing the dose of chlorine species necessary to achieve target brightness[7]. Other xylanase of this enzymatic system that has been cloned is xylanase B (Xyn10B). This xylanase presents special features that indicate that it's a cytoplasmatic enzyme. Additionally, it displays transglycosidase activity on oligosaccharides, suggesting that it can be potentially used for the synthesis of polymers[25]. Family 10 xylanase C, or Xyn10C, is yet another xylanase cloned from the BP-23 strain and is the main subject of the present work.

I.6.2. Family 10 xylanase C

The modular Xyn10C from P. barcinonensis was cloned and studied previously[5]. Conserved domain analysis showed that the catalytic domain, belonging to the GH10 family, is preceded by an N-terminal tandem domain, similar to others found in thermophilic microorganisms (**Fig I.2**). This kind of domain has been reported to increase the thermostability of enzymes, which was also observed for this endoxylanase[5]. Although these conserved domains were initially thought to be

thermostabilizing domains, this group was later categorized as family 22 carbohydrate binding module (CBM22, later known as X6), as several members of this family showed the ability to bind to xylan[13]. Inclusively, it was reported that in Xyn10B from *Clostridium thermocellum*, the thermostabilizing effect was due to the linker peptide between the CBM22 domain and the following catalytic domain and not to the CBM22 itself[21].

The GH10 catalytic site is also followed by a C-terminal family 9 carbohydrate binding tandem module (CBM9). Members of CBM9 were only described in xylanases and it is very common for this module to appear replicated in the C-terminal of the enzyme, especially in thermophilic and hyperthermophilic microorganisms. Division of CBM9 members in subclasses a) and b) was proposed[41]. Subclass a) CBM9 is composed of the inner domains of tandem CBM9 modules and lack carbohydrate binding ability due to mutations in structural and functionally important aminoacid residues. Subclass b) are the C-terminal domains of tandem CBM9 modules and CBM9 modules that appear alone. The conservation of important aminoacids throughout subfamily b) suggests that the carbohydrate binding function found in some characterized members is also conserved. Here, this designation is adopted for Xyn10C from *P. barcinonensis* and consequently the tandem will be addressed as CBM9a-b, while individually, the domains will be addressed as CBM9a and CBM9b.

In addition to these domains, the deduced aminoacid sequence of Xyn10C also shows an N-terminal region of 31 amino acids that resembles the typical *Bacillus* signal peptide (SP)[5]. Thus the complete architecture of Xyn10C can be described as follows, from the N to the C-terminal: SP-CBM22-CBM22-GH10-CBM9a-CMB9b

The activity of Xyn10C complete protein was previously studied. The substrate on which the enzyme was more active was birchwood glucuronoxylan followed by oat spelt arabinoxylan. Activity towards amorphous and microcrystalline cellulose was low. The ability to bind to microcrystalline cellulose was also demonstrated and suggested to be mediated through the CBM9a-b tandem.

To investigate the contribution of these domains to the catalytic activity of Xyn10C it is necessary to evaluate their function alone and in different combinations with each other. Inclusively, with that data, designing or optimizing the activity of the enzyme for a specific process, such as pulp bleaching, could be attempted. Besides, the complex structure of Xyn10C, with its several domains connected by flexible linkers, makes its crystallization difficult and it the case of obtaining crystals they would be of low resolution, as it has been seen in other enzymes of similar domain composition, thus compromising structural studies. It is therefore of high value the cloning and characterization of different variants of Xyn10C xylanase through genetic engineering to fully unravel its application potential and improve the general knowledge on modular enzymes, particularly xylanases.

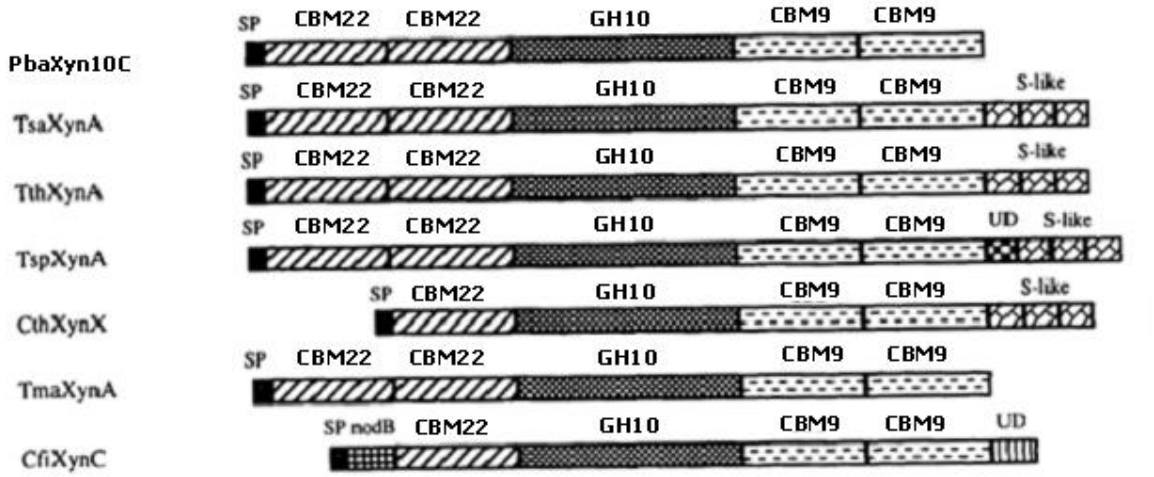

Figure I.2. Modular structure of Xyn10C and other homologous endoxylanases. Structures of *Paenibacillus barcinonensis* XynC (BspXynC), *Thermoanaerobacter saccharolyticum* XynA (TsaXynA), *Thermoanaerobacter thermosulfurigenes* XynA (TthXynA), *Thermoanaerobacter* JW/SL-YS 485 XynA (TspXynA), *Closteridium thermocellum* XynX (CthXynX), *Thermotoga maritima* XynA (TmaXynA) and *Cellulomonas fimi* XynC (CfiXynC). Boxes representing domains with similar function are filled with the same pattern. SP, signal peptide; CBM22, family 22 carbohydrate binding module; GH10, glycosyl hydrolase family 10 catalytic domain; CBM9, carbohydrate binding module; S-like, S-layer-like domain; nods, domain homologous to NodB proteins of Rhizobium spp.; UD, domain with unknownfunction.

Materials and Methods

M.1. Culture media and specific substances

M.1.1. Preparation of culture media

The nutrition media used for growth, screening, maintenance and storage of the different parental strains and recombinant clones was Luria-Bertani Broth[50]. Luria-Bertani (LB) broth is a rich medium suitable for most chemoheterotrophic microorganisms. This medium was sterilized by heat at 121°C for 20 min in autoclave. The LB medium composition used in this work was 10g bacterial tryptone, 5g yeast extract and 5g of NaCl per liter of distilled water. For LB-agar, 15g/L of agar were added.

M.1.2. Preparation of antibiotics

When necessary, sterilized culture media were supplemented with concentrated solutions of different antibiotics as seen in **Table M.1**. All antibiotic stock solutions were filtered through 22µm pore filters prior to storage and were always used in sterile conditions.

Table M.1. Antibiotic solutions used in the present work.

Antibiotic	Mode of action	Working concentration	Stock solution
Ampicillin (Amp)	Stops cell wall synthesis by inhibiting formation of the peptidoglycan cross-link	100 µg·mL-1	100 mg·mL-1 in H2O Stored at -20°C
Kanamycin (Kan)	Causes misreading of mRNA by binding to 70S ribosomes	50 µg·mL-1	100 mg·mL-1 in H2O Stored at -20°C
Tetracycline (Tet)	Inhibits protein synthesis by preventing binding of the amino-acyl tRNA to the ribosome A site	10 µg·mL-1 in liquid culture 12.5 µg·mL-1 in plates	12.5 mg·mL-1 in ethanol Stored protected from light at -20°C

M.1.3. Preparation of isopropyl-β-D-thiogalactoside

To promote the expression of the protein of interest by inducible strains, LB medium was supplemented with isopropyl-β-D-thiogalactoside (IPTG). Working solutions were prepared at a concentration of 1M (238.3 mg/mL) in bidistilled water, sterilized by 22µm filtration, and stored at -20°C. The final concentration used for optimal protein overexpression varied between 0.1 to 1mM, depending on the strain.

M.2. Microbial Strains

The parental strain from which all other recombinant clones were constructed was *Escherichia coli* 5K/pX31[5]. This strain contains the *P. barcinonensis* xylanase Xyn10C gene cloned in the pBR322 plasmid that was used as a template for the DNA amplification reactions that allowed the construction of the other recombinant strains in this work. The parental *E. coli* strains used to create new recombinant strains are listed in table **M.2**. The recombinant strains used and constructed in this work for protein expression are listed in table **M.3**.

Table M.2. *E. coli* strains used for cloning in the present work.

Strain	Genotype	Supplier
E. coli DH5α	F- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(rk^-, mk^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ-	Invitrogen
E. coli BLR (DE3)	F– *ompT* $hsdS_B$(r_B^-, m_B^-) *gal dcm* (DE3) Δ(srl-recA)306::Tn*10* (Tet^R)	EMD Chemicals
E. coli BL21 (DE3)	F– *ompT* $hsdS_B$(r_B^-, m_B^-) *gal dcm rne131* (DE3)	Invitrogen
E. coli Origami2 (DE3)	*Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL* F'[lac^+ $lacI^q$ *pro*] (DE3) *gor522*::Tn*10* *trxB* (Str^R, Tet^R)	EMD Chemicals

Table M.3. *E. coli* recombinant clones constructed and used in the present work.

Strain[a]	Features	Reference
E. coli 5K/pX31	Contains ***Xyn10C*** complete gene	Blanco et al, 1999
E. coli DH5α/pET11a:Link-CBM9a-b	Contains the **CBM9a-b** domains	Unpublished
E. coli BLR(DE3)/pET11a:Link-CBM9a-b		This work
E. coli DH5α/pET101:CBM9a-b	Contains the **CBM9a-b** domains fused to a C-terminal his-tag motif	This work
E. coli BLR(DE3)/pET101:CBM9a-b		This work
E. coli BL21 Star(DE3)/ pET101:CBM9a-b		This work
E.coli Origami2(DE3)/pET101:CBM9a-b		This work
E. coli DH5α/pET28a:CBM9b	Contains the **CBM9b** domain fused to an N-terminal his-tag motif	This work
E. coli BLR(DE3)/pET28a:CBM9b		This work
E. coli BL21Star(DE3)/pET28a:CBM9b		This work
E.coli Origami2(DE3)/pET28a:CBM9b		This work
E.coli DH5α/pGEX-4T-2:CBM9b[b]	Contains the **CBM9b** domain fused to an N-terminal GST domain	This work
E.coli DH5α/pET101:GH10-CBM9a-b	Contains the **GH10-CBM9a-b** domains fused to an C-terminal his-tag motif	This work
E. coli BLR(DE3)/pET101:GH10-CBM9a-b		This work
E. coli BL21Star(DE3)/pET101:GH10-CBM9a-b		This work
E.coli Origami2(DE3)/pET101:GH10-CBM9a-b		This work
E.coli DH5α/pET28a:GH10	Contains the **GH10** domain fused to an N-terminal his-tag motif	This work
E. coli BLR(DE3)/pET28a:GH10		This work
E. coli BLR(DE3)/pET28a:Xyn10CΔSP	Contains *Xyn10C* gene without the signal peptide	This work

Strain[a]: Recombinant strains presenting the (DE3) genotype are able to overexpress the cloned protein upon IPTG induction; E.coli DH5α/pGEX-4T-2:CBM9b[b]: Although this strain doesn't present the (DE3) genotype, the cloned protein is also overexpressed upon IPTG induction because of the *tac* promoter present in the plasmid.

M.3. Molecular methods

M.3.1. Plasmids

The list of plasmids used in this work is shown in **Table M.4**.

Table M.4. *E. coli* plasmids used in the present work.

Plasmid	Features	Supplier
pET11a	Amp^R, T7 promoter, *lacI*	Novagen
pET101-d-TOPO	Amp^R, T7 promoter, C-terminal Hisx6 fusion tag, *lacI*	Invitrogen
pET28a	Kan^R, T7 promoter, N-terminal Hisx6 fusion tag, *lacI*	Novagen
pGEX-T4-2	Amp^R, *tac* promoter, N-terminal GST fusion tag, $lacI^q$	GE Healthcare

Amp^r: Ampicillin resistant; Kan^R: Kanamycin resistant.

M.3.1.1. Plasmid isolation for downstream application

For the rapid isolation of high and low copy number plasmid DNA, the UltraPrep Plasmid DNA Kit (Lab Box) was used according to the manufacturer's specifications. This kit uses a plasmid DNA purification protocol involving an alkaline lysis procedure and a silica-based membrane for plasmid DNA separation. The isolated high-quality plasmid DNA was subsequently submitted to agarose gel electrophoresis visualisation, cloning, restriction digestion, PCR amplification or DNA sequencing.

M.3.1.2. Rapid plasmid isolation for screening purposes

A variation of method of rapid plasmid isolation described previously[4] was performed in this work as a screening technique for positive recombinant clones. Clones carrying the religated or non digested plasmid and clones carrying the plasmid with insert are distinguished by plasmid separation on agarose gel without enzyme digestion after cell lysis. The clones that presented a DNA band with an apparent size bigger than that of a negative control were considered positive clones.

Sterile pipette tips were used to retrieve the transformant colonies from the selection plate into a new master plate containing the appropriate antibiotic. Then, they were left in 1.5 mL Eppendorf® tubes containing 40µL of lysis buffer (10 mM Tris buffer, 10 mM EDTA, 100 mM NaOH and 1% [wt/v] SDS) and were incubated for 15 min at room temperature. The tubes were added 3µL of HCl 1M and 5µL of Blue/Orange Loading Dye 6X (Promega) and they were left at -80ºC for 20 min. The tubes were then centrifuged using a conventional table top centrifuge at maximum speed for 5 min and 5-10 µL of the supernatant were analysed by agarose gel electrophoresis.

M.3.2. PCR techniques

In this work, DNA amplifications were performed by polymerase chain reaction (PCR)[50] using a GeneAMP PCR System 2400 thermocycler (Perkin Elmer).

M.3.2.1 Standard Polymerase Chain Reaction (PCR)

Standard PCR was performed using different thermostable DNA polymerases:

***Taq* polymerase (Biotools)**: This highly-processive DNA polymerase doesn't show 3'→5' exonuclease proof-reading activity. This enzyme was used for routine amplifications to check presence or absence of known inserts, as used in colony PCR screening, or to optimize PCR conditions.

KAPAHiFi™ polymerase (Kapa Biosystems): It presents higher processivity than *Taq* DNA polymerase achieved through engineering as opposed to domain aggregation. This enzyme was preferentially used to amplify DNA fragments for further cloning. It is capable of generating blunt ends DNA fragments, compatible with pET101/d-TOPO cloning.

***Pfu* polymerase (Biotools)**: This enzyme is less processive than *Taq* DNA polymerase but it was used for high-fidelity amplifications due to its 3'→5' exonuclease proof-reading activity. Used to amplify DNA fragments for further cloning as an alternative to KAPAHiFi™ polymerase, it can also generate blunt ends DNA fragments, compatible with pET101/d-TOPO cloning.

The pX31 plasmid[5] was used as a template for the amplification of the truncated forms of Xyn10C. The proprieties of each PCR amplified DNA sequences are specified in **Table M.5** and the primers used in each case are listed in **Table M.6**. The standard conditions and programs used for DNA amplification using the different polymerases were as seen in **Tables M.7** and **M.8**.

Table M.5. Properties of the PCR amplified DNA corresponding to Xyn10C derivatives.

Xyn10C derivatives	Plasmid	Expected size (bp)	Polymerase	Annealing Temperature[a] (ºC)
CBM9a-b	pET101/d-TOPO	1131	KAPAHiFi	60
CBM9b	pET28a	608	KAPAHiFi	65
	pGEX-T4-2	625	KAPAHiFi	65
GH10-CBM9a-b	pET101/d-TOPO	2154	KAPAHiFi	65
GH10	pET28a	1083	KAPAHiFi	Touchdown
Xyn10CΔSP	pET28a	3217	*Pfu*	Touchdown

Annealing temperature[a]: If amplification was performed by touchdown PCR, see **Section M.3.2.2.**

Table M.6. Primers used for PCR DNA amplifications of Xyn10C derivatives.

Xyn10C derivatives	Primers used for PCR reactions
CBM9a-b	Cloning into pET101/d-TOPO®
	FW Link CBM9 TOPO: 5′ **C ACC** ATG ACG TTA CCC GTA TAT C 3′
	BW CBM9-2 No Stop: 5′ TTT CTT CAT CAG CAG CAA ACT AC 3′
	(CACC motif required for TOPO cloning in bold, ATG start codon underlined)
CBM9b	Cloning into pET28a
	FW CBD2 pET: 5′ C TTT **CAT ATG** AAG CAT GCC AAG GTT ATG T 3′
	BW CBD2 pET: 5′ T GGC **GGA TCC** GAG CTT GAA CA 3′
	(*NdeI* and *BamHI* restriction sites in bold)
	Cloning into pGEX-T4-2
	FW CBD2 pGEX: 5′ A CTG **GGA TCC** GAC TTA AAG CAT G 3′
	BW CBD2 pGEX: 5′ T AGT **GTC GAC** ATG GCC GTT CAG A 3′
	(*BamHI* and *SalI* restriction sites in bold)
GH10-CBM9a-b	Cloning into pET101/d-TOPO®
	FW GH10 TOPO: 5′ **C ACC** ATG GAC TTG GCC AAG AAG CTA G 3′
	BW CBM9-2 No Stop: 5′ TTT CTT CAT CAG CAG CAA ACT AC 3′
	(CACC motif required for TOPO cloning in bold, ATG start codon underlined)
GH10	Cloning into pET28a
	FW GH10 pET: 5′ G ATT **CAT ATG** GAA AAA AAT ATT CCG G 3′
	BW GH10 pET: 5′ TGC **GGA TCC** *CTA* ATT TCG ATA TAC G 3′
	(*NdeI* and *BamHI* restriction sites in bold, ATG star codon underlined and complementary reverse TAG stop codon in italic)
Xyn10CΔSP	Cloning into pET28a
	FW XBD pET: 5' GGA GAA **CAT ATG** GCA AGC GCA GCG 3'
	BW CBD2 pET: 5′ T GGC **GGA TCC** GAG CTT GAA CA 3′
	(*NdeI* and *BamHI* restriction sites in bold)

Table M.7. PCR reaction mixture used with each polymerase.

Reagent	*Taq*	*Pfu*	KAPAHiFi™
dNTPs mix	0.2 mM	0.2 mM	0.3 mM
PCR buffer	1X	1X	1X
FW/BK primers	0.3 µM (each primer)	0.3 µM (each primer)	0.3 µM (each primer)
Template[a]	≤10ng	≤10ng	≤10ng
DNA polymerase	2 U	0.5 U	0.5 U
H_2O	Up to 25 µL	Up µL to 25 µL	Up µL to 25 µL

Template[a]: The quantity of plasmid DNA used varied from amplification to amplification within the given range to obtain best results. Sometimes the quantity of DNA used reached values as low as <1ng per 25µL of total reaction volume.
Note: For colony PCR technique see ***Section M.3.2.2****.*

Table M.8. Typical PCR amplification program.

Cycles	*Taq* polymerase	*Pfu* polymerase	KAPAHiFi™ polymerase
1 cycle	5 min at 95ºC	5 min at 95ºC	4 min at 95ºC
25-30 cycles	30 s at 95°C 30 s at Tm[a] 1 min/kb at 72°C	30 s at 95°C 30 s at Tm[a] 1 min/kb at 72°C	20 s at 98°C 15 s at Tm[a] 30s/kb at 72°C
1 cycle	6 min at 72ºC	6 min at 72ºC	6 min at 72ºC
∞	4ºC	4ºC	4ºC

Tm[a]: The annealing temperature depends on the melting temperature of the primers used in the reaction. For the annealing temperature used in each case, see **Table M.5**.
Note: For touchdown PCR or colony PCR techniques see ***Sections M.3.2.2.*** *and* ***M.3.2.3****.*

M.3.2.2. Touchdown PCR

Sometimes, when conventional PCR wasn't effective in amplifying the target region, touchdown PCR was performed[23]. This technique is particularly useful in cases where a suitable annealing temperature for the given set of primers is hard to determine. Briefly, a first set of replication cycles is programmed to start at 65ºC, an elevated annealing temperature, which decreases over the cycles until it reaches 50ºC. The higher the annealing temperature, the harder it is for the primers to bond to the DNA chain but the more specific the pairing becomes. Minding

this, the temperature gradient ensures that the target sequence to which the primers are the most specific is amplified first than any other possible unspecific amplification, and that is also amplified on the next cycles. A second set of replication cycles is run with annealing temperature at 50 ºC to exponentially increase the number of copies of the target sequence that was preferentially amplified before.

The touchdown PCR technique was carried out with the DNA polymerases mentioned above and the PCR reaction mix were the same as specified in **Table M.7**. The PCR program can be seen in **Table M.9**.

Table M.9. Touchdown PCR amplification program.

Cycles	*Taq* polymerase	*Pfu* polymerase	KAPAHiFi™ polymerase
1 cycle	5 min at 95ºC	5 min at 95ºC	5 min at 95ºC
16 cycles	20 s at 95°C 20 s at 65 to 50ºC 1 min/kb at 72°C	20 s at 95°C 20 s at 65 to 50ºC 1 min/kb at 72°C	20 s at 98°C 15 s at 65 to 50ºC 30 s/kb at 72°C
14 cycles	20 s at 95°C 20 s at 50ºC 1 min/kb at 72°C	20 s at 95°C 20 s at 50ºC 1 min/kb at 72°C	20 s at 98°C 15 s at 50ºC 30 s/kb at 72°C
1 cycle	10 min at 72ºC	10 min at 72ºC	6 min at 72ºC
∞	4ºC	4ºC	4ºC

M.3.2.3. Colony PCR

Colony PCR was used in this work to screen positive recombinant clones. In this case, the DNA used as a template for the PCR comes directly from bacterial cells harbouring the target sequence that are lysed by heat during the first cycle of the amplification program. In order to do so, a 50µL PCR mix was divided by 5 PCR tubes and a small quantity of bacterial cells grown overnight on an agar plate was added to each one with a pipette tip. The tip was removed after some seconds and used to inoculate a new master plate with the respective antibiotic for later selection of positives. The polymerase used in the colony PCR was *Taq* polymerase and the conditions and program used are specified in **Table M.10**. The primers used were the respective sequencing primers of the plasmid used in the construction and are listed on **Table M.12**.

Table M.10. Colony PCR reaction mixture and amplification program.

Reagent	Final concentration	Colony PCR program	
dNTPs mix	0.2 mM	1 cycle	10 min at 97ºC
PCR buffer	1X	30 cycles	1 min at 95ºC
FW/BK primers	0.3 uM each primer		1 min at 57ºC
Template	Bacterial cells		3 min at 72ºC
Taq polymerase	1 U	1 cycle	10 min at 72ºC
H_2O	X µL to 50 µL	∞	4ºC

M.3.2.4. PCR Sequencing

The nucleotide sequence of DNA fragments was obtained using the dideoxy-mediated chain termination method which consists on the synthesis of DNA molecules in the presence of fluorescently labelled dideoxynucleotides (dNTPs) that terminate the DNA synthesis[52]. The BigDye® Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems) and the analytical system CEQ™ 8000 (Beckman-Coulter) available at the Serveis Científico Tècnics of the Universitat de Barcelona were used. The reaction mixture and the amplification program used for DNA sequencing was performed according to manufacturer's specifications and is specified in **Table M.11**. The primers used for sequencing reactions were included in each vector kit by the respective supplier and are listed in **Table M.12**.

Table M.11. Conditions for PCR sequencing.

Reagent	Quantity	Colony PCR program	
BigDye®Terminator v 3.1 Cycle reaction premix[a]	1 µL	1 cycle	3 min at 94ºC
BigDye® buffer (10X)	2 µL	30 cycles	10 s at 96ºC
Primer 10 µM	0.5 µL		5 s at 55 ºC
			4 min at 60ºC
Template	150-300 ng	1 cycle	10 min at 72ºC
H_2O	X µL to 20 µL	∞	4ºC

Premix[a]: Contains thermostable polymerase, buffer, dNTPs and ddNTPs.

Table M.12. Primers used for PCR sequencing of DNA inserts in recombinant plasmids

Vector	Primers used for sequencing
pET101	Primer FW: T7 promoter (5'- TAA TAC GAC TCA CTA TAG GG -3') Primer BW: T7 terminator (5'- TAG TTA TTG CTC AGC GGT GG-3') Reaction: amplification of inserts cloned into pET101 polylinker
pET28a	Primer FW: T7 promoter (5'-TAA TAC GAC TCA CTA TAG GG-3') Primer BW: T7 terminator (5'- GCT AGT TAT TGC TCA GCG G-3') Reaction: amplification of inserts cloned into pET28 polylinker
pGEX-T4-2	Primer FW: 5'pGEX-T4-2 Sequencing Primer (5'-GGG CTG GCA AGC CAC GTT TGG TG-3') Primer BK: 3'pGEX-T4-2 Sequencing Primer (5'-CCG GGA GCT GCA TGT GTC AGA GG-3') Reaction: amplification of inserts cloned into pGEX-T4-2 polylinker

M.4. Nucleic acid manipulation

M.4.1. Gel electrophoresis

Plasmid DNA preparations, PCR products, digested DNA and other DNA samples were analyzed by agarose gel electrophoresis following the specifications described previously[50].

DNA samples were prepared for electrophoresis by adding Blue/Orange Loading Dye (Promega) 6X to a 1X final concentration. They were loaded into an agarose gel 0.8% in TAE buffer 1X (40 mM Tris buffer, 40 mM acetic acid and 1 mM EDTA). Electrophoresis was performed in a horizontal Mini-Sub®Cell GT tray or Wide Mini-Sub®Cell GT tray for bigger gels (Bio-Rad) using a voltage of 100 volts generated by a PowerPac™ Basic power supply (Bio-Rad).

M.4.1.2. Nucleic acid staining, visualization and size estimation

DNA molecules separated in agarose gels were visualized by gel staining for 15-30 min in distilled water containing 0.75 $\mu g \cdot mL^{-1}$ ethidium bromide (Sigma-Aldrich) or 0.75 $\mu g \cdot mL^{-1}$ GelRed™ (Biotarget). Alternatively, after cooling down to just before solidification, 25mL of gel solution were supplemented with 1µL of Nancy-520 5000x (Sigma-Aldrich) and run normally. The stained agarose gels were irradiated with UV-light (310 nm) using an Image Master® VDS (Pharmacia Biotech) for viewing the nucleic acid bands.

DNA sizes were estimated by comparison of their migration with respect to that of GeneRuler™ 1 kb DNA ladder (Fermentas) composed of fourteen chromatography-purified individual DNA fragments of sizes 250, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 5000, 6000, 8000 and 10000bp.

M.4.1.3. Isolation of DNA fragments

To get suitable DNA samples for cloning, sequencing, etc., a purification procedure was performed to remove enzymes, salts, DNA stains and other impurities from DNA samples or agarose gels. The commercial kit Ilustra GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare) was used for fast purification of DNA either from a solution or from electrophoresis separated DNA bands, following the manufacture's specifications. This kit contains a silica membrane assembly for binding DNA in high-salt buffer and elution was performed with water, if DNA was meant for sequencing, or with a low-salt buffer for other downstream applications. Because DNA purification from agarose gels shows lower yield, when appropriate, DNA was preferentially purified directly from the solution after checking the expected band number and size in agarose gel, if needed. Generally, digested plasmids were purified from the agarose matrix after electrophoresis to prevent contamination with non digested plasmids, since they migrate differently. In case the plasmid DNA was present in low quantities, purification was made directly from the mixture and the circular non recombinant plasmids were later linearized (**Section M.4.3.1.**).

M.4.2. Quantification of nucleic acids

Nucleic acid concentration and the degree of its purity were measured using the Spectrophotometer ND-100 NanoDrop®, which uses a sample retention technology that employs surface tension alone to hold the sample in place. It only needs a small amount of sample (1-2µL), which is pipetted onto the end of an optic fiber cable (the receiving line). Then, a second optic fiber cable (the source fiber) is brought into contact with the liquid sample causing the liquid to bridge the gap between the optic fiber ends, forming a liquid sample column. A pulsed xenon flash lamp provides the light source to measure the absorbance of the sample using a 0.2 mm path length. The sample concentration is based in OD_{260nm}, being the detection limit 2 ng/µL. The degree of purity of samples is assessed by the ratio of absorbance at 260 and 280 nm, being accepted ~1.8 for DNA samples.

M.4.3. Enzymatic treatment of DNA

All enzymatic treatments of nucleic acids were performed following the manufacturer's specifications.

M.4.3.1. Digestion with restriction enzymes

Double digestions were carried out using the most suitable buffer for both enzymes. The digestions were generally used to create adhesive ends on both the insert and the plasmid vector. The maximum quantity of DNA in the digesting reactions was 1000ng for PCR amplicons and 500ng for plasmid vectors. Incubation was at 37ºC for different time lapses. Sometimes, just before transforming *E. coli* cells with the result of a ligation reaction, the mixtures were incubated

for 1-2 hours with an enzyme that cleaves between the two restriction sites used for cloning to prevent contamination with circular non recombinant vector.

M.4.3.2. Alkaline phosphatase treatment

Alkaline phosphatase was used to remove the 5' phosphate groups from digested DNA fragments in order to avoid re-circularization of plasmid vectors in ligation reactions. Dephosphorylation reactions were carried out after one third of the digestion time had passed. To do so, 1U of alkaline phosphatase and 1X alkaline phosphatase buffer were added directly into 10-50 µL of restriction enzyme digestion of ~500ng of double stranded DNA of 4-6 kb.

M.4.4. Ligation of DNA molecules

Ligation reactions between vectors and insert molecules were performed with T4 DNA ligase, according to the manufacturer's specification (BioLabs). This enzyme catalyzes the formation of phosphodiester bonds between neighbouring 3'-hydroxyl- and 5'-phosphate ends in double-stranded DNA. The ligation reaction was performed at different vector:insert ratios from 1:0, used as a negative control, up to 1:12. The tool found at http://www.gibthon.org/ligate.html was employed to calculate the appropriate amounts of the components in the reaction to get the desired ratios. The amount of vector in the several ligation reactions was never superior of that used in the negative control. The ligation reactions were done at 16ºC for 16h (overnight).

When the amplified fragment was cloned into vector pET101/D-TOPO®, the specifications of the instruction's manual were followed.

M.4.5. Transformation of bacteria with exogenous DNA

Strains of *E. coli* were transformed using a simple and rapid variation of the technique described by Cohen *et al* (1972)[15]. Competent *E. coli* cells were prepared prior to transformation.

M.4.5.1. Preparation of **E. coli** *competent cells using calcium chloride*

E. coli cells were grown in LB supplemented with the appropriate antibiotic for 16 h at 37°C and 0,4 mL culture was transferred into 20 mL fresh LB and maintained under agitation at 37°C until the OD_{600nm} was approximately 0.5 (mid-exponential phase). Then, bacterial culture was divided by transferring 5 mL to 4 sterile, disposable, ice-cold 10 mL tubes and cooled down to 0°C by storing the tubes on ice for 10 min. After this incubation, cells were harvested at 3000 x *g* for 5 min at 4°C, and suspended in 10 mL of ice-cold 50 mM $CaCl_2$. The pellet was carefully resuspended by softly stroking the tube with the fingernail or a pen. After 20 min incubation on ice, the cell-harvesting and resuspension steps were repeated using 1330µL of ice-cold 50 mM $CaCl_2$ supplemented with 15% glycerol. This cell suspension was distributed into 150µL aliquots that were either directly used for transformation (as described below) or and stored at -80°C for later usage.

M.4.5.2. Transformation of E. coli competent cells

To transform $CaCl_2$-treated competent cells, 75 µL competent cell suspensions were transferred to a sterile 1.5 mL Eppendorf® tube and the DNA of interest was added (no more than 12.5 ng of DNA in a volume of 10 µL or less). The appropriate positive and negative controls were always included in the assays. The resulting mixtures were incubated in ice for 15 min, further heated at 42°C for 45 s without shaking, and rapidly transferred to ice. After 10 min, 500 µL LB medium were added to each tube and the mixture was incubated for 60-90 min at 37°C to allow cells to recover and express the antibiotic resistance marker encoded by the plasmid. After this incubation period, 50 µL of each sample were spread on LB agar plates supplemented with the proper antibiotic for selection of the transformed/recombinant cells. Plates were incubated at 37°C, expecting transformed colonies to appear in 12-16 h.

M.5. Growth conditions of induced recombinant strains

After confirming the DNA sequence and correct orientation of the insert contained in the plasmid of a recombinant clone, protein expression was essayed with an appropriate inducible strain. Recombinant strains were grown at 37ºC for 16h (overnight) in LB liquid media supplemented with the appropriate antibiotic. An half full Erlenmeyer flask with new culture media containing the same antibiotic was inoculated at an inoculum:culture volumes proportion of 1:20 with the overnight culture. When the cultures reached an OD_{600nm} of 0.5 to 0.8, roughly 2h30 to 3h later, protein expression was induced using IPTG at different concentrations ranging from 0,1 to 1 mM for 1 to 3 hours at 37º or 16h at 16ºC, depending on the recombinant strain. The optimum conditions were essayed in small volumes of 4-5mL and when the target protein was found to be overexpressed in the soluble fraction (**Section M.6.1.**), the same conditions were used in a scale-up to 500 or 1000 mL, depending on protein yield.

The properties of the truncated forms of Xyn10C that were successfully cloned and expressed as described above can be seen in **Table M.13**.

M.6. Protein techniques

M.6.1. Preparation of crude cell extracts

Preparation of cell extracts from *E. coli* recombinant strains was adapted from the protocol described by Ruiz et al., (2002). Generally, grown cell cultures were centrifuged at 4,000 x *g* for 10 min at 4°C and the supernatants discarded. In most cases, when the objective was only the visualization of the protein profile and, if suitable, enzymatic activity, pellets were resuspended in 50 mM sodium phosphate buffer (NaPi) buffer pH 7.4 to concentrate cells 10 times. If the protein extract was meant to further purification by high-performance liquid chromatography (HPLC), the

buffer in which the pellet was resuspended and the concentration might have been different, according with the suppliers indications.

Cell lysis was performed using two methods: for extracts up to 2 mL, lysis was carried out by sonication with the tube on ice in a Labsonic 1510 sonicator (B. Braun) performing 2 cycles of 0.9 s pulses at 50 watts for 1 min, separated by 15 s of repose to avoid excessive heating. In case of higher extract volumes, lysis was performed by 2 pressure-cycles at 500 PSIG using a cold French Pressure Cell Press SLM (AMINCO).

After cell lysis, samples were centrifuged at 10,000 x *g* for 15 min at 4°C to recover cleared cell lysates, which was considered to be cytoplasm fractions (crude cell extracts), whereas the final pellets were considered as the insoluble cell-debris fractions. If the objective was to determine whether the overexpressed protein is in the soluble or insoluble fraction of the lysate, the pellets were also resuspended in the same volume of 50 mM NaPi pH 7.4 as the volume used initially to resuspend the cells. The soluble and insoluble cell extracts could be stored at 4ºC for later analysis.

Table M.13. Properties of the cloned Xyn10C variants

Recombinant plasmid	Expected size (Da)	Xyn10C aminoacids	Features
pET101/CBM9a-b	45415.2	712 to 1086	C-terminal his-tag motif
pET28a/CBM9b	23450.2	888 to 1086	N-terminal his-tag motif
PGEX-T2-4/CBM9b	48185.1	886 to 1086	N-terminal GST domain
pET101/GH10-CBM9a-b	83784.6	370 to 1086	C-terminal his-tag motif
pET28a/GH10	42220.7	365 to 718	N-terminal his-tag motif
pET28a/Xyn10CΔSP	117378.7	29 to 1086	N-terminal his-tag motif

M.6.2. Protein electrophoresis

M.6.2.1. SDS-polyacrylamide gel electrophoresis

Analytical electrophoresis of proteins was carried out by polyacrylamide gel electrophoresis containing the strongly anionic detergent sodium dodecyl sulfate as a denaturing agent (SDS-PAGE), basically as described by Laemmli (1970)[38]. These conditions ensure dissociation of the proteins into their individual polypeptide subunits, minimizing aggregation. The denatured polypeptides bind SDS and become negatively charged, forming SDS-polypeptide complexes which migrate through polyacrylamide gels in accordance with the size of the polypeptide.

Protein electrophoresis was carried out using a Mini-Protean® Tetra Electrophoresis System (Bio-Rad), powered by PowerPac® Basic Power Supply (Bio-Rad). Acrylamide gels were hand cast in an appropriate apparatus included in the electrophoresis system, using 30% Acryl-Bisacrylamide Mix (Bio-Rad) and TEMED (Roth) in accordance with the suppliers' instructions. Acrylamide concentration of the resolving gel varied from 8% to 15% in order to obtain better definition from greater to lower molecular weight ranges, respectively. If enzymatic activity was to be evaluated by zymogram analysis, the resolving gel was further supplemented with 2% Birchwood xylan to a final concentration of 0.15%. Acrylamide concentration of stacking gel was 5%. The buffers and gels were prepared as stated in **Table M.14**.

Prior to electrophoresis, loading buffer 3X was added to all samples and they were usually heated for 5 min at 100°C in order to disaggregate proteins. Separation was performed in running buffer (**Table M.14**) at 200 volts through the stacking gel, and at 180 volts through the resolving gel. After SDS-PAGE separation, gels were directly stained in a Coomassie solution (**Section M.6.2.2.**) or subject to zymogram analysis as described below (**Section M.7.1.**). The molecular weight (MW) of protein bands was determined by comparison of their migration in SDS-PAGE gels with that of the commercial MW standards. The MW of the standards contained in each commercial product, expressed in kDa, are: Roti®-Mark Standard (Carl-Roth): 14.3 (lysozyme), 20 (trypsin inhibitor), 29 (carbonic anhydrase), 43 (ovalbumin), 66 (serum albumin), 118 (β-galactosidase), 212 (myosin); PageRuler™ Unstained Protein Ladder (Fermentas): 10, 15, 20, 25, 30, 40, 50 (denser), 60, 70, 85, 100, 120, 150, 200 (recombinant proteins).

M.6.2.2. Staining SDS polyacrylamide gels

The proteins separated by polyacrylamide gel electrophoresis were visualized by Coomassie Brilliant Blue staining (Bio-Rad). Once performed the separation, gels were kept for 1 h in a Coomassie solution and distained by keeping them in a distaining solution, both prepared as described in **Table M.15**. Because this chemical compound binds non-specifically to proteins but not to the gel, after distaining, the proteins can be seen as discreet blue bands within the translucent matrix of the gel.

M.6.3. Protein purification by column chromatography

Grown cell cultures of *E. coli* recombinant clones, induced to express the desired enzyme, were used as a source of protein for purification. The crude cell extract of these cultures was obtained as described above but buffer used for cell resuspension and concentration of the extract might have been different (Section **M.6.1.**). Protein purification was carried out using an ÄKTAFPLC™ apparatus designed for high-performance liquid chromatography (HPLC), in association with a Monitor UPC-900 and Frac-900 for fractionation of the samples. The software used to control and monitor the purification was UNICORN™ (GE Healthcare). This kind of system allows the purification of proteins by its size, ionic charge or affinity, depending on the chromatography column used, while fractioning the extract for subsequent analysis.

Table M.14. Gel and buffer compositions used in SDS-PAGE

Gel Components	Stacking Gel	Resolving Gel			
	5%	8%	10%	12%	14%
30% Acryl-Bisacrylamide	420 µL	1.6 mL	2.0 mL	2.4 mL	2.8 mL
Stacking/Resolving Buffer	625 µL	1.56 mL	1.56 m	1.56 mL	1.56 mL
Distiled water	1.4 mL	2.8 mL	2.4 mL	2 mL	1.6 mL
APS 10% (w/v)	17.5 µL	30 µL	30 µL	30 µL	30 µL
TEMED	5 µL	5 µL	5 µL	5 µL	5 µL

Buffer Solutions			
Stacking buffer	**Resolving buffer**	**Running buffer (pH 8.3)**	**Loading Buffer (3X)**
0.5 M Tris-HCl (pH 6.8) 0.4% SDS (w/v)	1.5 M Tris-HCl (pH 8.8) 0.4% SDS (w/v)	25 mM Tris 192 mM glycine 1% SDS (w/v)	62.5 mM Tris-HCl (pH 6.8) 10% glycerol (v/v) 2.3% SDS (w/v) 2.5% β-mercaptoethanol (v/v) 0.02% bromophenol blue (w/v)

Table M.15. Composition of Coomassie staining solution

Coomassie staining solution	Distaining solution
0.25% (w/v) Brilliant Blue Coomassie® R250 10% (v/v) acetic acid 45% (v/v) metanol	10% (v/v) acetic acid 45% (v/v) metanol

Several chromatography columns were used in this work and the suppliers' instructions were strictly followed for each one. Detailed information is described below. Before loading, samples were further clarified by centrifuging for 5 min at maximum speed on a table top centrifuge followed by 22µL pore filtration and buffers were filtered through 45µL pore filters prior to purification. The HPLC software calculates the protein concentration of each fraction by measuring the optical density with UV light at 280nm. After purification was complete, size and purity of the proteins in the collected fractions were checked by SDS-PAGE and fractions from several purification rounds of the same protein were pooled together and used for further analysis or were aliquoted in 1 mL samples, quickly frozen by tube immersion in -80ºC ethanol and stored at -80ºC.

M.6.3.1. Immobilized metal ion affinity chromatography

Immobilized metal ion affinity chromatography (IMAC) was carried out using 1 mL HisTrap™ HP columns (GE Healthcare) that permit purification of proteins containing a his-tag motif (six consecutive histidines) due to its affinity to Ni^{2+} ions embedded in the column's agarose matrix[34]. This his-tag can be fused to a protein of interest by cloning its DNA sequence in frame with the adjacent tag's sequence existent in vectors like pET28a or pET101, both used in this work. The bound proteins were eluted with an imidazole gradient.

The crude cell extracts were prepared as mentioned above (**Section M.6.4.**) but in this case the cells were concentrated 40 times in 20 mM NaPi pH 7.4, 500 mM NaCl binding buffer before lysis, according to the supplier's indications. After clarification of sample, 2 mL were loaded into the system and made pass through the column at 1 mL/min. Then the column was washed with 2 to 5 column volumes (CV) of binding buffer with 5% of elution buffer (20 mM NaPi pH 7.4, 500 mM NaCl and 500 mM imidazole [Roth]). This maintains an imidazole background that avoids unspecific binding to the Ni^{2+} ions in the matrix. The bound proteins were eluted with a elution buffer gradient. This chemical compound competes with the his-tag protein motif to the Ni^{2+} ions in the matrix such that when its concentration reaches a certain level, the bound protein is eluted into the collected fractions.

M.6.3.2. Glutathione Affinity Chromatography

Purification of proteins with affinity to glutathione was carried out using a 1 mL GSTrap™ FF column (GE Healthcare). Proteins cloned in the pGEX system, such as the CBM9b domain cloned in pGEX-T4-2, become fused to the *Schistosoma japonicum* glutathione-S-transferase (GST) domain. This naturally occurring 26kDa protein has the ability to bind to glutathione molecules embedded in the agarose matrix of the column[55]. After binding, the target protein is recovered through isocratic elution with a low concentrated reduced glutathione solution, preserving protein conformation and activity.

The protein extracts were obtained as described above (**Section M.6.1.**) from a 40 times concentrated cell suspension in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) binding buffer, following the supplier's instructions. After cellular lysis and clarification, 2 mL were loaded and made pass through the chromatography column at 0.2 mL/min. It is important to have a low flow rate in this step due to the relatively slow binding kinetics between GST and glutathione. The column was then washed with 5 CV of binding buffer at 1 mL/min. Then, isocratic elution was carried out at 1 mL/min with a 1:1 ratio of binding buffer and elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0) solution and 1 mL fractions were recollected. Elution takes place as the GST domain becomes bound to the reduced glutathione present in the elution buffer instead of the glutathione molecules of the column's matrix.

M.6.3.3. Ion exchange chromatography

Ion exchange chromatography was used in this work as a second step to obtain purified CBM9b from the GST-CBM9b purified extract after thrombin digestion (**Section M.6.4.**). The chromatography column employed was the Tricorn™ Mono Q™ 5/50 GL column (GE Healthcare), an anionic exchange column with high resolution, which allows separation of proteins depending on their electric charge.

Since the buffer of the protein extract after the first purification and during the thrombin digestion was not the advised for the ion exchange chromatography, the buffer of the protein extract had to be exchanged to the starting buffer (20 mM Tris-HCl, pH 8.0) suggested by the supplier (**Section M.6.6.**). Then, 0.5 mL of the extract was passed through the column with a flow rate of 2 mL/min. After washing the column with 5-10 CV of starting buffer, an increasing gradient of the elution buffer (20 mM Tris-HCl, 1.0 M NaCl, pH 8.0) was used to increase the ionic strength up to 0.5 M NaCl. During this gradient, different proteins are eluted at different NaCl concentration due to their different ionic charge and overall properties allowing distinct proteins to be analyzed separately, even with similar sizes.

M.6.4. Protein digestion

The GST-CBM9b protein, cloned with the pGEX-T4-2 plasmid, presented a thrombin recognition site between both domains. Thrombin digestion was performed to separate the domains and recover the CBM9b domain with a posterior purification step. The digestion was performed at room temperature for 16 h using the Thrombin Capture kit (Novagen), according to the supplier's instructions.

M.6.5. Protein concentration

When required, protein solutions were concentrated by centrifugation 5,000 x *g* and 4°C for the time needed by ultrafiltration using Amicon® Ultra-15 Centrifugal Filter Units (Millipore) with volumes up to 15 mL. The molecular weight exclusion size used was 30 kDa.

M.6.4.1. Protein concentration determination

To determine the concentration of solubilised proteins in samples, the Bio-Rad Protein Assay kit based on the Bradford method was used. This assay involves addition of an acidic dye to the protein solution and subsequent absorbance measurement at 595 nm with a spectrophotometer. Comparison to a standard curve provides a relative measure of protein concentration.

Due to the high amount of samples to be analyzed, the assay was performed with a microplate reader (Bio-Rad Model 3550), used according to the manufacturer's specifications. Two methods were employed for different protein concentration ranges. To generally quantify the protein concentration of purified samples, a standard curve of bovine serum albumin (BSA, Sigma) at concentrations ranging from 0.125 to 2.0 mg/mL was used as a reference. In this case, often

the protein extract was measured at different dilutions to obtain values within the range of the standard curve. On the contrary, to measure the protein concentration of fractions collected when performing the Avicel binding essay by the chromatography column method, a smaller scaled standard curve was used with BSA concentration ranging from 1.25 to 25 µg/mL.

M.6.6. Buffer exchange

When needed, protein samples were desalted and small molecular weight solutes were removed or exchanged by exchanging the buffer using PD-10 Desalting Columns (GE Healthcare) strictly following the manufacturer's indications. The kit can be used by the gravity or the centrifugation method. In this work, only the gravity method was employed which involves a final sample dilution of approximately 1.4 times.

M.7. Enzymatic assays

M.7.1. Qualitative analysis by zymogram

The xylanase activity was qualitatively measured by zymogram analysis of separated protein bands by SDS-PAGE. Often, this was done simultaneously with Coomassie staining for size comparison with the protein marker and/or zymographic staining of enzyme activity of the separated protein bands of the sample. In such cases, protein samples were loaded in duplicated lanes in gels previously supplemented with 0.15% birchwood xylan (**Section M.6.2.1.**), and, after the electrophoresis was performed, gels were split in two halves that were Coomassie stained or alternatively developed as zymograms.

The zymogram assay was carried out as previously described[5]. Immediately after performing the electrophoresis, gels were soaked for 30 min at room temperature (RT) in 2.5 % (v/v) Triton X-100® and washed in 20 mM TrisHCL buffer (pH 7.0) for 30 min to remove the SDS and for the protein to refold. The 20 mM Tris buffer was renewed and gels were incubated for 10-15 min at 55ºC for the enzyme to digest the xylan embedded in the gel. After this, the gels were immersed in a 0.1% (wt/v) Congo red aqueous solution for 15 min. The Congo red is a chemical that binds irreversibly to the carbohydrates in the Acrylamide matrix. The gels are then washed in 1M NaCl to remove the exceeding Congo Red. If the enzyme is able to digest xylan polymers, a clear band becomes visible in the red colored gel and its size can be compared to protein standards if Coomassie staining was done simultaneously. To preserve the zymogram result, the gels were quickly immersed in 10% (v/v) acetic acid, which reacts with the Congo Red chemical producing a dark precipitate, and were extensively washed with running water.

M.7.2. Quantitative analysis by reducing sugar quantification

The enzymatic activity of the complete Xyn10C and its GH10 catalytic domain was quantitatively measured by the Somogyi-Nelson method as described previously[57]. This method permits the quantification of the amount of reducing sugars, like the reducing end of a xylan polymer, thus allowing evaluation of endoxylanase activity, since new reducing ends are created with the hydrolysis of the polysaccharide. Therefore, the more active the enzyme is, the more reducing sugars will be detected. The reducing sugars, when heated with alkaline copper tartrate, reduce the copper from the cupric (Cu2+) to cuprous state (Cu+) and thus cuprous oxide is formed. When cuprous oxide is treated with arsenomolybdic acid, the reduction of molybdic acid to molybdenum blue takes place. The blue color developed can be compared with a set of standards by basic photometry to infer the quantity of reduced sugars.

The xylan degradation was performed in 100 µL volumes by mixing 33uL of 4.5% (wt/v) xylan substrate, 33uL of 150 mM acetate buffer pH 5.0 and 34uL of protein extract on a 1.5mL Eppendorf® tube. For each situation, triplicates and a control tube were done. The control was prepared without the protein extract, which was only added later in the essay. The tubes were incubated at 45ºC with 300 rpm for 30 min, after which 150 µL of ice-cold water were added to a final volume of 250 µL and the tubes were kept in ice for at least 5 min. In some situations, the protein extract used for the essay had to be previously diluted and/or the incubation time adjusted in order to obtain values within the calibration curve range. During ice incubation, the copper reagent was prepared by mixing Solution I and Solution II in a 4:1 proportion (**Table M.16)**. Then, 250 µL of this reagent was added to the essay tubes. The protein extract was subsequently added to the control tubes since the copper reagent inhibits enzyme activity. The tubes were vortexed and heated at 99ºC for 10 min. After the tubes cooled down on ice, 250 µL of a solution containing arsenomolibdate reagent (**Table M.16)** and H_2SO_4 1,5 N in a 1:2 proportion was added and the tubes were carefully vortexed to liberate the gas produced in the reaction. Finally, 750 µL of water was added to the tubes and these were centrifuged for 5 min at max speed in a conventional table top centrifuge. The content of the tubes was transferred to appropriate photometry wells and the optical density of the solutions was measured at 520 nm.

The standard curve was prepared by diluting known quantities of xylose, ranging from 3 to 48 ug, in 250 µL of water. Each standard measurement was performed in duplicates. Then, 250 µL of the copper reagent was added and the remainder of protocol was followed as described above. The curve was repeated a number of times during several days until consistent results were obtained. A correspondence between absorbance and the amount of reducing sugars present in the solution was established. This relation was used to determine the amount or reducing sugars released by the enzyme on a specific xylan substrate. The control measurement was subtracted from the average of the triplicates to remove interference from the protein extract and/or xylan substrates.

If the quantity of the enzyme extract was previously determined, the specific activity of the enzyme was expressed by U/µmol of protein. In this work, 1 U/µmol is defined as 1 µmol of reducing xylose sugar per minute per µmol of protein. For the activity measurements between the soluble or insoluble fractions of oat spelt and beechwood xylan (**Section M.7.2.2.**), the activity on the insoluble fraction was expressed relatively to the activity on the soluble fraction. Unpaired Student's T-test was applied to evaluate the significance of the differences observed.

Table M.16. Reagent composition and preparation for the Somogyi-Nelson assay

Solution I		**Solution II**		**Arsenomolibdate reagent**	
Potassium sodium tartrate	15 g	$CuSO_4.5H_2O$	5 g	$(NH_4)_6Mo_7O_{24}.4H_2O$	25 g
Na_2CO_3	30 g	Na_2SO_4	45 g	H_2SO_4 (96%)	21 mL
$NaHCO_3$	20 g			$Na_2HAsO_4.7H_2O$	3 g
Na_2SO_4	180 g				
Dissolve the potassium sodium tartrate and Na_2CO_3 by heating. Add $NaHCO_3$ and Na_2SO_4 already dissolved.		Dissolve the compounds in distilled water. Heat if necessary.		Dissolve the $(NH_4)_6Mo_7O_{24}.4H_2O$ in water then add H_2SO_4 (96%) and $Na_2HAsO_4.7H_2O$, previously dissolved in water. Incubate at 37 ºC for 48 h and store in an opaque bottle.	

M.7.2.1. Xylan substrates

Enzymatic activity was measured towards different xylan substrates derived from different plant organisms and therefore have distinct chemical compositions. Beechwood, birchwood and oat spelt xylan were purchased from Sigma-Aldrich, and rye and wheat xylan were purchased from Megazyme. The composition of this natural substrates was specified by the suppliers as follows. Xylose monomers were considered to represent 90% of the sugar composition of beechwood and birchwood glucuronoxylan. Since these plant organisms are hardwoods, the main chain of these xylan compounds was considered to be highly acetylated as well. Oat spelt arabinoxylan contained minimally 70% xylose and maximally 10% arabinose and 15% glucose. Wheat and rye arabinoxylans contained approximately 60% and 50% of xylose residues and 40% and 50% of arabinose residues, respectively.

M.7.2.2. Soluble and insoluble xylan

Enzymatic activity between the soluble and insoluble fractions of beechwood and oat spelt xylan was performed. To obtain the soluble fraction of the xylan substrates, a 4.5% aqueous solution was vigorously vortexed and the supernatant was carefully recovered after centrifugation at maximum speed using a conventional table top centrifuge. To obtain the insoluble fraction of xylan substrates, a modification of the alkali extraction method performed by Numm *et al* (1985)[42] was employed. Briefly, 1 g of the dry xylan was diluted in 20 mL of bidistilled water and 10 M

NaOH was added until the solution was at pH 10.0. The tube was incubated for 1 h while rotating vertically at 12 rpm. The solution was centrifuged at 3,000 x *g* for 5 min and the pellet was washed with water twice and once with acetate buffer pH 5.0. The pellet was finally resuspended in 10 mL of 96% (v/v) ethanol and filtered by a 45 µm pore filter. The recovered fraction was dried by dry heat at 125ºC for 3 h and weighted. The insoluble fraction of beechwood and oat spelt xylan was calculated to be 16.5% and 81.0% based on the initial and final weights of the substrates. The 4.5% (wt/v) solutions for the insoluble fractions used in the Somogyi-Nelson assay were prepared normally using the obtained insoluble xylan. On the other hand, for the soluble solutions, the dry substrate was mixed in water to a final concentration of 4.5% (wt/v) according to the determined proportions of soluble xylan of the susbtrate.

M.8. Binding assays

Proteins expressed in this work were evaluated in regard to their ability to bind to microcrystalline cellulose, under the commercial name of Avicel (Fluka). This was done by two different protocols: the centrifugation method and the column chromatography method.

M.8.1. Avicel binding by centrifugation method

This method was performed as described previously. Briefly, crude cell extracts were incubated with 5% (wt/v) Avicel in volumes no bigger than 500µL for 1 h at 4 ºC while rotating vertically at 12 rpm. The tubes were centrifuged at maximum speed with a conventional table top centrifuge and the pellets were sequentially washed with one volume of 50 mM NaPi pH 7.4 and one volume of 20 mM NaPi pH 7.4. Bound proteins were recovered by incubating the pellet with bidestilled water for one hour at room temperature and rotating vertically at 12 rpm and analyzed by SDS-PAGE.

M.8.2. Avicel binding by chromatography column method

For this method, a handmade Avicel column was employed. To do so, aluminum foil pieces were placed inside one Pasteur glass pipette in order to block the thin end of it. While in a vertical position, 2 mL of a 10% (wt/v) Avicel solution on 20 mM Tris-HCl were loaded into the column. The buffer should be able to pass through it while the Avicel remains retained by the aluminum. The column was equilibrated by 5-10 mL of the same buffer as the buffer of the extract. After equilibration, a sample of 500 µL of the extract was loaded and the eluted fraction was recovered. The column was then washed 4 times with 500 µL with the sample buffer and each fraction was recovered. Then, elution of the bound material was done by washing the column 10 times with 250 µL of water, while recovering each fraction. The protein concentration of each fraction was quantified using the small scaled Bradford method described above (**Section 6.5.1.**) and the fractions were analyzed by SDS-PAGE.

M.13. Protein crystallization

Pure protein samples of the GH10 catalytic domain of Xyn10C were sent to Rocasonalo Institute (Madrid) for three dimensional studies by crystallization methods.

Equal amounts of protein solution (2.5 mg/ml Xyn10C-GH10 in 20 mM Tris-HCl, pH 7) and precipitant buffer composed of 35% (v/v) 2-methyl-2,4-pentandiol (MPD), 17% (w/v) polyethylene glycol 3,350 (PEG 3350) in 100 mM Tris, pH 7 were mixed and vapor-equilibrated against a reservoir containing the latter solution. Crystallization experiments were carried out at room temperature and crystals were obtained after 2 to 3 days.

M.14. Computational analysis of DNA and protein sequences

The nucleotide and deduced amino acid sequences obtained in this work were analyzed for conserved motifs and homology by means of the complete non-redundant protein sequence database at NCBI through BLAST application (http://www.ncbi.nlm.nih.gov/blast/).

The software Contig Express licensed by Invitrogen was used to assemble DNA sequences.

The ProtParam tool found at ExPASy proteomics server was used to analyze the physico-chemical parameters (http://us.expasy.org/tools /protparam.html).

Results

R.1. Cloning, expression and purification of truncated forms of Xyn10C

R.1.1. CBM9a-b

The C-terminal CBM9 tandem domain of Xyn10C was previously cloned into pET11a plasmid vector and transformed into *E. coli* BLR(DE3) strain (unpublished data). In the present work, this *E. coli* BLR(DE3)/pET11a:LinkCBM9a-b recombinant strain was used as a first attempt to produce the CBM9a-b tandem. Several conditions and inducible strains were tested but no significant amounts of the target protein were observed in the crude cell extracts (results not shown).

This truncated enzyme was then cloned into the pET101/d-TOPO vector, in order to fuse it with a C-terminus his-tag motif. Just like in the pET11a construction, besides the CBM9 tandem, the 9 aminoacid link between the CBM9a domain and the preceding GH10 domain was included. Screening of the *E. coli* Top10 transformed cells was made by colony PCR (**Fig. R.1.A**) and positive clones were selected for plasmid sequencing to check the correct sequence and orientation of the insert. The plasmid isolated from one selected clone was used to further transform IPTG inducible *E. coli* strains. Best results were obtained with *E. coli* BLR(DE3)/pET101:CBM9a-b recombinant strain when induced with 1mM IPTG at 37 ºC for 1 h with 200 rpm. A larger culture was grown and induced under the same conditions and protein purification proceeded. Since the engineered enzyme produced by this strain had a C-terminus his-tag motif, protein purification was carried out using HisTrap chromatography column. However, it was observed that when the purification was performed under standard conditions, the protein was not able to bind efficiently to the column and was being collected in the washing fractions (**Fig. R.1.B**).

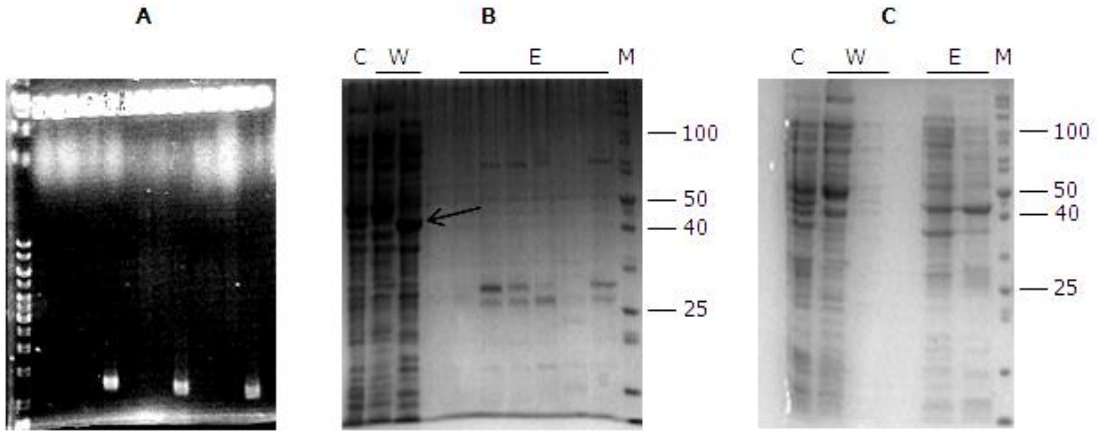

Figure R.1. Steps of CBM9a-b production cloned in the pET101 vector plasmid. A) Colony PCR of transformed *E. coli* Top10. Three positive clones are visible. B) IMAC purification performed under standard conditions. CBM9a-b (48 kDa) was washed in the unbound fraction (arrow). C) IMAC purification with the binding step performed without the standard 5% elution buffer. CBM9a-b is visible in both elution fractions collected. Lanes (B and C): C, crude cell extract; W, unbound fractions; E, imidazole eluted fractions; M, protein ladder wixh sizes (kDa).

The IMAC purification was repeated without the 5% elution buffer in the binding step, thus removing the imidazole background that prevents unspecific bindings to the column. As expected, the purified fractions obtained contained not only the target protein but also considerable amounts of other proteins (**Fig. R.1.C**). Due to time constrains, no optimization of the purification was performed and two of these IMAC fractions were used for microcrystalline cellulose binding assay.

R.1.2. CBM9b

Two cloning strategies were employed to overexpress the CBM9b domain of Xyn10C, the closest domain to the C-terminus of the protein. Since the link between this domain and the CBM9a domain is of only 2 aminoacids, a minimal number of upstream aminoacids from where the CBM9b domain hypothetically starts was included. This was made to prevent incorporation of possible functional motifs of the preceding domain to the CBM9b construction.

The first attempt to express CBM9b was to clone it into the pET28a vector. This vector also allows the incorporation of a his-tag motif but in the N-terminal end of the protein. Although a CBM9b expressing strain was obtained, the target protein was being poorly expressed in the soluble fraction of the extract. IMAC purification of this extract was attempted using HisTrap chromatography column but purification was unsuccessful (results not shown).

The production of the CMB9b domain was then attempted using the expression vector pGEX-T4-2 that allows the N-terminal fusion of the protein with GST. Screening of positive clones was performed by rapid plasmid isolation using *E. coli* DH5α cells (**Fig R.2.A**). After confirming the correct sequence and insert orientation by DNA sequencing, protein expression assays took place with the obtained *E. coli* DH5α/pGEX-T4-2:CBM9b recombinant strain since the plasmid alone is IPTG inducible thanks to its *tac* promoter located upstream of the cloned insert.

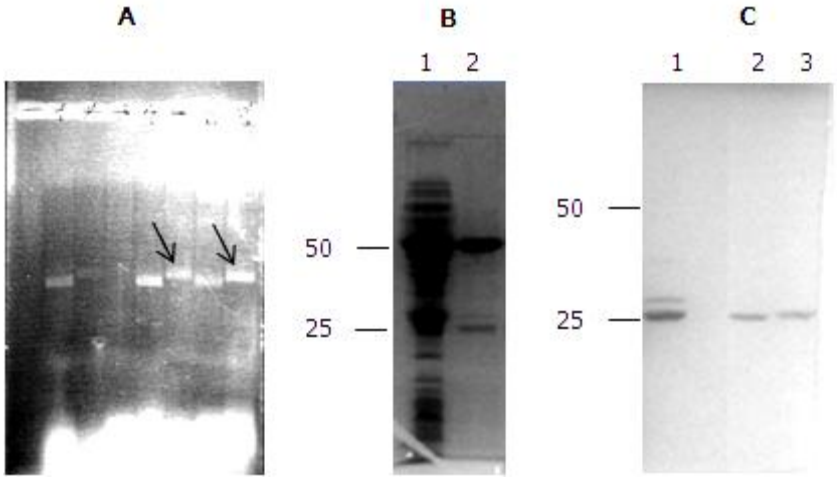

Figure R.2. Steps of CBM9b production cloned in the pGEX-t4-2 vector plasmid. A) Rapid plasmid isolation of DH5α transformed clones. Positive clones are marked with arrows. B) Glutathione affinity purification of GST-CBM9b containing extract. Coomassie staining of the gel showed the GST-CBM9b (48 kDa) was successfully purified but the sample also contained other impurities, including a band with the apparent size of GST alone (26 kDa). Lanes: 1, crude cell extract; 2, purified sample. C) Ion exchange purification of purified GST-CBM9b after thrombin digestion. Two bands with slightly different apparent size around 25 kDa were purified. Lanes: 1, thrombin digested extract; 2 and 3, collected purified protein sample.

Great amounts of the target protein were produced in the soluble fraction when the strain was induced by 1mM IPTG for 3 h at 37 ºC and 300 rpm. Consequently, the crude cell extract of a larger induced culture was used to perform affinity chromatography using a GSTrap FF column. The protein was obtained although there were some protein impurities in the sample (**Fig R.2.B**). The protein samples were then subject to thrombin digestion for 16 h to release the CBM9b domain, since the peptide link between this and the GST domain contains a thrombin recognition site. After digestion, both GST and CBM9b were expected to be present in the digested sample. These two molecules, presenting sizes of 26 kDa and 22 kDa respectively, could be separated by ion exchange chromatography. This second purification step was performed with a Mono Q 50/5 column. Of the fractions analyzed, only two contained proteins with the similar apparent sizes of GST or CBM9b (**Fig. R.2.C**). Identification of these proteins was then attempted based on the hypothetical cellulose binding capacity of CBM9b (**Section R.3.3.**).

R.1.3. GH10-CBM9a-b

The truncated form of the Xyn10C xylanase containing the catalytic domain GH10 coupled with the CBM9a-b tandem was cloned into pET101/d-TOPO as a first approach to express it. The cloned vector was transformed into *E. coli* Top10 strain and only one positive clone was screened from colony PCR (**Fig. R.3.A**). The recombinant plasmid containing the insert had its sequence confirmed by DNA sequencing and it was transformed into various *E. coli* IPTG inducible strains. The best results were obtained with *E. coli* BL21(DE3)/pET101:GH10-CBM9a-b recombinant strain incubated with 0,5 mM IPTG at 16ºC for 16h, although only small amounts of protein were overexpressed solubly (**Fig. R.3.B**). Despite that, a larger culture was incubated with the same

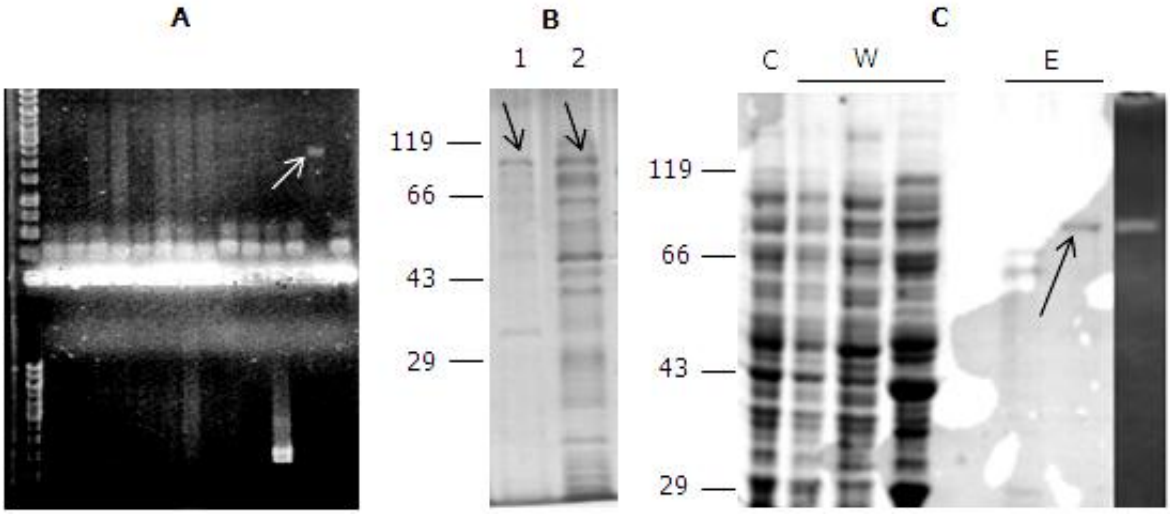

Figure R.3. Steps of GH10-CBM9a-b production cloned in the pET101 vector plasmid. A) Colony PCR of *E. coli* DH5α transformed clones. One positive clone is identifiable (arrow). B) Protein extract of GH10-CBM9a-b producing *E. coli* recombinant strain. A faint protein band with the apparent size of GH10-CBM9a-b (84 kDa) can be seen (arrows). Lanes: 1, insoluble fraction; 2, soluble fraction. C) IMAC purification of GH10-CBM9a-b containing extract. Coomassie staining after SDS-PAGE reveals a protein band with GH10-CBM9a-b apparent size in one of the eluted fractions. This protein exhibits xylanase activity as evidenced by zymogram analysis. Lanes: C, crude cell extract; W, unbound fractions; E, imidazole eluted fractions.

conditions and was submitted to protein purification by IMAC, using a HisTrap column. Fractions corresponding to absorbance peaks were analyzed (**Fig R.3.C**). One of the fractions contained a fainted band with an apparent molecular weight corresponding to GH10-CBM9a-b that was also represented by zymogram analysis. Despite that, the quantity of protein obtained was insufficient to proceed to enzymatic assays. Therefore, this recombinant plasmid was considered unviable for protein production.

Cloning of this truncated form of Xyn10C was attempted in pET28a but clones were never obtained.

R.1.4. GH10

The plasmid vector pET28a was used to clone the catalytic domain GH10 fused to an N-terminal his-tag. Screening for positive recombinant clones was carried out by rapid plasmid isolation using *E. coli* DH5α transformed cells. All the analyzed clones appeared to be positive (Fig. R.4.A). After confirmation by DNA sequencing checking, *E. coli* BLR strain was transformed with the recombinant plasmid from one clone and tested for protein production. This recombinant strain produced great amounts of the target protein under the induction conditions of 1mM IPTG for 3h at 37ºC (**Fig. R.4.B**). A larger culture was induced with the same conditions and submitted to IMAC purification using HisTrap column. Fractions corresponding to a great absorbance peak contained great amounts of the target protein alone and the zymogram showed it had xylanase activity (**Fig. R.4.C**). The pure protein extract was submitted to crystallography studies to determine the tridimensional structure of this GH10 domain (**Section R.4.**). The extract was also used for the biochemical characterization of its catalytic properties (**Section R.2.**).

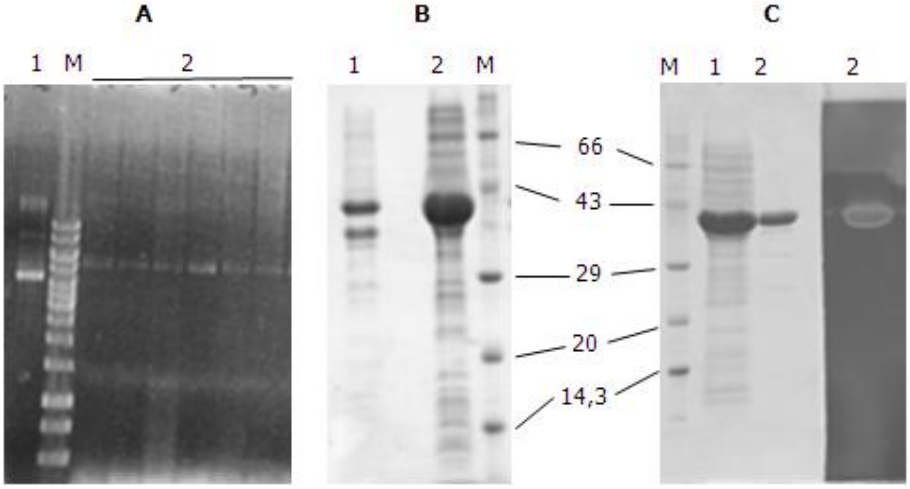

Figure R.4. Steps of GH10 production cloned on pET101 vector plasmid. A) Rapid plasmid isolation screening of *E. coli* DH5α transformant clones. Every analyzed clone (lanes 2) presented a DNA band above that of the control (lane 1). B) Protein extracts of GH10 producing *E. coli* recombinant strain. A highly overexpressed protein band with the apparent size of GH10 (42 kDa) is visible by Coomassie staining. Lanes: 1, insoluble fraction; 2, soluble fraction; M, protein ladder with sizes (kDa). C) IMAC purification of GH10 containing protein extract. One purified protein band with the apparent size of GH10 is seen by Coomassie staining. Xylanase activity is revealed by zymogram analysis. Lanes: M, protein ladder with sizes (kDa); 1, crude cell extract; 2, purified sample.

R.1.5. Xyn10CΔSP

The complete Xyn10C xylanase was cloned into the pET28a plasmid. In order to maintain the his-tag motif in the N-terminus of the expressed protein, the first 28 aminoacids of the sequence were not included in the DNA amplification since they correspond to the signal peptide which is cleaved in *E. coli* hosts during protein translation. Due to time constrains, positive screening was made directly by protein analysis of *E. coli* BLR transformed cells. Grown cultures of 1mL were incubated with 1mM IPTG for 1h30 at 37ºC and the corresponding crude cell extracts were analyzed (**Fig. R.5.A**). One clone was producing the target protein analysis and was consequently chosen for protein purification, after sequence confirmation of the recombinant plasmid. A large volume culture of the selected clone was induced under the same conditions and the crude cell extract obtained was used for IMAC purification with HisTrap chromatography column. SDS-PAGE analysis of the collected fractions showed a protein band with the target protein's expected size and a faint extra band of smaller size (**Fig. R.5.B**). These two bands showed xylanase activity upon zymogram analysis and for this reason the smaller band was considered to be a proteolysis product from the target protein, similarly to what has been described previously (ref). Fractions corresponding to this absorbance peak were pooled together and used for enzymatic studies (**Sections R.2.** and **R.3.1.**).

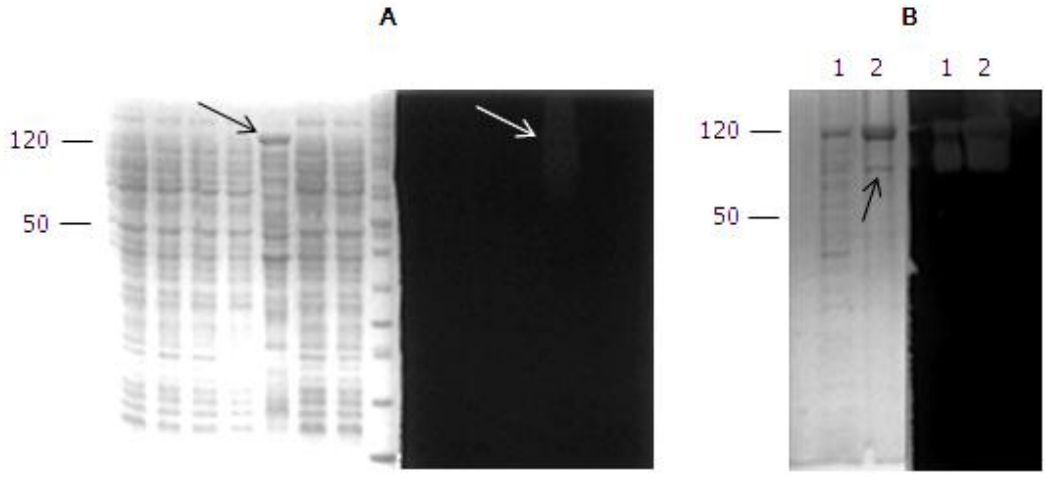

Figure R.5. Steps of Xyn10C production cloned on pET28a vector plasmid. a) Analysis of crude cell extracts from *E. coli* BLR(DE3) transformed clones. Each lane of the Coomassie staining of the gel corresponds to a single clone, doubled for zymogram analysis. One clone presents an active protein with the apparent size of Xyn10C (117 kDa) (arrows). B) IMAC purification of Xyn10C containing protein extract. Coomassie staining evidences a protein band with Xyn10C apparent size and a secondary protein band (arrow). The two proteins present xylanase activity as seen by zymographyc staining. Lanes: 1, crude cell extract; 2, Xyn10C purified sample.

R.2. Enzymatic activity

The wild type form of the Xyn10C xylanase and its catalytic GH10 module were cloned and purified as described above from *E. coli* BLR(DE3)/pET28a:Xyn10CΔSP and *E. coli* BLR(DE3)/pET28a:GH10 recombinant strains, respectively and were submitted to quantitative enzymatic studies. Specific activity, expressed as U/μmol, was measured towards xylan samples

extracted from birchwood, beechwood, oat spelt, wheat and rye (**Fig. R.6.A**). Both enzymes presented low activity towards birchwood xylan (16.5 U/µmol for Xyn10C; 19.1 U/µmol for GH10) and high activity towards rye xylan (317.0 U/µmol for Xyn10C; 367.0 U/µmol for GH10). Moreover, the activity towards wheat arabinoxylan was significantly inferior for the GH10 truncated enzyme (158.0 U/µmol for GH10; 241.5 U/µmol for Xyn10X; *p*-value<0,01) in respect to the activity of the complete Xyn10C. The same assay was performed with the soluble and insoluble fractions of birchwood and oat spelt xylan. The activity towards the insoluble fractions of the xylans was expressed in relation to the activity towards the soluble fractions (**Fig. R.6.B**). The two enzymes showed similar activity against the soluble and insoluble fractions of birchwood xylan. However, the activity of both enzymes was greatly inferior on insoluble oat spelt xylan than the activity measured for the corresponding soluble fraction. This reduction was found to be significantly superior for the GH10 catalytic domain enzyme that for Xyn10C (8.7% for GH10; 35.7% for Xyn10C; *p*-value < 0,01).

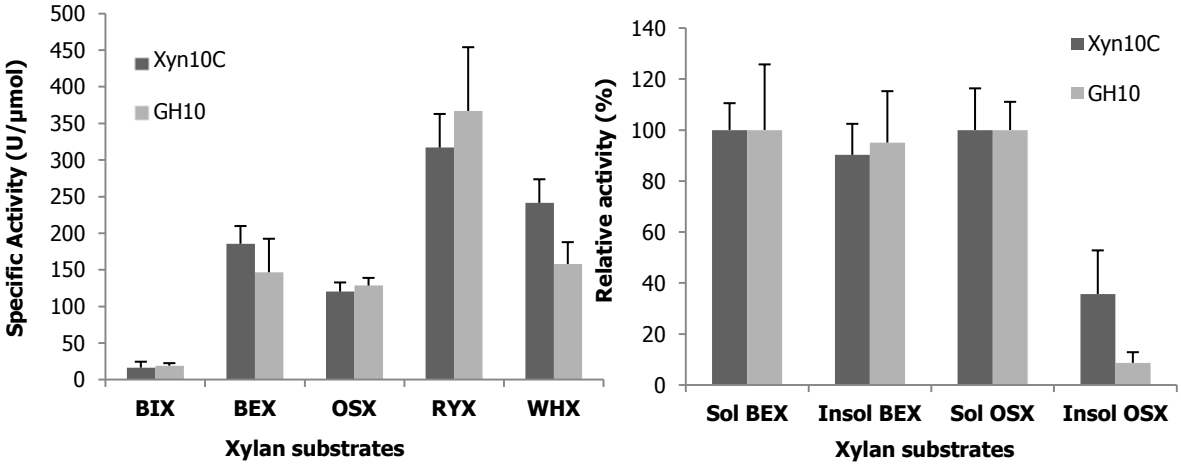

Figure R.6. Comparison of enzymatic activities between XynC10 (wt) and GH10. A) Specific activity profiles of the xylanases towards several xylan substrates. BIX: birchwood xylan, BEX: beechwood xylan, OSX: oat spelt xylan, RYX: rye xylan, WHX: wheat xylan. B) Activity profiles of the xylanases towards soluble and insoluble xylan substrates. Sol/Insol BEX: soluble/insoluble beechwood xylan, Sol/Insol OSX: soluble/insoluble oat spelt xylan.

R.3. Cellulose binding

R.3.1. Xyn10C

The binding essay of the complete form of the Xyn10C (wt) enzyme was initially performed by the centrifugation method using the crude cell extract of an E. coli BLR(DE3)/pBR322:Xyn10C culture, which contains the enzyme's sequence in its native form. The crude cell extract, the washed fraction and the Avicel bound fraction were analysed by SDS-PAGE/zymogram (**Fig. R.7.A**). The protein could not be observed in the crude cell extract, probably because it is not cloned in an inducible system and is being produced in low quantities, constitutively. On the other hand, zymogram analysis revealed three distinct active proteins indicating that the protein was

being expressed and active as it hydrolyzed xylan polymers. Coomassie staining of the gel didn't show Xyn10C on the Avicel bound fraction. However, when this fraction was analyzed by zymogram, the highest molecular weight protein of the 3 active bands observed in crude extract was detected. The apparent molecular weight of this protein was 119 kDa, the same as the Xyn10C. The other two proteins showed molecular sizes of around 100 kDa and 80 kDa. Despite all this, the Avicel unbound fraction revealed also high activity on the zymogram, evidencing that the majority of the protein didn't bound to microcrystalline cellulose.

The cellulose binding capacity was tested again with crude cell extracts from the overexpressing recombinant strain *E. coli* BLR(DE3)/pET28a:Xyn10CΔSP by the column method. Although the protein was being overexpressed, no improvement of the Avicel binding capacity was observed (result not shown).

R.3.2. CBM9a-b

Crude cell extracts from induced *E. coli* BLR/pET101:CBM9a-CBM9b culture were used to evaluate the Avicel binding capacity of the CBM9a-b tandem. Two fractions from the resulting IMAC purification that were thought to contain this protein, although only partially purified, were tested through the centrifugation method and analysed by SDS-PAGE (**Fig. R.7.B**). On the Avicel bound fraction, both extracts presented a fainted band with the protein's size but most of the protein was in the unbound fraction. The same essay was repeated with 16h of incubation with the microcrystalline cellulose instead of the standard 2h, but there was no improvement on the results (results not shown). The Avicel binding assay was also performed by the column method directly from the crude extracts of the same recombinant strain but there was still no improvement on these results (results not shown).

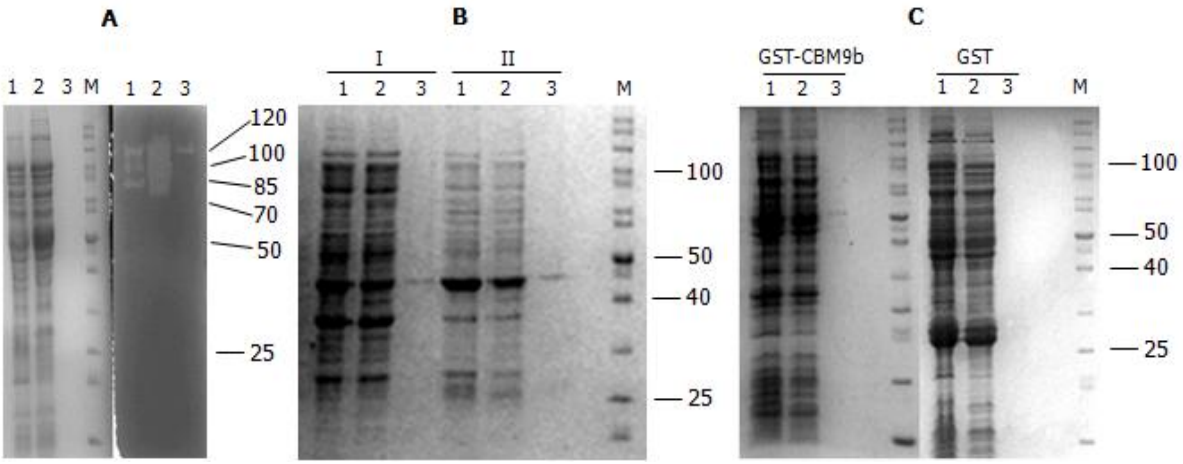

Figure R.7. Microcrystalline cellulose binding assay of different Xyn10C derivatives. A) Xyn10C binding assay. The enzyme is not visible by Coomassie staining. Of the three bands visible in the zymogram, only the band that appears to be the complete Xyn10C (117 kDa) is seen in the Avicel bound fraction. The other two present sizes of around 100 and 80 kDa. B) CBM9a-b binding assay. A fainted band showing the CBM9a-b expected size (48 kDa) is visible on the Avicel bound fraction of both samples (I and II) obtained from the IMAC purification. C) CBM9b binding assay. A faint band with the apparent size of GST-CBM9b is visible in the Avicel bound fraction while no protein band is detected when the GST domain was evaluated alone. Lanes: 1, protein sample/extract; 2, unbound fraction; 3, Avicel bound fraction.

R.3.3. CBM9b

Both fractions collected from the IMAC purification of the thrombin digested GST-CBM9b protein extract were submitted to Avicel binding assay through the centrifugation method. Each of these fractions contained purified protein samples thought to correspond either to GST (26kDa) or the CBM9b (22kDa) proteins (**Section R1.2.**). By analysing the cellulose binding capacities of these enzymes, the CBM9b domain could be identified because of its putative cellulose binding capacity. This would allow the protein to be unequivocally produced and consequently submitted for crystallography and a thorough carbohydrate binding analysis. However, none of these proteins showed Avicel affinity (Results not shown).

However, the microcrystalline cellulose affinity of the CBM9b domain was assessed through other strategy. The binding assay was repeated by column chromatography with crude cell extracts from induced *E. coli* DH5a/pGEX-T4-2:CBM9b recombinant strain culture. If, in fact, the CBM9b domain has the capacity to bind to Avicel, that feature would also be present by the overexpressed GST-CBM9b fusion protein. To exclude the hypothesis that the GST domain could have itself binding capacity, crude cell extract from the GST producing *E. coli* DH5a/pGEX-T4-2 strain was used as a negative control. A fainted protein band with the GST-CBM9b expected size was visible on the Avicel bound fraction while no band was detected in the bound fractions of the extract containing the GST protein (**Fig R.7.C**). This indicates that CBM9b presents low affinity to microcrystalline cellulose.

R.4. Protein crystall from GH10

The extract containing the pure GH10 catalytic domain of Xyn10C was submitted to structural studies through crystallography methods and protein crystals were successfully obtained (**Fig. R.8**). However, up to this point, this analysis is still at a very early stage and is not sufficient to provide insight on its structural conformation.

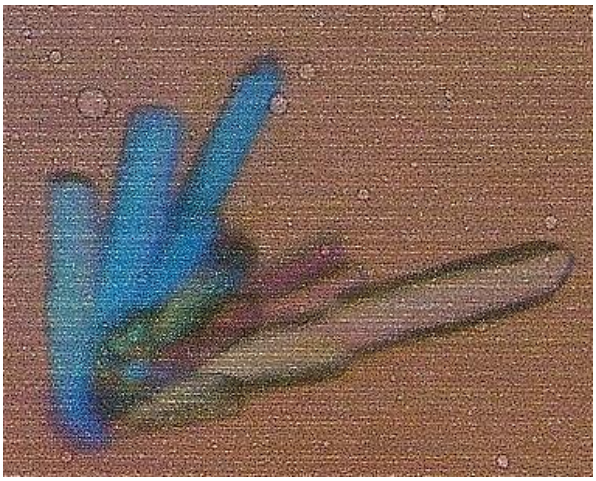

Figure R.8. Protein crystals of GH10 catalytic domain.

Result Analysis

A.1. Different variants of Xyn10C were cloned and produced

In order to obtain more insight on how Xyn10C xylanase from P. barcinonensis functions and what are the roles of the various modules that compose it, several truncated forms of Xyn10C xylanase of *P. barcinonensis* were successfully cloned. The complete protein without the signal peptide, here designated by Xyn10CΔSP or simply Xyn10C, the catalytic domain GH10 and the CBM9b domain were produced fused to an N-terminal his-tag motif using pET28a vector plasmid. Contrarily, the adjacent domains GH10-CBM9a-b and the CBM9a-b tandem were successfully fused to a his-tag motif on the C-terminus of the protein using the pET101 vector. Additionally, the CBM9b domain was also cloned in pGEX-T4-2, fusing its N-terminus to a GST domain.

The effect of the N-terminal tag in the case of the GH10 construction was striking. High amounts of soluble protein were produced by the resulting recombinant strain and only one purification step by IMAC with a Ni2+ based column was necessary to obtain high quantities of pure enzyme from the cell extract of a 1L induced culture. On the other hand, the results obtained with the pET28a construction that contained the full enzyme also fused with an N-terminal his-tag, were less satisfactory, although the cloning, expression and purification steps were analogous to the GH10 truncated enzyme. In this case, lower levels of expression were observed which could have contributed to a less effective purification. However, it is necessary to remark that in this case, due to time constrains, no optimization of the expression conditions was performed and induction was made for 1h while the GH10 protein expression was induced for 3h. In addition, the complete enzyme Xyn10C is substantially larger, around three times, than the GH10 domain, which implies that it needs more time and more cell resources to be produced. Together, these two facts can be responsible for the registered lower yield in Xyn10C production.

The truncated forms of the enzyme produced using the pET101 vector plasmid were more problematic as the target proteins were being produced mostly in the insoluble fraction of the extracts. Production of GH10-CBM9a-b truncated enzyme, when cloned in this vector, was complicated because, despite the serious effort in trying different induction conditions, no adequate soluble overexpression was obtained in order to conduct a proper purification. A very important difference between this construction and the GH10 construction, apart from being cloned in different plasmids, pET101 or pET28a, is that the latter contains the 5 aminoacids previous to the first aminoacid of the GH10 predicted domain, thus harboring a portion of the link between this domain and the preceding one, and the N-terminal 21 aminoacid long his-tag motif. The lack of these extra 5 amino acids from the linker, the his-tag or even both at the N-terminus can have critically compromised the correct folding of the GH10-CBM9a-b. Probably the protein was being produced too trimmed to the GH10 domain, containing functional important amino acids, thus leading to low yields of soluble overexpressed protein. Regardless, purification of the

GH10-CBM9a-b protein was performed from the soluble fraction was possible as the active protein was visible by zymographic analysis, evidencing that despite the small productivity of soluble protein, the C-terminus his-tag motive proved to be an effective strategy for IMAC purification of the CH10-CBM9a-b construction.

The family 9 carbohydrate binding domain tandem, here referred to as CBM9a-b, had been cloned before into pET11a vector plasmid twice, once without the 9 aminoacid liker between the domain and the previous GH10 domain and once including it (unpublished data). In this work, the construction that included the linker, pET11a:Link-CBM9a-b, was used to produce the desired protein but despite all efforts, none of the inducing conditions tested were effective to overexpress the enzyme in the soluble fraction, although great amounts were produced insolubly. As an alternative, a different construction of the Link-CBM9a-b protein was performed. In this way, the pET101 vector plasmid and the addition of a C-terminal his-tag was tested. This gave rise to a successful production in the soluble fraction that would allow the subsequent protein purification. However, while other IMAC purifications were successful using 5% of the elution buffer to prevent unspecific bindings, in this case, this imidazole background had to be removed so that the protein could bind to the column. Consequently, although the IMAC purified fractions contained high quantities of the target protein, they also contained a great number of contaminant proteins that bound unspecific, thus rendering the purification under these conditions unfeasible.

The CBM9b, thought to be the actual functional domain of the tandem CBM9a-b, was cloned into the pET28a vector thus fused to an N-terminal his-tag motif. It was previously seen that this peptide might have enhanced the solubility of GH10, but contrarily, it didn't have a positive effect for CBM9b. The majority of CBM9b enzyme was produced insolubly and only a small, but distinguishable, band was seen upon analysis of the correspondent soluble fraction. However, when this soluble extract was submitted to IMAC purification, no protein appeared in the eluted samples. This was accounted for the low quantities of desired protein present in the extract and no optimization of the purification was done. However, the CBM9b and the Link-CBM9a-b truncated enzymes share a feature that separates them from all the other successfully purified proteins (Xyn10C, GH10 and GH10-CBMa-b). The calculated isoelectric point (pI) of the two former proteins is 7.86 and 8.98, respectively while the three latter proteins have pI values between 5.53 and 6.15. Since the pH of the buffers advised for the standard conditions of the purification is 7.4, it can be that proteins like the CBM9b and the Link-CBM9a-b, which present a relatively close pI, lose some of its net charge, hindering the histidine hexapeptide affinity to the Ni^{2+} ions of the column matrix.

Given that the soluble production of CBM9b with the pET28a plasmid as well as the purification of the protein itself were not successful, this last domain of Xyn10C xylanase was cloned into pGEX-T4-2, fusing it with an N-terminal GST motif. This construction was effective in producing great amounts of the target protein fused with the GST tag in a soluble state and this fusion protein was successfully purified through a glutathione affinity column, exemplifying the advantages of this type of protein cloning strategy.

A.2. Ancilliary CBM modules assist arabinoxylan hydrolysis

A first characterization of Xyn10C from *P. barcinonensis* has been previously done[5]. There, it was reported that the enzyme was highly active on xylan compounds while the activity on amorphous cellulose (CMC) and crystalline cellulose (Avicel) was low. The enzymatic activity, measured with the crude cell extracts and not with purified protein, was higher on birchwood xylan than on oat spelt xylan. In the present work, apart from the birchwood xylan and oat spelt xylan, beechwood, wheat and rye xylan were used as substrates in the enzymatic activity assay using the purified proteins Xyn10C (wt) and GH10, both obtained using a pET28a construction. The two forms of the enzyme displayed a similar enzymatic profile and inclusively the data from this study appears to contradict the previous one in the sense that the activity on birchwood was the lowest of all the substrates tested, even comparing with the activity on oat spelt xylan. Probabily, the fact that the earlier characterization was performed with the crude cell extract is in the base for the registered dissidence. On the other hand, both enzymes showed far more activity on the highly arabinose branched rye xylan than on any other substrate. Interestingly, it seems that there is a positive correlation between Xyn10C activity and the frequency of branches presented by the xylan polysaccharride (**Fig. A.1**). However, beechwood xylan appears to be an exception to it. Although this substrate is very similar to birchwood xylan, beechwood xylan was reported to contain 3 times more hexuronic acid[31] and frequent glucuronic acid ramifications resulting in xylose contents of 83%[24]. While these differences can account for the observed difference in Xyn10C activity between these hardwood xylans, the biochemical reason remains unclear and different analysis need to be performed in order to understand the profile activity of Xyn10C.

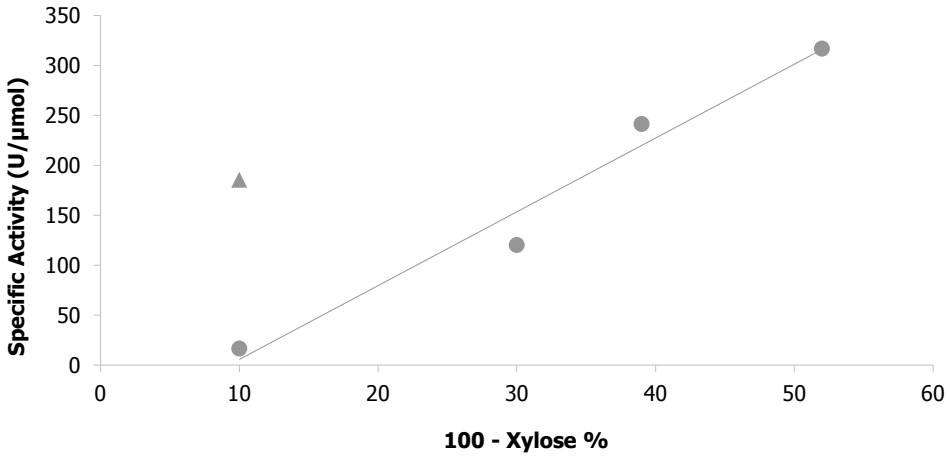

Figure A.1. Correlation between Xyn10C activity and content of polysaccharides other than xylose. The specific activities of Xyn10C were plotted against the level of branching of the main chain of different xylan substrates, calculated as the content of sugars other than xylose. Dots, from left to right, represent the activity values obtained beechwood, oat spelt, wheat and rye xylans. The triangle represents the activity value obtained for birchwood xylan. Regression line: y=7,3877x - 68,176; R^2=0,968 (Dots only).

The comparison of the activities of the complete Xyn10C and its GH10 truncated form suggested that the carbohydrate binding domains, absent in the latter, were not essential for the hydrolysis of beechwood and birchwood glucuroxylan and oat spelt and rye arabinoxylan. Nevertheless, the activity of GH10 truncated enzyme on wheat arabinoxylan was statistically inferior than the activity of whole Xyn10C, showing that in some cases, the hydrolytic activity of the catalytic domain is in fact enhanced by the presence of these binding domains.

To further address to role of the non catalytic domains, the activity assay was performed with the soluble and insoluble fraction of xylan substrates. Only two substrates were selected, beechwood xylan and oat spelt xylan, because both displayed an important amount of visible insoluble fractions and because they represent two distinct types of xylan, hardwood xylan (glucuronoxylan) and softwood xylan (arabinoxylan), respectively. Birchwood xylan was not tested because the enzyme has low activity towards it, while wheat and rye xylans were not tested because they don't display any apparent insoluble fraction when diluted in aqueous solution. Interestingly, the two forms of the enzyme displayed the same activity levels on both insoluble and soluble beechwood xylan while a great reduction in the enzymes' activity was found on the insoluble fraction of oat spelt. This activity reduction between the soluble and insoluble fractions of oat spelt xylan was more acute for the GH10 truncated enzyme, again suggesting that the non catalytic modules presented by this enzyme enhance the activity of the GH10 catalytic domain against arabinoxylan. It is possible that this difference was not observed without separating the soluble and the insoluble fractions of oat spelt due to the discrepancy between the enzyme's activity on both substrates.

A.3. CBM9b plays a role in cellulose binding

In the previous work, Xyn10C of *P. barcinonensis* was shown to bind to microcrystalline cellulose (Avicel)[5]. Here, those results were reproduced and just like the previous study, of the several protein bands detected by zymogram, only the band with the greater MW was visible on the Avicel bound fraction. This suggests that Xyn10C suffered proteolysis losing functional domains involved in cellulose binding but not in xylan hydrolysis. Post translational proteolysis of xylanases has been reported before to occur naturally[1]. However, in this work, Xyn10C is being expressed heterologously on *E.* coli strains so it is impossible to say if this constitutes or not an intrinsic feature of *P. barcinonensis*. Since the two secondary proteins observed in the zymogram display an apparent MW of 100 and 75 kDa, one can hypothesize that they correspond to Xyn10C devoid of CBM9b and CBM9a-b as the corresponding expected size would be 94 and 72 kDa, respectively, thus reflecting the expected role of CBM9b to cellulose binding. On the other hand, Coomassie staining of the acrylamide gel didn't reveal any protein on the Avicel bound fraction, not even when the binding assay was repeated with an overexpressing strain, indicating that the binding kinetic of the protein towards crystalline cellulose is weak. The binding assay of the two semi purified fractions of CBM9a-b strengthened this hypothesis as the corresponding band, that was clearly

visible on both samples' protein profile, was only barely visible on the Avicel bound fraction, while most of it was present in the unbound fraction. The binding ability of CBM9b to Avicel had to be evaluated while fused to the GST N-terminal domain because when the two domains, GST and CBM9b, were separated by thrombin digestion, none of the resulting proteins presented binding ability, suggesting that CBM9b has lost its right configuration, and function, upon separation from the GST motif. Thus, the assay was performed with the fused GST-CBM9b protein resulting in the same outcome as with CBM9a-b. Together, these results hint that the domain responsible for the microcrystalline cellulose binding of Xyn10C is CBM9-b as no difference in Avicel binding was observed between this protein (fused with GST), the CBM9a-b tandem and the complete protein and that the CBM9-b mediated Xyn10C affinity to Avicel is low under the essayed conditions.

Discussion

The architecture exhibited by Xyn10C of *P. barcinonensis* is commonly found in family 10 xylanases among thermophilic bacteria. This clearly hints that the non catalytic domains play a significant role in the enzyme, important enough to be conserved throughout time and evolution. Regardless, the role of the non catalytic domains is still unclear. It is interesting to think that the duplication of domains with carbohydrate binding capacity, resulting in that only the most distant from the catalytic domain is active, serves to further separate the hydrolytic site from the anchor domain. Since linker peptides are known to be flexible regions on modular proteins, this can be an adaptation to further increase the flexibility of the catalytic domain while the carbohydrate binding domain is anchored on the substrate. On the other hand, these linker domains can have a thermoprotective effect on the catalytic domain, crucial for high temperature efficiency, probably not because of any active functional proprieties but due to its proximity and because they maintain the protein's overall structure.

The study described in the present work indicates that the microcrystalline cellulose binding capacity of Xyn10C is mediated through the affinity of subfamily b CBM9, the domain most distant from the catalytic site of a tandem family 9 carbohydrate binding module (CBM9a-b) situated at the C-terminus. This is consistent with the hypothesis mentioned above and is corroborated by other reports based on Xyn10A of thermophilic *Thermotoga maritima*[41]. However, the CBM9b domain of Xyn10A was thoroughly studied and was shown to bind to a broad array of carbohydrates, including microcrystalline cellulose (Avicel), and interestingly, to the reducing end of polysaccharides[9]. Crystallography studies reveal that the three-dimensional structure of this CBM9b consists of two antiparallel b-sheets, each made up of 5 β-strands to form a β -sandwich fold and that cellobiose molecules bind to a cleft created by two parallel tryptophan residues, Trp70 and Trp175[9]. The involvement of aminoacids with aromatic side chains is a common trait of CBM binding sites[8]. Alignment of Xyn10C-CBM9b of *P. barcinonensis* with the Xyn10A-CBM9b from *T. maritima* revealed that, while the structural features and components remain conserved, the cleft at the binding site is composed of a histidine residue and a tyrosine residue, His70 and Tyr175[41]. This means that the two nonpolar indole aromatic side chains found in tryptophan residues are substituted by the imidazole group of a histidine and the benzene group of a tyrosine, both polar side chains. This differences observed in the binding site can account for the low microcrystalline cellulose affinity and, at the same time, it suggests that this protein domain can present different proprieties than those found for *T. maritima* Xyn10A-CBM9b. Curiously, while the Trp71 residue of Xyn10A-CBM9b is little conserved throughout subfamily b CBM9, Trp175 residue is highly conserved in CBM9b domains that appear in association with members of subfamily a CBM9. However, CBM9b members that appear alone present a conserved tyrosine residue instead of the tryptophan[41], much like the CBM9b domain studied in the present work. In any case, a more thorough study of the binding profile presented by Xyn10C-CBM9b is required.

Overall, with the present work, the type of xylan to which Xyn10C presents more xylanolytic potential became clear. However, the pH and temperature at which the enzymatic assays were performed were in agreement with the estimated optimum conditions by in the previous work. Since those results appeared to be uncertain in light of the results here obtained, the optimum conditions for Xyn10C activity would have to be determined again to learn the actual specific activity of this enzyme towards the different xylan substrates. Regardless, this study indicates that Xyn10C presents much higher activity towards highly arabinose substituted xylan polymers, like those present in cereal lignocellulosic tissues, than glucuronoxylans derived from hardwood species. It is also proposed that the carbohydrate binding domains of this enzyme represent adaptations to the hydrolysis of some arabinoxylan that would be otherwise difficult substrates. This positive effect was seen on the hydrolysis of wheat xylan and on the hydrolysis of insoluble oat spelt xylan. On the other hand, the presence or absence of these domains was irrelevant for the hydrolysis capacity of glucuronoxylan substrates. Thus, it seems that Xyn10C is an enzyme that was evolutionarily design to present more affinity towards grass xylan, mainly composed of arabinoxylan, and that the inclusion of the carbohydrate binding domains in this enzyme confers an advantage to the host since *P. barcinonensis* BP-23 was isolated from a rice field and xylan extracted from rice shows an arabinose/xylose ratio up to 0.98[37].

The application of this enzyme in the paper industry can be investigated, since arabinose branched xylan is also an important component of softwood species. The usage of this enzyme is particularly advantageous due to its cellulase-free xylanolytic activity. However, the very high hydrolytic capacity towards arabinoxylans exhibited by Xyn10C makes it a very strong candidate for the biodegradation of cereal lignocellulosic material, a major by-product of agro-textile industries. This process is the basis of a much more rentable way to obtain energy from plant material than through combustion: sugar fermentation[28]. It is also a very important step to reduce the price of the feedstock to biorefinaries of the polysaccharides contained in the lignocelluloses biomass in order to obtain bioethanol[56]. Furthermore, this study also perspectives the utilization of this enzyme in the enhancement of animal feed, namely poultry feed, since the degradation of arabinoxylan polymers present in wheat and rye enhances digestion and feed intake by diminishing the viscosity in the small intestine[2].

The continued study of Xyn10C from *P. barcinonensis*, and the evaluation of its effectiveness as a xylanase towards raw material will allow the confirmation of its potential in the industrial applications here proposed.

Bibliography

1. **Badhan A.K., Chadha B.S., Kaur J., Saini H.S., and Bhat M.K.** 2007. Production of multiple xylanolytic and cellulolytic enzymes by thermophilic fungus *Myceliophthora* sp. Bioresource Technology 98: 504-510.
2. **Bedford M.R.** 1995. Mechanism of action and potential environmental benefits from the use of feed enzymes. Animal Feed Science and Technology 53: 145-155.
3. **Biely P., Vršanská M., Tenkanen M., and Kluepfel D.** 1997. Endo-β-1,4-xylanase families: differences in catalytic properties. Journal of Biotechnology 57: 151-166.
4. **Birnboim H., and Doly J.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research 7: 1513–1523.
5. **Blanco A., Díaz P., Zueco J., Parascandola P., and Pastor F.I.J.** 1999. A multidomain xylanase from a *Bacillus* sp. with a region homologous to thermostabilizing domains of thermophilic enzymes. Microbiology 145: 2163-2170.
6. **Blanco A., and Pastor F.I.J.** 1993. Characterization of cellulase-free xylanases from the newly isolated *Bacillus* sp. strain BP-23. Canadian Journal of Microbiology 39: 1162-1166.
7. **Blanco A., Vidal T., Colom J., and Pastor F.** 1995. Purification and properties of xylanase a from alkali-tolerant *Bacillus* sp. Strain BP-23. Applied and Environmental Microbiology 61: 4468-4470.
8. **Boraston A.B., Bolam D.N., Gilbert H.J., and Davies G.J.** 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochemical Journal 3: 769–781.
9. **Boraston A.B., Creagh A.L., Alam M.M., Kormos J.M., Tomme P., Haynes C.A., Warren R.A.J., and Kilburn D.G.** 2001. Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. Biochemistry 40: 6240-6247.
10. **Brett C.T., and Waldron K.** Physiology and biochemistry of plant cell walls. Chapman & Hall, pp. 32-36.
11. **Buchert J., Tenkanen M., Kantelinen A., and Viikari L.** 1994. Application of xylanases in the pulp and paper industry. Bioresource Technology 50: 65-72.
12. **Cantarel B.L., Coutinho P.M., Rancurel C., Bernard T., Lombard V., and Henrissat B.** 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research 37: D233-D238.
13. **Charnock S.J., Bolam D.N., Turkenburg J.P., Gilbert H.J., Ferreira L.M.A., Davies G.J., and Fontes C.** 2000. The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: Structure and biochemistry of the *Clostridium thermocellum* X6b domain. Biochemistry 39: 5013-5021.
14. **Chiriac A., Cadena E., Vidal T., Torres A., Diaz P., and Javier Pastor F.** 2010. Engineering a family 9 processive endoglucanase from *Paenibacillus* barcinonensis displaying a novel architecture. Applied Microbiology and Biotechnology 86: 1125-1134.
15. **Cohen S.N., Chang A.C.Y., and Hsu L.** 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proceedings of the National Academy of Sciences 69: 2110-2114.
16. **Collins T., Gerday C., and Feller G.** 2005. Xylanases, xylanase families and extremophilic xylanases. Fems Microbiology Reviews 29: 3-23.
17. **Cutillas-Iturralde A., Zarra I., and Lorences E.P.** 1993. Metabolism of cell wall polysaccharides from persimmon fruit. Pectin solubilization during fruit ripening occurs in apparent absence of polygalacturonase activity. Physiologia Plantarum 89: 369-375.
18. **de Vries R.P., Kester H.C.M., Poulsen C.H., Benen J.A.E., and Visser J.** 2000. Synergy between enzymes from *Aspergillus* involved in the degradation of plant cell wall polysaccharides. Carbohydrate Research 327: 401-410.
19. **de Vries R.P., and Visser J.** 2001. *Aspergillus* Enzymes Involved in Degradation of Plant Cell Wall Polysaccharides. Microbiololy and Molecular Biology Reviews 65: 497-522.
20. **Deng W., Jiang Z.Q., Li L.T., Wei Y., Shi B., and Kusakabe I.** 2005. Variation of xylanosomal subunit composition of *Streptomyces olivaceoviridis* by nitrogen sources. Biotechnology Letters 27: 429-433.
21. **Dias F.M.V., Goyal A., Gilbert H.J., Prates J.A.M., Ferreira L.M.A., and Fontes C.** 2004. The N-terminal family 22 carbohydrate-binding module of xylanase 10B of

Clostridium themocellum is not a thermostabilizing domain. Fems Microbiology Letters 238: 71-78.

22. **Dobozi M.S., Szakacs G., and Bruschi C.V.** 1992. Xylanase activity of *Phanerochaete chrysosporium*. Applied Environmental Microbiology 58: 3466-3471.
23. **Don R.H., Cox P.T., Wainwright B.J., Baker K., and Mattick J.S.** 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research 14: 4008.
24. **Ebringerova A., Alfoldi J., Hromadkova Z., Pavlov G.M., and Harding S.E.** 2000. Water-soluble p-carboxybenzylated beechwood 4-O-methylglucuronoxylan: structural features and properties. Carbohydrate Polymers 42: 123-131.
25. **Gallardo O., Diaz P., and Pastor F.I.J.** 2003. Characterization of *Paenibacillus* cell-associated xylanase with high activity on aryl-xylosides: a new subclass of family 10 xylanases. Applied Microbiology and Biotechnology 61: 226-233.
26. **Gallardo O., Fernandez-Fernandez M., Valls C., Valenzuela S.V., Roncero M.B., Vidal T., Diaz P., and Pastor F.I.J.** 2010. A xylanase of family GH5 with activity on neutral oligosaccharides: Characterization and evaluation as a pulp bleaching aid. Applied and Environmental Microbiology: 871-10.
27. **Gao D., Uppugundla N., Chundawat S.P., Yu X., Hermanson S., Gowda K., Brumm P., Mead D., Balan V., and Dale B.E.** 2010. Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. Biotechnology and Biofuels 4.
28. **García-Aparicio M.P., Ballesteros M., Manzanares P., Ballesteros I., González A., and Negro M.J.** 2007. Xylanase contribution to the efficiency of cellulose enzymatic hydrolysis of barley straw. Applied Biochemistry and Biotecnology. In: Mielenz JR, Klasson KT, Adney WS, and McMillan JD, (eds.). Humana Press, pp. 353-365.
29. **Helm Richard F.** 1999. Lignin - Polysaccharide Interactions in Woody Plants. Lignin: Historical, Biological, and Materials Perspectives. American Chemical Society, pp. 161-171.
30. **Henrissat B.** 1991. A classification of glycosyl hydrolases based on amino-acid-sequence similarities. Biochemical Journal 280: 309-316.
31. **Hespell R., and Cotta M.** 1995. Degradation and utilization by *Butyrivibrio fibrisolvens* H17c of xylans with different chemical and physical properties. Applied and Environmental Microbiology 61: 3042-3050.
32. **Holtman K.M., Chang H.-M., Jameel H., and Kadla J.F.** 2003. Elucidation of lignin structure through degradative methods: comparison of modified DFRC and thioacidolysis. Journal of Agricultural and Food Chemistry 51: 3535-3540.
33. **Ishihara M., and Shimizu K.** 1988. Alpha (1-2) glucuronidase in the enzymatic saccharification of hardwood xylan .1. Screening of alpha-glucuronidase producing fungi. Mokuzai Gakkaishi 34: 58-64.
34. **Jerker P.** 1992. Immobilized metal ion affinity chromatography. Protein Expression and Purification 3: 263-281.
35. **Jorgensen H., Kristensen J.B., and Felby C.** 2007. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels, Bioproducts and Biorefining 43: 119-134.
36. **Kavoosi M., Meijer J., Kwan E., Creagh A.L., Kilburn D.G., and Haynes C.A.** 2004. Inexpensive one-step purification of polypeptides expressed in *Escherichia coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T. maritima*. Journal of Chromatography B 807: 87-94.
37. **Kormelink F.J.M., and Voragen A.G.J.** 1993. Degradation of different (glucurono)arabino xylans by a combination of purified xylan-degrading enzymes. Applied Microbiology and Biotechnology 38: 688-695.
38. **Laemmli U.K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.
39. **Lazaridou A., and Biliaderis C.G.** 2007. Molecular aspects of cereal β-glucan functionality: Physical properties, technological applications and physiological effects. Journal of Cereal Science 46: 101-118.
40. **McKendry P.** 2002. Energy production from biomass (part 1): overview of biomass. Bioresource Technology 83: 37-46.

41. **Notenboom V., Boraston A.B., Kilburn D.G., and Rose D.R.** 2001. Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. Biochemistry 40: 6248-6256.
42. **Nummi M., Perrin J.M., Niku-Paavola M.L., and Enari T.M.** 1985. Measurement of xylanase activity with insoluble xylan substrate. Biochemical Journal 2: 617–620.
43. **Pastor F.I.J., Gallardo O., Sanz-Aparicio J., and Díaz P.** 2007. Xylanases: Molecular proprieties and applications. In: Polaina J and MacCabe AP, (eds.). Industrial enzymes: structure, function and applications. Springer, pp. 70-73.
44. **Pastor F.I.J., Pujol X., Blanco A., Vidal T., Torres A.L., and Díaz P.** 2001. Molecular cloning and characterization of a multidomain endoglucanase from *Paenibacillus* sp BP-23: evaluation of its performance in pulp refining. Applied Microbiology and Biotechnology 55: 61-68.
45. **Peter B.** 1985. Microbial xylanolytic systems. Trends in Biotechnology 3: 286-290.
46. **Poutanen K., Sundberg M., Korte H., and Puls J.** 1990. Deacetylation of xylans by acetyl esterases of Trichoderma reesei. Applied Microbiology and Biotechnology 33: 506-510.
47. **Raven P.H., R.F. E., and S.E. E.** 1999. Biology of plants. W.H. Freeman and Company Worth Publishers, pp. 40–60.
48. **Sá-Pereira P., Paveia H., Costa-Ferreira M., and Aires-Barros M.** 2003. A new look at xylanases. Molecular Biotechnology 24: 257-281.
49. **Saha B.C.** 2000. Alpha-L-arabinofuranosidases: biochemistry, molecular biology and application in biotechnology. Biotechnology Advances 18: 403-423.
50. **Sambrook J., and Russel D.W.** 2001. Molecular Cloning: A Laboratory Manual. . CSH Laboratory Press.
51. **Sánchez M.M., Fritze D., Blanco A., Spröer C., Tindall B.J., Schumann P., Kroppenstedt R.M., Diaz P., and Pastor F.I.J.** 2005. *Paenibacillus barcinonensis* sp. nov., a xylanase-producing bacterium isolated from a rice field in the Ebro River delta. International Journal of Systematic and Evolutionary Microbiology 55: 935-939.
52. **Sanger F., Nicklen S., and Coulson A.R.** 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences USA 12: 5463-5467.
53. **Schulze E.** 1891. Chemische Berichte: 2277.
54. **Singh S., Dutt D., and Tyagi C.H.** 2011. Environmentally friendly totally chlorine free bleaching of wheat straw pulp using novel cellulase-poor xylanases of wild strains of *Coprinellus disseminatus*. BioResources 6: 3876-3882.
55. **Smith D.B., and Johnson K.S.** 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67: 31-40.
56. **Smith W.A., Thompson D.N., Thompson V.S., Radtke C.W., and Carter B.** 2009. Assessment of xylanase activity in dry storage as a potential method of reducing feedstock cost. Applied Biochemistry and Biotechnology 154: 287-301.
57. **Spiro R.G.** 1966. The Nelson-Somogyi copper reduction method. Analysis of sugars found in glycoprotein. Advanced Microbiology and Physiology 37: 1-81.
58. **Subramaniyan S., and Prema P.** 2002. Biotechnology of Microbial Xylanases: Enzymology,Molecular Biology,and Application. Critical Reviews in Biotechnology 22: 33.
59. **Tepfer M., and Taylor I.E.P.** 1981. The permeability of plant-cell walls as measured by gel-filtration chromatography. Science 213: 761-763.
60. **Uffen R.L.** 1997. Xylan degradation: a glimpse at microbial diversity. Journal of Industrial Microbiology & Biotechnology 19: 1-6.
61. **Viikari L., Kantelinen A., Sundquist J., and Linko M.** 1994. Xylanases in bleaching: From an idea to the industry. FEMS Microbiology Reviews 13: 335-350.
62. **Waeonukul R., Kyu K.L., Sakka K., and Ratanakhanokchai K.** 2008. Effect of carbon sources on the induction of xylanolytic-cellulolytic multienzyme complexes in *Paenibacillus curdlanolyticus* strain B-6. Bioscience Biotechnology and Biochemistry 72: 321-328.
63. **Wong K.K.Y., Tan L.U.L., and Saddler J.N.** 1988. Multiplicity of beta-1,4-xylanase in microorganisms - functions and applications. Microbiological Reviews 52: 305-317.
64. **Ye X., Zhu Z., Zhang C., and Zhang Y.H.** Fusion of a family 9 cellulose-binding module improves catalytic potential of *Clostridium thermocellum* cellodextrin phosphorylase on insoluble cellulose. Applied Microbiology and Biotechnology 52: 1-10.