EUSKAL HERRIKO U NIBERTSITATEA - UNIVERSIDAD DEL PAIS VASCOMATERIALEN FISIKA SAILA - DEPARTAMENTO DE FÍSICA DE MATERIALES

eman ta zabal zazu

Universidad del País Vasco   Euskal Herriko Unibertsitatea

Design of Ancestral Enzymes for Engineering Nanobiomaterials

Leire Barandiaran Larrea
PhD Thesis

Thesis supervisors:
Dr. Raul Perez Jimenez
Dr. David de Sancho

2019

1

SUMMARY

Nanomaterials have attracted scientific community attention as they have unique characteristics due to their increased specific surface area by volume. Two of the most interesting nanomaterials are nanocellulose and nanochitin, which are obtained from biomass, being biocompatible, biodegradable and renewable. These properties make them very promising biomaterials in biomedicine and pharmacology. But the obtention and modification of these materials still implies chemicals and processes that endanger their biocompatibility and sustainability, so a ne method for their non-invasive and sustainable obtention is needed. Enzymes showed up as an alternative to nowadays methodologies for these nanomaterial engineering, nevertheless implementation of enzymes in material science is still a challenge, as enzymes need to fulfill harsh condition environments and improve their hydrolysis yields in comparison with chemical methodologies obtention yields. To obtain biocatalysts with improved performance at industrial conditions several strategies such as *de novo* enzyme design and directed evolution have

been implemented. But we need to improve several catalytic properties of enzymes (catalytic efficiency, pH and T range.) in a cost-efficient manner. Here we used Ancestral Sequence Reconstruction (ASR), to successfully reconstruct biomass-degrading enzymes. The reconstruction of ancestral enzymes is carried out from an extant enzyme, the one we call Query.

We have improved a set of enzymes that can hydrolyze cellulose and chitin in a wide range of conditions. The activity of these reconstructed enzymes was compared to the Query. We reconstructed a β-1,4 Endo xylanase and a Lytic polysaccharide monooxygenase (LPMO) to, in combination with an ancestral endoglucanase (Anc EG), boost nanocellulose obtention yields from lignocellulosic biomass, and even modifying its structure. To consolidate the capability of enzymes to obtain nanomaterials from biomass we reconstructed a chitin degrading LPMO and obtain nanochitin. In all cases, ancestral enzymes showed considerably higher specific activities than modern enzymes.

LABURPENA

Azken hamarkadetan material berrien garapenean hobekuntza handia ekarri du nanoteknologiak, besteak beste, nanomaterialen sorreran. Azken horien artean nabarmenenak nanozelulosa eta nanokitina dira: lehena zelulosatik eratortzen da eta bigarrena kitinatik, eta biak biomasan dauden elementurik oparoenak dira. Biomasatik eratorritako nanomaterial hauek berezi egiten dituzten ezaugarri batzuk dituzte: biobateragarritasuna eta jasangarritasuna. Lehena garrantzitsua da biomedikuntzan ezartzen diren materialekiko osagarria delako; bigarrena, hondakin organikoetatik erator daitekeelako: nanozelulosa hondakin lignozelulosikoetatik, eta nanokitina krustazeoen hondakinetatik besteak beste.

Material lignozelulosikoa hiru osagaiz osatuta dago: lignina, hemizelulosa eta zelulosa. Lehena molekula aromatikoz sortuta dagoen material rekalzitrantea da; bigarrena, pentosaz eta hexosaz sortutako polisakaridoa; eta, azkena, D-glukosaz osatutako polisakaridoa. Lehen bien artean hidrogeno loturak egin eta Van der Waals indarrak sortzen dira, zelulosa kokatzen den sare konplexu bat sortuz. Azken hau isolatu ahal izateko, ezinbestekoa da beste bi osagaietatik askatu ahal izatea.

Hondakin hauetatik nanomateriala isolatzeko orain arteko prozesuek, kimikoak erabiltzeaz gain, materialen egitura kimikoa eraldatzen dute. Ordea, materialak organikoak izaki, entzimak erabil daitezke haien hidrolisirako eta, beraz, nanomateriala lortzeko.

Entzimek izaera oso selektiboa dute eta jatorriko baldintza naturaletan bakarrik egiten dute lan. Beraz, helburua industrian lan egingo duten entzimak lortzea bada, lanerako baldintza gogor horiek jasango dituzten entzima berriak sortu behar dira. Aipatzekoa da ikergai honek duen berebiziko garrantzia. Ez zioten, ausaz, iazko Kimika Nobel saria gai hau landu zuen F. Arnoldi eman. Hala ere, orain arte garatu diren entzima bidezko metodoek (directed evolution, rational design), garestiak izateaz gain, denbora gehiegi hartzen dute eta entzimaren ezaugarrietatik bakarra hobetzen dute. Gure laborategiak, ordea, azken urteetan eboluzioa ikertzeko erabili den ARS metodoak entzima horiek berreraikitzeko balio dezakeela erakutsi du (1.irudia).

Metodo horrek egungo espezie askoren proteina beraren sekuentziak alderatu eta haien arteko harremana aztertzen du, zuhaitz filogenetikoak sortuz. Horrela, eta método bioinformatikoa erabiliz, espezieen arbasoen proteina-sekuentziak jakin ditzakegu. Gehien bat bakteriekin egiten dugu

lan hori, denboran are atzerago egitea ahalbidetzen digutelako, prekanbrikora iritsi arte. Garai hartako bizi-baldintzak oraingoak baino askoz ere latzagoak ziren (itsas-tenperatura 60-70 gradukoa zen, adibidez), eta beraz, espezieak extremofiloak. Halako ezaugarriak dituzten proteinak birsortu ditugu gure laborategian, biologia molekularra erabiliz, eta industria-baldintzetan lan egin dezaketela frogatu.
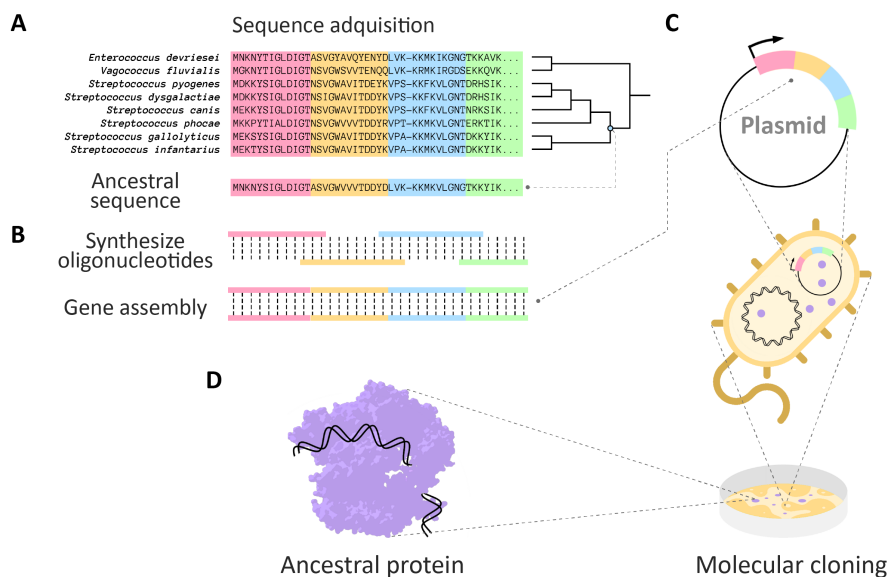
1. irudia: Antzinako enztima bat eraikitzeko burutu behar diren pausoak. A) ASR metodo bioinformatikoen bidez antzinako entzimaren sekuentzia lortu. B) Antzinako entzimaren genea sintetizatu. C) Entzimaren genea expesio bakteria batean sartu ingenieritza genetikoko teknikak erabiliz. D) Antzinako entzima laborategian purifikatu.

Horretarako, lehenik, endoglukonasa berreraiki eta nanozelulosa laborategiko sustratu idealetatik lortu ahal izan dugu. Jarraian, sustratu

errealetatik, hau da hondakin izan daitezkeenetatik, nanozelulosa isolatzeko Xylanasa eta LPMO (Lytic Polyssacharide Monooxigenase) entzimen laguntza behar dugu. Xylanasari dagokionez, *Bacillus subtillisen* Endo b 1,4- xylanasatik abiatuta, 3 mila milioi urteko xylanasa bat lortu dugu. LPMOaren kasuan, *Streptomyces viridisporusen* LPMO batetik abiatuta 2.8 mila milioi urteko aitzin-entzima bat (2.irudia). Ondoren, entzima berreraiki hauen ezaugarriak karakterizatu ditugu gure laborategian, jatorrizko entzimekin konparatuz, eta antzinakoek propietate termiko, pH- eta kinetiko hobeak aurkeztu dituzte: esaterako, horietako zenbaitek aktibitatea mantentzen dute tenperatura 12 gradu igota ere. Prekambriko garaiko ingurugiro baldintzekon bat eginez.

Hurrengo pausua nanozelulosa hondakin lignozelulosikoetatik lortzea izan da, BKP (Bleached Kraft Pulp) eta UBKP (Unbleached Kraft Pulp) sustratuetatik, zeintzuek, zelulosaz gain, lignina eta hemizelulosa ere baduten. Antzinako entzima berreraikiekin nanomaterial gehiago lortu dugu eta, gainera, LPMOak modu oxidatzailean lan egiten duenez, lehen aldiz nanozelulosa oxidatua lortu dugu, entzimen erabilera hutsarekin.

Amaitzeko, entzimen erabilera biomasan oparoen aurkitzen den bigarren materialarekin ere frogatu nahi izan dugu: kitinarekin. Kitina modu hierarkiko batean ordenatuta dagoen N acetyl glukosaminaz osatutako polisakarido bat da, modu hierarkiko batean ordenatzen diren zuntzak osatuz. Naturalki duen ordenazio horrek zuntz handietatik txikienetara kitina eskala txikitzen joatea ahalbidetzen digu, nanokitina lortu arte.
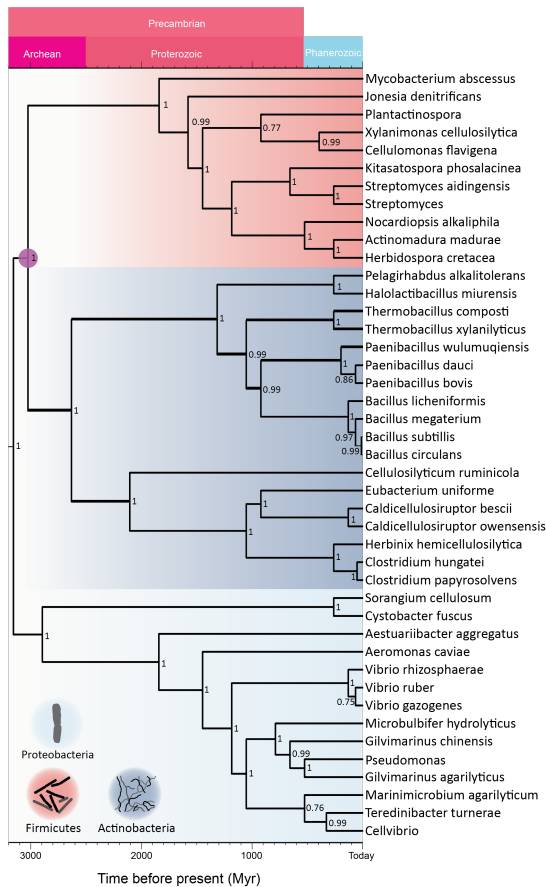
2. irudia: zuhaitz filogenetikoa. Zuhaitz honetan Aktinobakteria, Firmikutes eta Proteobakteria filoaen LPMO sekuentziak konparatzen ditugu zuhaitz filogenetikoa lortzeko. Geroago, nodoa aukeratu eta anztinako entzima eraikiko dugu.

Industrian prozesu azido, mekaniko zein termikoen bitartez egiten da prozesu hau, nanobiomaterial honen biobateragarritasuna zalantzan jarriz. Horregatik, tesi honetan, eta aurretik egindako lana bermatzeko, kitina hidrolizatzen duen aitzin-entzima bat berreraiki dugu: 2,8 mila milioi urtetako *Bacilus thurigiensis* bakteriaren LPMO entzima. Laborategian bere aktibitatea gaur egungo bere baliokidearekin alderatu

dugu eta berreraikiak, besteetan bezala, temperatura, pH eta kinetika hobea aurkeztu du, nanokitina ekoizpena handitzearekin batera.
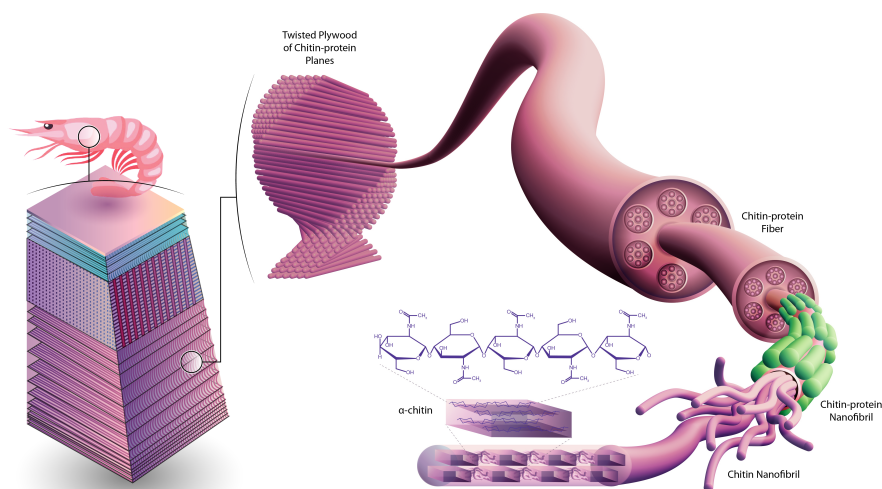
3. irudia. KItinaren egitura. Kitina, bigarren material oparoena biomasa krustazeoen exoeskeletoak aurkitzen da besteak Beste. Irudian bere ordenazio hierarkikoa nabarmentzen dugu, fibra handietatik, makrofibrak, fibra txikienera, nanofibrak.

Lan honetan, antzin-entzimek industrian lan egiteko duten balioa azpimarratu nahi izan dugu: hondakin lignozelulosikoetatik nanozelulosa atera daitekeela erakutsi dugu, xylanasa eta LPMO entzimen antzin-berreraikitzea eginez.

INDEX

PART I

Chapter 1 Introduction

In the past decades, nanotechnology has made great improvements in the development of new materials that have attracted the attention of researchers and industry. A prominent example is graphene, but many other materials such as carbon based fullerene, carbon nanotubes [1], 2D and Van der Waals materials [2], titanium nanoparticles [3], nanocomposites [4] and liquid crystals [5] [6], have also shared great interest. These materials are commonly used in many applications such as electronics [7], construction, aeronautics[8], [9], water treatment[10], and coating [11], among others. However, material scientists are facing new challenges, especially in bio-related disciplines that force them to incorporate biologically relevant materials, also in the nanoscale, that are known as nanobiomaterials. These materials have their origin in biological system and as such, often require treatment methods that also involve biosystems. Two notable examples of these materials are nanocellulose [12], [13] and nanochitin [14], [15]. Nanocellulose is obtained from cellulose and nanochitin from chitin. Luckily, cellulose and chitin are the first and second most abundant polymers on Earth,

respectively. These materials have been introduced in the past decade and active research is currently ongoing at a global level. These materials are the center of study of this thesis, where I have worked in their obtention, development and applications.

Nanobiomaterials have found and area of application in bio-related disciplines such as Biomedicine. Traditional biomedical applications have started benefiting from the use of nanomaterials in multiple areas, such as biosensors [16], [17], tissue engineering [18], [19], controlled release systems [20], drug delivery systems and nanocomposites used in implant design [21], [22]. These applications require materials with high biocompatibility and also physico-chemical properties that often need to approach living cells and tissues. This is still a challenge that can only be approached through multidisciplinary research involving biology, chemistry, physics, physiology, bioengineering and all related disciplines. The great effort required in such designs has remained a barrier that it is been overcome only in recent years.

Some of the most promising nanobiomaterials are the ones obtained from biomass due to their biocompatibility and easy tunability towards biomedical applications. The above-mentioned nanocellulose and nanochitin are currently investigated for such applications and both have been predicted to have a bright future in biomedical research. However, the obtention of these materials implies mechanical, thermal and chemical treatments that may affect not only their physico-chemical properties but also their biocompatibility, thus affecting their applications in biomedicine. This motivated us, in our laboratory, to develop a novel method for nanobiomaterial obtention based on enzyme catalysis. We use molecular evolution as the driving force for the design of enzymes with capabilities that go beyond catalysis, i.e., material transformations. In this thesis, I have contributed in developing these

methods for the design of enzymes capable of rendering nanobiomaterials such as nanocellulose and nanochitin.

The most abundantly raw material on Earth is lignocellulosic biomass from plants. Lignocellulosic biomass is composed of carbohydrate polymers (cellulose and hemicellulose), and aromatic polymer (lignin). These carbohydrate polymers contain different sugar monomers (five and six carbon sugars). Cellulose is the most abundant biopolymer; it is a linear homopolymer composed by D-glucose units bonded together by β-1,4-glycosidic bonds, where each unit is rotated 180º to the next one. The smallest unit is cellobiose, formed by two D-glucose units, with a size of 1,03 nm (Figure 1.1).
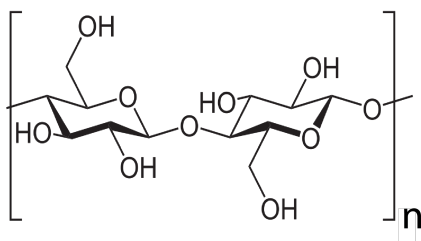
Figure 1.1. Cellulose chemical structure. Two glucose molecules bonded together by a β-1,4-glycosidic bond. Glucose is the repeating unit of cellulose polymer and cellobiose is form by two D-glucose units. Each unit is rotated 180° to the next and has three hydroxyl groups that made the polymer very reactive.

This monomer polymerizes forming linear cellulose chains and the hydroxyl groups, present in glucose monomers, help to form hydrogen bonds between chains. These hydrogen bonds can happen intramolecularly (between different cellulose chains), giving stiffness to cellulose, or intermolecularly, ordering chains to form a crystalline structure. Cellulose is a semi crystalline material, divided into crystalline

and amorphous fractions. The amount and orientation of inter and intramolecular hydrogen bonds vary depending on cellulose source and on the treatment used to its extraction. The native, and most abundant form of cellulose is type I cellulose, which shows a parallel organization of the chains. Cellulose II is the second most abundant polymorph, where chains have an antiparallel organization, allowing hydrogen bonds among neighbor's groups improving interlayer attraction forces.

In nature, cellulose is produced as an individual long chain that during the biosynthesis process is organized in a hierarchical structure forming the fibers. The polymeric chains are packed together through hydrogen bonding forming the elementary fibrils with a diameter of 3-5 nm and a length of 2-20 nm depending on the source [23]. At the same time, these fibers are ordered into bigger fibers by van der Waals forces and more intra and intermolecular hydrogen bonds. These microfibrils have a diameter around 30 nm and a length of several micrometers. Microfibrils are packed in larger structures called macrofibrils that together form the main cellulose fibers (Figure 1.2).
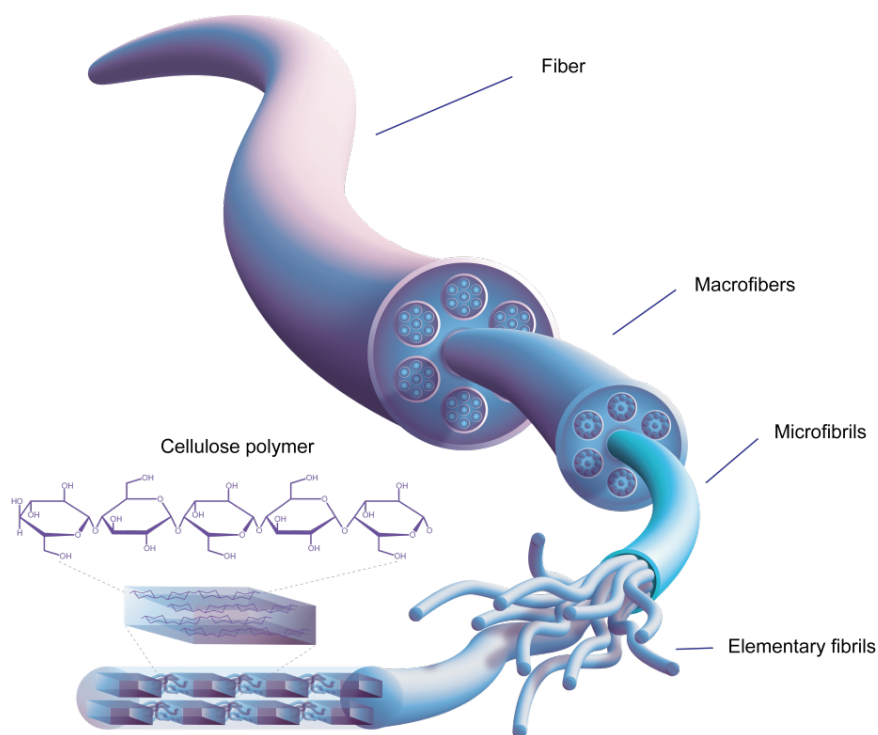
Figure 1.2. Hierarchical cellulose structure. Cellulose polymers are packed together forming elementary fibrils. These tiny fibers are aggregated by hydrogen bonding and van der Waal forces into microfibrils and these are assembled in macrofibers; the macrofibers are the last unit that forms the main cellulose fibers.

The hierarchical structure of cellulose from glucose chain to cellulose fibers permits a top down approach to obtain cellulose nanocrystals (CNC). Until now, the most common method to isolate CNC is by chemical treatment that includes sulfuric acid hydrolysis (H_2SO_4) [29], [30], or 2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO) oxidation [26]. This hydrolysis is carried out in combination with thermal or mechanical pretreatments. Some works have shown the possibility of using enzymes, but always in combination with other pretreatments such as mechanical

and acid [13][27][33][34]. However, **common methods to isolate nanocellulose from cellulose are a barrier to the full implementation of nanomaterials in biotechnological applications**. Learning from nature, we propose a full enzymatic treatment to isolate not only nanocellulose from biomass but to isolate nanomaterials with the only action of enzymes.

Microorganisms, like bacteria and fungi, use enzymes for the degradation of cellulose. Cellulases are the most common enzymes used by organisms to break down the β-1,4-glycosidic bonds of cellulose. For the total hydrolysis of cellulose into glucose monomers three cellulases are needed: Endoglucanase (EG), which cleaves β-1,4-glycosidic bond randomly producing oligomers with reducing ends and preferably attacking the amorphous region of the material; Exoglucanase (CBH), which breaks down crystalline cellulose degrading the previous reducing oligomer into cellobiose units; and β-glucosidase (BG), which hydrolyzes cellobiose into glucose.

The recent discovery of the Lytic Polysaccharides MonoOxygenases [30] (LPMO) has attracted the attention of the lignocellulosic research community, since it has been proved that they are able to break down recalcitrant polysaccharides such as cellulose, LPMOs are produced by fungi and bacteria, and are even found in some viruses and insects [31]. LPMOs are enzymes containing a cupper atom in their catalytic domain, and they have been described to produce oxidative cleavage on the glycosidic bonds. They were first studied for their activity in crystalline

chitin [32] and cellulose [33], but they also showed activity on other polymers such as hemicellulose [34] or starch [35].

The most described oxidation mechanism of LPMO to produce chain cleavage is the one made on C_1 or C_4 carbons of glucose (Figure 1.3), but in some cases it even oxidized C_6 [36]. C_1 oxidation produces soluble oligosaccharides with an aldonic acid in the reducing end and C_4 oxidation generates a ketoaldose in the non-reducing end [33], [37]. Since LPMO discovery, several reports of using them in combination with cellulases in biomass conversion have been published [38]–[40]. Addition of LPMO to cellulose cocktails boost their activity as LPMO helps cellulases to attack the crystalline cellulose [38].
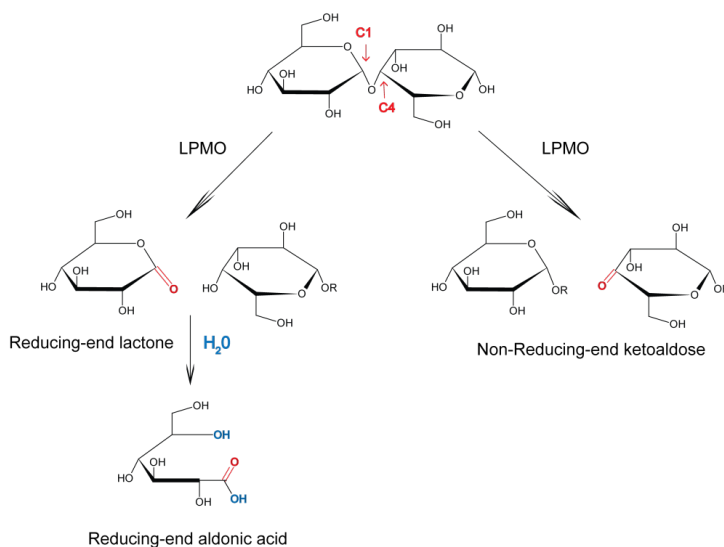
Figure 1.3. LPMO catalytic activity. LPMO can break the cellulose polymer by oxidation of C_1 or C_4 carbon in general. C_1 oxidation leads to soluble oligosaccharides with an aldonic acid in the reducing end and C_4 oxidation produces a ketoaldose in the non-reducing end.

Cellulose does not come alone in nature, it is embedded into a complex matrix form of lignin and hemicellulose, the other two components of lignocellulosic biomass. Hemicellulose is linked to cellulose by hydrogen bonding and covalently to lignin, giving stiffness to the lignocellulose matrix [44],[45],[43], [44]. In Figure 1.4 we see how cellulose (in blue) is trapped in a matrix of hemicellulose (in green) and lignin (in red). Cellulase and LPMO hydrolyze cellulose, but in order to get better yields of cellulose from raw materials, lignocellulosic biomass enzymes hydrolyzing lignin-hemicellulose matrix are needed [44]–[46]. In this work, we propose the degradation of hemicellulose in parallel to cellulose, to have more chances for cellulases to reach cellulose.
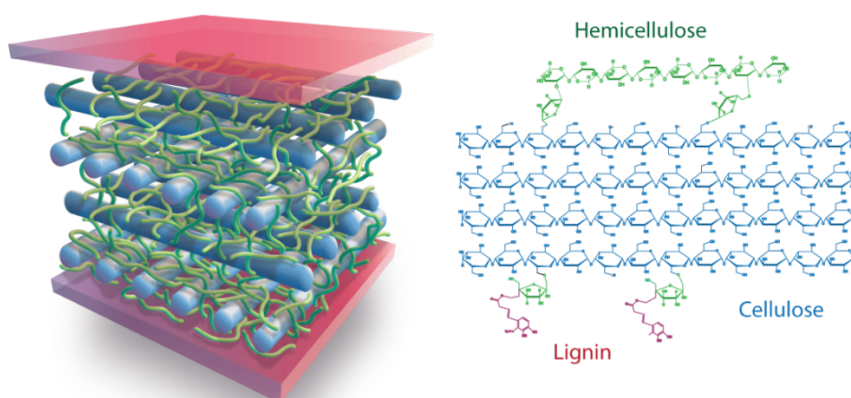
Figure 1.4. Lignocellulosic biomass. Cellulose is located in a matrix of hemicellulose and lignin. Hemicellulose is composed by monosaccharides of 5 or 6 sugars binded together by β-glycosidic bond, and lignin is formed by phenylpropanoid precursors. Hemicellulose is bonded to cellulose and lignin by hydrogen and covalent bonding producing a very recalcitrant and stiff

structure. This structure needs to be degraded in order to liberate cellulose fibers for further cellulose conversion into nanocellulose or sugars.

Hemicellulose structure consists of different carbohydrate polymers; the main polymer is xylan or glucomannan but has other sugars like xyloses, arabinoses, glucose, galactose, mannose, and sugar acids (Figure 1.5). Due to the short size of its lateral chains and its low molecular weight is easier to degrade than cellulose [38]. The hemicellulose can be extracted from biomass with sulfuric acid [47], alkaline pretreatment [48], steam explosion [49], and mechanical treatments [49], but due to its polysaccharide structure, it can also be removed by enzymatic hydrolysis. The most used enzyme is the Endo-1,4-β-xylanase that degrades the xylan polymer into oligosaccharides by breaking the β-glycosidic bonds between the monomers. Xylanase is produced by bacteria or fungi [50] and has been used in biomass degradation or in paper pulp bleaching processes [51][52].
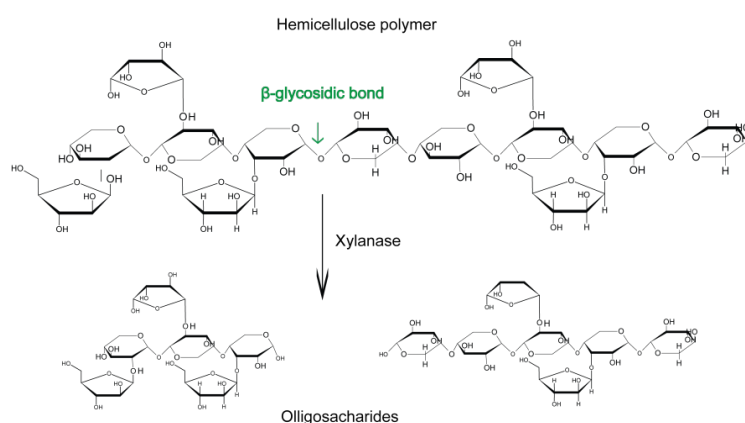
Figure 1.5. Xylanase catalytic reaction. Xylanase is able to break down the β-glycosidic bond between the xylan polymers and produces small oligosaccharides as product.

Lignin is the third main component of lignocellulosic biomass embedding cellulose. It is a physical barrier that protects cellulose fibers due to its structural complexity, high molecular weight, and insolubility. The linkage between lignin, cellulose, and hemicellulose are believed that inhibit the enzyme activity [53]. Chemical and physicochemical treatment can be used to disrupt lignin structure [54], [55]. There are two main families of enzymes that can depolymerize lignin too: peroxidases and laccase, produced mainly by the lignolytic white-rot fungi [56] and higher plant, but also found in bacteria [57].

In this thesis, we propose the addition of xylanase and LPMO to EG in order to hydrolyze biomass and get more accessibility to cellulose. In this way, together with other enzymes developed in our group at CIC Nanogune, we can isolate enzymatically nanomaterials from lignocellulosic biomass.

As we propose, enzymes seem to be a promising alternative for nanomaterial obtention from biomass. Nevertheless, they are adapted to their natural biological conditions, which often are far from industrial or even laboratory conditions required for biotechnological applications. These conditions often demand pH and temperature resistance that out of the reach of enzymes (Figure 1.6). This limitation has put the scope of biotechnology into enzyme engineering for industrial applications. In the past decades, researchers have focused on the improvement of enzymes properties, paying special attention to the enhancement of thermal stability, increase of specific activity, improvement of their substrate promiscuity and increase of the production rate [58]–[62].

The use of enzymes from extremophiles has been a solution for some industrial demands, producing satisfactory results, but there are situations in which even these extremophile enzymes need to be improved. Development in protein engineering research over the past decade has enable enzymes to be evolved in vitro for properties that favor the required process conditions, and also to obtain enzyme variants with altered substrate specificity or enantioselectivity [63], [64]. That is the significance of these techniques for our society, that last year (2018) Nobel Prize in chemistry was awarded to Prof. Frances Arnold for her groundbreaking contribution in protein engineering.
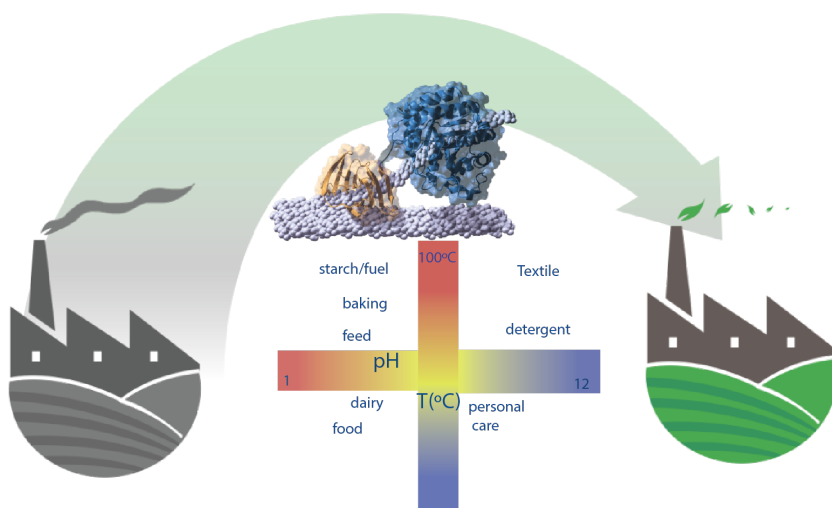
Figure1.6. Scheme of the industrial conditions that enzymes have to resist. In order to fulfill industrial conditions enzymes, need to be modified.

Several strategies ranging from rational and computational design to *de novo* enzyme design and directed evolution, have been implemented to improve enzyme biocatalysis [65], [66]. Rational design involves site directed mutagenesis introducing a specific amino acid into a target gene

[67]. The structure and function of the enzyme will be determinant on the selected amino acid, but detailed structural knowledge of an enzyme is often unavailable. DNA shuffling consists on the fragmentation of gene parents, followed by some PCR (polymerase chain reaction) cycles to obtain different mutants of a gene. Error-prone PCR inserts random mutants into any piece of DNA, based on the well-founded PCR [68]. Another technique, directed evolution in vitro [66], [69] mimics the process of natural selection to evolve proteins or nucleic acids towards a user-defined condition. Directed evolution technique creates thousands of variants of the enzyme of interest with different random mutations, generating a library of mutants. These mutants are exposed to the desired conditions and the variants that best perform under these conditions are identified and selected for commercial exploitation. The likelihood of success in a directed evolution experiment is directly related to the total library size, as evaluating more mutants increases the chances of finding one with the desire properties [70]. In any case, this protocol is time and money consuming, since the generation of a mutant library and subsequent screening has to be done experimentally.

Limitations of these engineered enzymes are still a serious barrier for the chemical industry. Yet no methodology seems to be able to enhance, for instance, temperature and pH operability, the expression level or the specific activity of enzymes, all at once. In this regard, developing a strategy able to improve catalytic properties of enzymes in a cost-efficient manner may revolutionize the biotechnology and chemical industry. In our group we work using ancestral sequence reconstruction (ASR) technique, that has been used to study evolution of genes, proteins and enzymes [71]–[74], it have been proved that ancestral enzymes

display enhanced thermal and pH stability, improved activity and higher expression level, chemical promiscuity and in some cases, all of this at once[71], [74],[75]–[79].

ASR (ancestral sequence reconstruction) is based on the evolutionary theory, which states that groups of organisms change over time so that descendants differ structurally and functionally from their ancestor. Nowadays, ARS combined with biophysical and biochemical state-of-the-art techniques allows us to study and compare features of extinct proteins and genes that are otherwise inaccessible [71]–[73].

In this regard, phylogenetic methods applied to genomic information have made possible to establish evolutionary relationships among different living organisms, including the possibility of inferring the putative sequences of the genes of their already extinct ancestors [80], [81]. Since 1837, when Charles Darwin sketched an evolutionary tree, the scope of phylogeny has been to reconstruct the correct relationship between organisms and estimate the time of divergence since they share a common ancestor. This is carried out combining the phylogeny (relationship between organisms) and the time of divergence, which together form a phylogenetic tree scaled to time. This way, Ancestral sequence reconstruction allows determining the sequences of genes and proteins of ancestors from extant species [82].

The resurrection of an ancestral gene involves five steps (Figure 1.7). First, we select modern species genes of the desired gene or protein; these sequences are the descended from the ancestral protein we want to reconstruct. We acquired from Internet databases. After selection we

align them, and the tree of their relationship is inferred. Phylogenetic methods such as Maximum likelihood are then used to infer the best estimate of the ancestral state for each sequence site given the present-day sequence data. Then DNA sequence for the ancestral protein is inferred. To express ancestral protein, we cloned into a plasmid that allows high-level expression, and the plasmid is then transfer to a high expression host bacterium in order to produce the ancestral protein. All these steps are explained in chapter 2 and chapter 3.
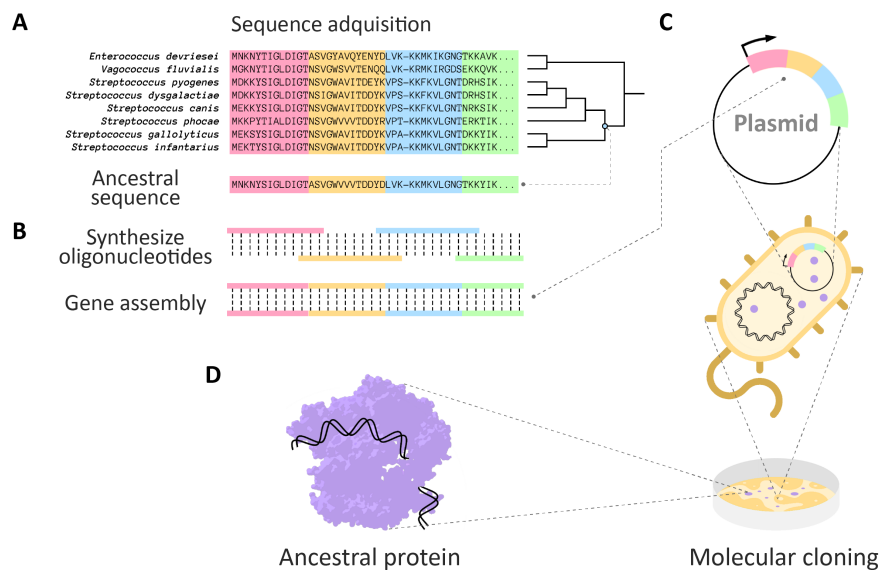
Figure 1.7 Ancestral sequence reconstruction. (A) Amino acids sequences from extant organisms are aligned and a phylogenetic tree is inferred with bioinformatics tools, (B) Sequences from the ancestral proteins are calculated from the nodes of the tree. (C) Expression plasmid with the gen from the ancestral protein and the bacterial host for protein expression. (D) Ancestral proteins expressed from the bacterial host.

The first ancestral protein was reconstructed in 1994 by Shindyalov et al [83]. Since them, numerous studies have used ASR, providing information about physiological and metabolic features [84] and about the environmental conditions that hosted ancestral organisms [73]. More recently, a study carried out by Perez-Jimenez et al. demonstrated the thermochemical evolution of thioredoxin enzymes covering a span between 4 to 1.4 Gyr, using single molecule force spectroscopy confirming a paleotemperature trend [74]. The deduction of the environmental conditions of different geological eras is an important application of ASR.

In our group we have implemented ASR as a biotechnological approach to improve the physico-chemical features of enzymes. Ancestral proteins and enzymes often show exceptional thermal stability, and also superior chemical properties due to the fact that our planet has been subjected to extreme environmental conditions throughout history and organisms and their molecules were adapted to these conditions. Moreover, it has been shown that ancestral enzymes are promiscuous, being able to work with lower selectivity and therefore with different types of substrates [70], [75], [84].

In our laboratory we reconstructed cellulases belonging the different ancient species: LFCA EG (Last Firmicutes Common Ancestor), LCCA EG (Last Clostridium Common Ancestor) and LACA EG (Last Actinobacteria Common Ancestor) and compared their activity (Figure 1.8) with modern *Thermotoga maritima* EG (Tm EG) and *Bacillus subtillis* EG (Bs EG). We demonstrated that the activity of cellulase follows a trend parallel to that of the cooling trend of the ocean, highlighting the tight relation between enzymes activity and

environmental temperature[85] as reported by Robert and Chaussidon [86], who demonstrated that the temperature on ancient oceans cooled down over 30 ºC during the period of 3.5-0.5 Gyr ago.The EG of the late Archean eon efficiently work in a broad range of temperatures (30-90ºC), pH values (4-10) and processes different lignocellulosic substrates showing processive activity and even doubling the activity of modern enzymes in some conditions [85].
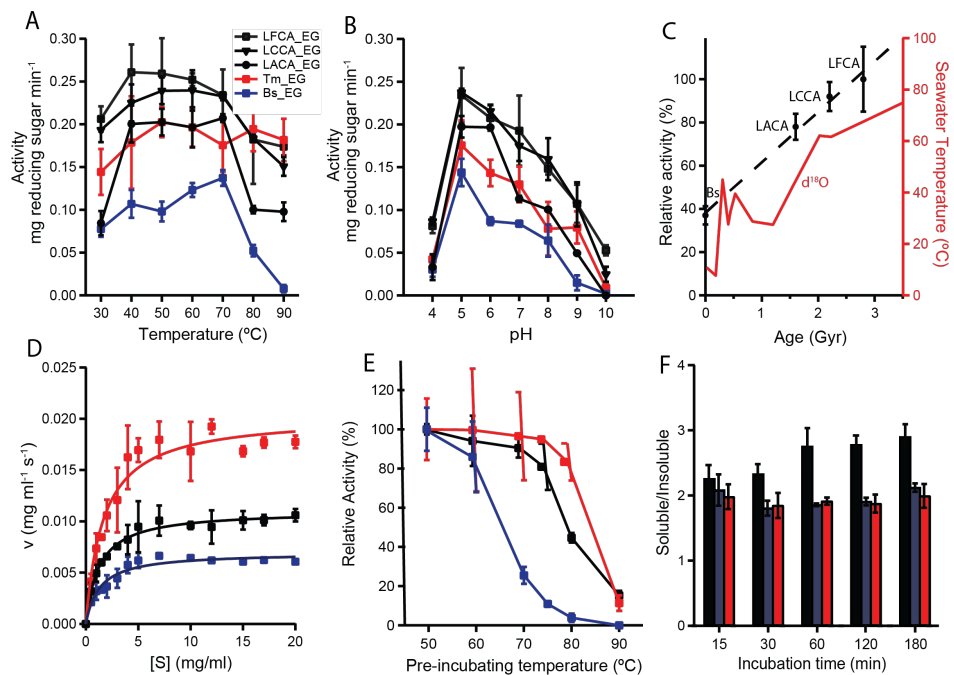
Figure 1.8. A published work of our laboratory showing A) activity assays for different ancestral endoglucanase as function of temperature: LFCA EG, LCCA EG and LACA EG against modern Tm EG and Bs EG at pH 4.8 and 5mg enzyme/g of glucan. B) EG activity as function of pH at 70 ºC. C) Relative activity of the ancestral cellulases and Bs_EG measured at 50 °C plotted vs. evolutionary time. The right axis, seawater paleotemperature trend as inferred from $\delta^{18}O$ isotopes in seawater cherts vs. time. D) Kinetics of enzyme catalysis

fitted to Michaelis-Menten equation. E) Pre-incubation experiments at different temperatures conducted for 30 min. F) Ratios of soluble to insoluble reducing sugars generated using PASC as a substrate by LFCA EG, Tm EG, and Bs EG,respectively.

Ancestral enzymes seem suitable for industrial applications, but nowadays there are relatively few studies in that field. These studies have shown that ancestral enzymes not only have a great interest from an evolutionary point of view but can also have multiple applications in areas such as bioengineering and biomedicine. These include studies about thioredoxin [74], lactamases [75] or cellulases [94]. Enzymes in industry have been well established as biocatalyst in various industrial processes and can be found in various products of daily life. Nevertheless, their application in material science is still modest, as their conventional role in industry did not imply material production [87], [88]. In our group, besides **giving a biotechnological scope to ASR, we proved that ancestral enzyme features make them suitable for material production and modification**.

Using an ancestral EG, we successfully isolated CNC from cellulose (Figure 1.9). This ancestral EG degrades cellulose chains in random locations to generate oligomers with reducing ends [85]; preferably attacking the amorphous regions of the fibers. We demonstrated that enzymatically obtained CNC kept cellulose native polymorph, whereas CNC isolated with other protocols does not maintain cellulose natural polymorph [89], where glucose chains are ordered in parallel. Chemical hydrolysis of cellulase, most commonly using sulfuric acid, leaves sulfur groups attached to the CNC surface. Similarly, the TEMPO treatment

renders oxidized CNC [90]–[92]. Enzyme hydrolysis emerges as an alternative given that CNC are obtained chemically pure, although usually enzyme yields are lower in comparison to chemical methods. In figure 1.8 and 1.9 we can observe the activity and hydrolysis capabilities of a the resurrected EG, which always shows better performance that commercial EGs.
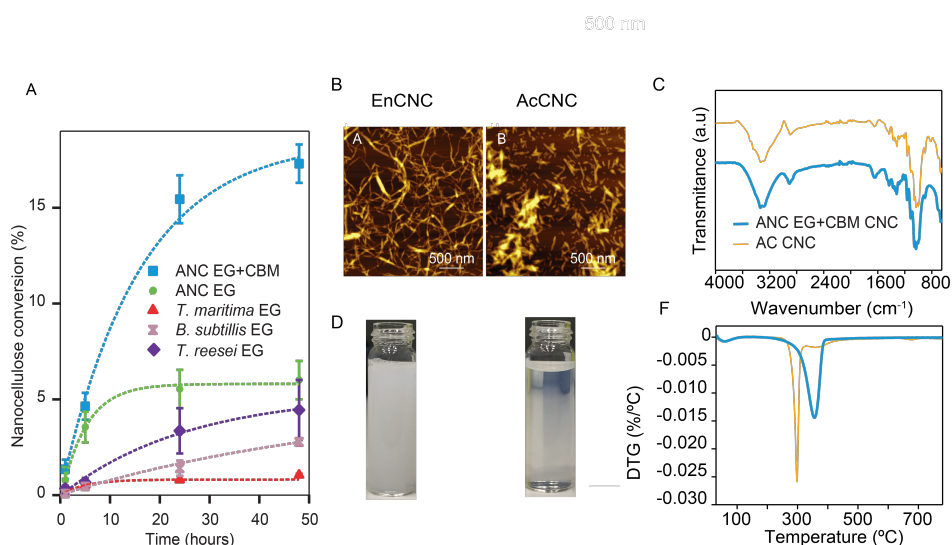
Figure 1.9: Enzymatic isolation of CNC from filter paper using ancestral EG+CBM. A) Nanocellulose total conversion percentage using different enzymes: ANC EG+CBM, ANC EG, T maritime EG, B. subtillis EG and T. reseei EG. ANC EG+CBM shows 3 times better yields from the next enzyme after 24 hours. B) AFM images of CNC obtained using ANC EG +CBM (EnCNC) and acid (AcCNC). C) FTIR of CNC obtained by ANC EG +CBM (blue) and Avid (yellow). D) Pictures of CNC suspension in water, on the left ANC EG+CBM CNC that has a better particle dispersion on water than, on the left AC CNC that show poor dispersion on water due to the negative charges of its surface. F) Derivative from thermogravimetric curve of CNC obtained by ANC EG+CBM has better thermo stability than CNC obtained by Acid.

In this thesis, I consolidate the process of enzymatic isolation of nanomaterials by reconstructing an ancestral xylanase and an ancestral LPMO to help nanocellulose isolation from lignocellulosic biomass. Moreover, as LPMO is an oxidative enzyme, we think that the action of this enzyme in nanocellulose may modify its surface; hence we can obtain an oxidized CNC enzymatically without the need of chemicals such as TEMPO.

To reinforce the capability of enzymes in nanomaterial obtention, in this thesis I have used enzymes to isolate nanochitin from the second most abundant biopolymer on Earth: chitin. It is a linear homopolymer composed of acetyl -D-glucosamine linked by β-1,4-glycosidic bonds (Figure 1.10). It is a main component of cell walls in fungi [93], the exoskeletons of arthropods, such as crustaceans and insects [94], mollusks [9] and the scales of fish among others. It is biocompatible, biodegradable and bio absorbable, with antibacterial and wound-healing abilities and low immunogenicity [95]. Moreover, it is transparent, flexible and has high thermal and mechanical resistance. These has made of chitin and its derivatives a very valuable feedstock for industry with a broad range of applications in different fields such as food technology [96], material science [97], agriculture [98], wastewater treatment [99], drug delivery systems, tissue engineering or nanobiotechnology.

The structure of chitin can be compared to that of cellulose. But, unlike cellulose, chitin can be a source of nitrogen as well as carbon [100]. It was discovered in 1811 by French professor Henri Braconnot but it was in 1843 when Lassaigne demonstrated the presence of nitrogen on chitin [101]. So, the presence of the amyl group in the C_6 carbon of the sugar

ring instead of the hydroxyl group creates a bigger dipole in the molecule than in glucose, as nitrogen will be positively charged whereas the double bonded oxygen of the carboxyl is negatively charged. This dipole increases hydrogen bonding in chitin and thus, chitin strength. Therefore, chitin or chitin-combined composites resulting structures are very strong.
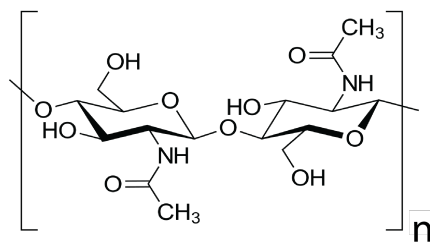
Figure 1.10. Chitin chemical structure. b) N-Acetyl-glucosamine is the repeating unit of chitin; the second most abundant biopolymer on earth, instead of a hydroxyl group contains an amine group. Each unit is rotated 180° to the next and has three hydroxyl groups that made the polymer very reactive.

Chitin chains are packed together by intra and inter molecular hydrogen bonding and van der Waals forces, about 18-25 chitin molecules arrange to form crystalline nanofiber structure of about 0.1-10 nm length [102]. Similarly, nanofibers form inter and intra molecular hydrogen bonds and van der Waals forces to pack them into bigger fibbers that in nature are wrapped with proteins, this chitin-protein fibers are 10-100 nm long and about 20 nm diameter [103]. Chitin nanofibers are arranged into bigger fibers of about 100 nm-10 μm length forming horizontal planes that are packed together in a twisted plywood structure (Bouligand) [104]. At the macro scale cuticles consists of three layers: endocuticle, exocuticle and

outer thin waxy layer, the epicuticle. The hierarchical chitin structure is shown in Figure 1.11.
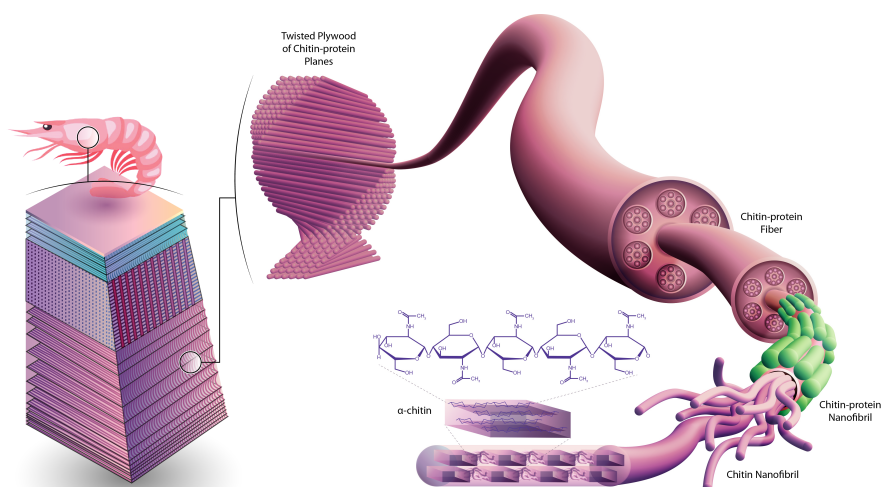
Figure 1.11: hierarchical structure of chitin. From the N-acetyl glucosamine molecules bonded by glycosidic bonds creating chitin chains, to nanofibrils (pink), nanofibrils wrapped in protein (green), microfibrils creating horizontal planes in a twisted plywood structure. At the macroscale three cuticles consists of three layers: endocuticle, exocuticle and outer thin waxy layer, the epicuticle, from a shrimp exoskeleton.

The hierarchical organization of chitin from N-acetyl -D-glucosamine chains to chitin fibers permit a top down approximation to obtain nanoparticles. As mentioned before, nanoparticles from biomass is focusing research attention because of its extraordinary capabilities like biocompatibility, renewability; it is sustainable, nanometer size, high

aspect ratio, flexible, electrical thermal and mechanical properties.

Chitin is predominantly present in fibrillar crystalline material. Based on X-ray diffraction data and infrared spectroscopy (FTIR) native chitin can organized in three crystalline forms [105]: α-chitin, β-chitin and γ-chitin depending on their origin. α-chitin is the most abundant existing in crabs, lobsters, krill and shrimp shells, insect's cuticles, and fungal yeast cells walls. It has and antiparallel orientation of the N-acetyl-glucosamine chains, with strong intermolecular bonding, having a crystallinity higher than 80% [106]. β-chitin, present in squid pens and tube worms [107] is the second most abundant chitin polymorph in which the molecules are arrange in parallel. Due to this molecular arrange intermolecular interactions are weaker than those in α-chitin. γ-chitin is the less abundant chitin polymorph in which the molecules are arranged in both parallel and anti-parallel fashion.

The amorphous domains of chitin can also be removed under certain conditions such as acidosis to give rise to crystalline nanochitin (CNCh) in bibliography is also called chitin whiskers, CHW). They can be obtained by hydrochloric hydrolysis [108], [109],[110], [111], TEMPO mediated oxidation [106], gelation, ultra-sonication [112], electro spinning [113] or mechanical grinding treatment [114]. However, HCl hydrolysis is the most common method and the others are usually carried out in combination of acid treatment [115], [116].

A

B
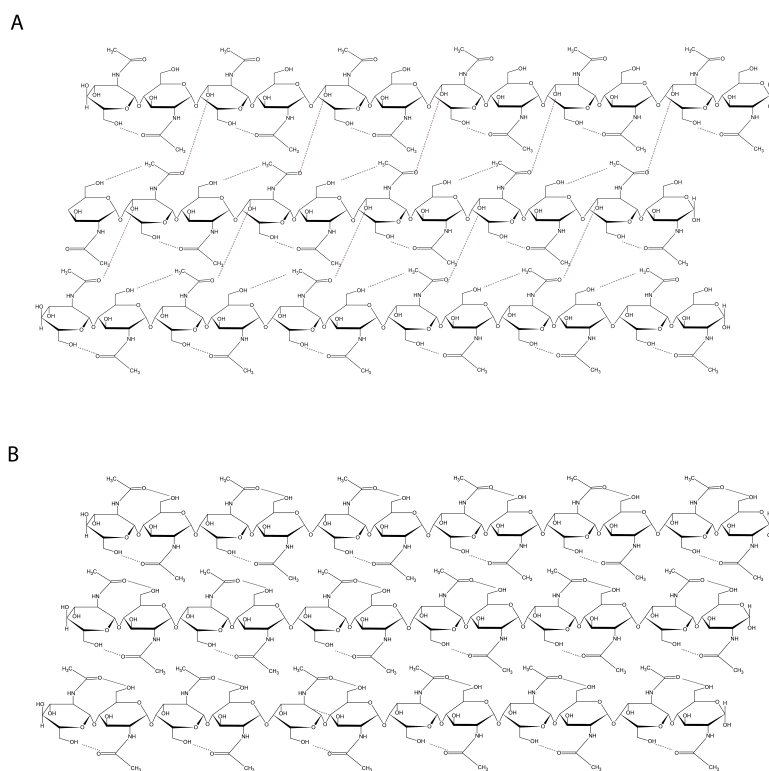
Figure 1.8: chitin polymorphs A) α-Chitin, chains are oriented antiparallel favoring intra and intermolecular bonds. B) β-Chitin, chains are oriented parallel.

There are different enzymes that catalyses chitin hydrolysis. For instance, the endochitinase activity is defined as the random cleavage at internal positions in the chitin chain. The exochitinase activity is defined as the progressive action starting at the non-reducing end of chitin with release of chitobiose or N-acetyl glucosamine [117]. But, this enzyme acts in colloidal chitin and there are no reports of chitinases acting in crystalline chitin [118], [119].

CNCh has very interesting features regarding biomaterial technology, such as easy availability, nontoxicity, biodegradability, low density plus

the ones mentioned before. For this reason, they have been widely used as substitutes to inorganic nanoparticles in reinforcing polymer nanocomposites [120], In packaging for food [121], agriculture, water treatment or nanopore filter. Biomedical applications have also been reported as tissue regeneration and antibacterial material. In spite of its increasing studies about CNCh the research and development of CNCh area goes slower than CNC [116].

Organic nanoparticles such as CNC and CNCh show many advantages over traditional nanocomposites and together are being studied for multiple applications such as agriculture, medicine, tissue engineering, cellular scaffold, and conductive bio-inks. It has been reported that films cast with both nanomaterials effectively enhanced the mechanical performance of the film with high transparency and flexibility [122].

In this thesis, we will reconstruct biomass degrading enzymes in order to obtain better yield and structurally modified CNC by the only action on enzymes. Besides we will reconstruct an ancestral LPMO and, for the fists time, will enzymatically isolate CNCh from chitin. We will compare the activity of all the reconstructed enzymes in this work with their modern homologous.

In order to characterize these new nanomaterials and compare them with commercially available CNC and CNCh we will use X-ray diffraction, FTIR spectroscopy, NMR and AFM to study morphology.

Objectives

The main objectives of this thesis were the reconstruction of ancestral enzymes for the degradation of lignocellulosic biomass and chitin and to confirm the capability of these enzymes to isolate nanocellulose and nanochitin from biomass.

1. Reconstruction of an ancestral Endo 1,4-beta xylanase from *Bacillus subtillis* and a Lytic polysaccharide monooxygenase (LPMO) from *Streptomyces viridosporus* for the degradation of lignocellulosic biomass. Reconstruction of a LPMO from Bacillus thurigiensis for the degradation of chitin.

2. Characterization of the reconstructed enzymes and their extant homologous. We use molecular biology tools to first express the genes, then we purify the protein at our laboratory and finally we perform different activity assays for each enzyme.

3. Nanocellulose isolation from complex substrates from lignocellulosic biomass. Help an ancestral endoglucanase

(EG) reconstructed in our laboratory to isolate nanocellulose from complex lignocellulosic substrates by adding the reconstructed xylanase and LPMO in this work. Study the effect of an oxidative enzyme (LPMO) in the enzymatic method on the obtained nanocellulose.

4. Nanochitin isolation from α-chitin using ancestral LPMO. We want to consolidate the capability of ancestral enzymes for the obtention of nanobiomaterials from different biomass substrates. The characterization of this enzymatically produced nanochitin and its comparison with a nanochitin obtained with HCl hydrolysis.

This work has been carried out to consolidate the potential of blending two different fields in nanobiomaterials engineering: Material science and Enzyme chemistry. This work was made in CIC Nanogune under the direction of Raul Pérez Jiménez and the co-direction of David de Sancho from DIPC (Donostia Physics Centre) with Nanogune findings.

PART II

Chapter 2: Phylogenetic analysis and ASR

In this chapter, I will describe the methodology to build a phylogenetic tree. Moreover, I will explain how to track the evolutionary history of genes to obtain information of extinct species using ancestral sequence reconstruction (ASR).

2.1 Introduction

Ancestral reconstruction is the extrapolation back in time, using statistic methods, from measured characteristics of individuals to their common ancestor. It can have non-biological applications such as the evolution of phonemes or vocabulary of ancient languages [123], oral traditions of extinct cultures [124] or ancestral marriage practices [125]. But in the context of evolutionary biology it can be used to study the evolutionary relationship among individuals, populations or species, in **phylogenetics** [126]**.** In this scenario, either protein or nucleic acids will be the characters to compare. All these data come from extant species that have been sequenced. In order to recover the most accurate ancestral state ASR relies on a realistic statistic model of evolution. In order to determine the route of the evolution, the genetic information already obtained through methods such as phylogenetic is used[127].

Linus Pauling and Emilie Zuckerkandl in 1963 suggested the possibility of reconstructing ancestral proteins from analyzing and comparing extant proteins. 25 years later the paleochemistry they suggested became true. Thanks to the increasing known about gene and protein sequences, more feasible statistical methods and improvements on genetic engineering techniques. This idea began in 1955, when Frederick Sangerce started developing techniques for sequencing the primary structure of proteins. This achievement lead Zuckerkandl and Pauling to propose that, based on the amino acid sequence of extant proteins, it is possible to infer the phylogeny of that protein and the sequences of all the common ancestors, including the earliest point of the tree [128], [129]. But, it was in 1971 by Walter M.Fitch when the first algorithm for ancestral reconstruction

using the principles of maximum parsimony was developed [130]. Nowadays, by means of the so- called ancestral sequence reconstruction, we are able to reconstruct ancestral biological macromolecules; genes and proteins.

2.2 Theory of phylogenetic analyses

To organize the diversity of life, during all the evolution, there are two main tools: taxonomy and phylogeny. Taxonomy classifies organisms in three domains (Bacteria, Achaea and Eukaryotes). Whereas phylogeny is an estimation of the organism evolutionary history where a most recent common ancestor splits to form two species.

We need to determine the route of the evolution to reconstruct an ancestor, for that, we will use the genetic information already obtained through phylogeny, constructing a phylogenetic tree. This will be a hypothetical tree or cladogram showing the order in how species are correlated between each other. In a rooted phylogenetic tree, each node with descendants represents the inferred most recent common ancestor of those descendants. The very first ancestor of the tree, and the starting point for the branches, is the last universal common ancestor (LUCA). The end of the branches or the terminal nodes are the extant species. At the end, all the species and, therefore, all the evolutionary lines converge in LUCA (Figure 2.1)
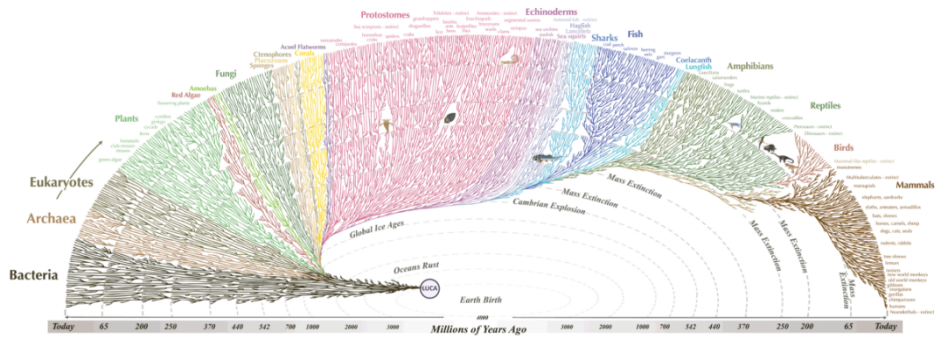
Figure 2.1. Phylogenetic tree of all the living species with a Universal Common Ancestor. This Figure wsource is Evogeneao, the tree of life.

2.2.1 Methods

As the computing power has grown exponentially, literature about different methods for analyzing phylogenetically structured data has also grown. First, a method to infer the tree came up: Maximum parsimony, where a set of extant sequences are used to minimize the amount of mutations that are necessary to match the available data. Since maximum parsimony method was described a lot of changes have happened. In 1975, David Sankoff added a cost to the mutation [131] optimizing the existing algorithms. In 1989, thanks to this work David. L. Swofford could develop the first phylogenetics program called PAUP [132]. At the same time, the exponential increase of the computing power made the implementation of much more complex algorithms possible, such as maximum likelihood approaches [133]–[135]or Bayesian methods [135]

The method will change depending on the question they seek to answer. Therefore, many different methods have been described.

2.2.1.1 Maximum parsimony

This method attempts to reconstruct ancestral states by minimizing the number of trait changes between the ancestor and the present descendants. To infer the tree, it uses a set of extant sequences minimizing the amount of mutations that are necessary to match the available data. Due to its low computational costs and high efficiency for large datasets and when ab initio phylogenies are needed [136] this method is very useful. But, parsimony methods impose general assumptions that are not always valid such as the assumption that all character states are equally likely to occur, which is unrealistic.

2.2.1.2 Maximum likelihood

The maximum likelihood approach for phylogenetic analysis seeks the tree and associated evolution model parameters that maximize the probability of producing the given set of genomes. In other words, this method assumes that the ancestral states are those, which are statistically most likely, based on the input data, the extant genomes. So, maximum likelihood provides probabilities of the sequences given a model of their evolution on a particular tree.

To use a model of evolution we need to understand that not all events are equally likely to take place. So, the more probable the sequences given the tree, the more the tree are preferred. Anyway, this does not mean that they need to happen just because they are more likely to occur. It can happen that the event with less probability occurs. In those cases, maximum parsimony may actually be more accurate because it is more willing to make large, unlikely leaps than maximum likelihood.

Maximun likelihood is really reliable in reconstructing character states. But, is not good giving accurate estimations of the stability of proteins as overestimated, since it assumes that the input proteins where the more stable and optimal [137].

In Maximum likelihood Markov process models the evolution of the sequence. In this model we assume that all mutations are independent [138] and the likelihood of the model is calculated from a sum of intermediate probabilities of the nodes for the tree. This way, in each ancestral node, the likelihood of the descendants is computed to obtain a maximun posterior probability (equation1):

$$[1] Lx = \sum_{Sx \in \Omega} P(Sx) \left(\sum_{Sy \in \Omega} P(Sy|Sx, txy) Ly \sum_{Sz \in \Omega} P(Sz|Sx, txz) Lz \right)$$

Where node x is the ancestor of y and z. Si represents the sequence of the i-th node, tij refers to the branch length from i to j. Ω is the set is the set of all the possible combinations (for example, the four nucleotides or the 20 basic amino acids). The ultimate objective of the reconstruction is to search for the best configuration in all the nodes before to obtain the maximum likelihood for their ancestor in a given tree. Therefore, the aim of an ancestral reconstruction is to find the assignment for all x internal nodes that maximizes the likelihood of the input data for the given tree.

Each descendant obtains a likelihood value. In order to find the most probable evolutionary lineage to the common ancestor two different conventions have been proposed (Figure 2.2):

- Joint reconstruction: where one can consider the probabilities of all the descendants for a certain ancestor and calculate the joint combination with the maximum likelihood.

- Marginal reconstruction: where instead of calculating the global likelihood, one can successively select the most likely ancestor for every node.
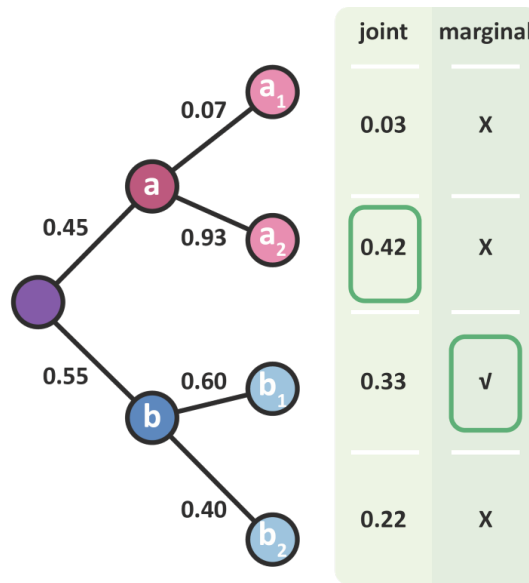
Figure 2.2. Scheme of Joint and marginal reconstructions for a phylogenetic tree. Pale pink and pale blue circles are wxtant species and dark pink and dark blue circles their extants. Purple circle is Last Universal Common Ancestor.

2.2.1.3 Bayesian

It is a statistical inference method in which Bayes theorem is used to seek the probability for a hypothesis in where the more information or evidence the better. It employs both the likelihood of the experimental data and a prior knowledge about the possible solutions. Thus, the aim of ancestral sequence reconstruction is to obtain the posterior probabilities

for every internal node of a known tree. Moreover, the posterior probabilities can be combined with the posterior distributions over the parameters for a given evolutionary model and the structure of all possible trees, giving the following applications of Bayes theorem (equation 2):

$$[2]\ P(S\,|D,\theta) = \frac{P(D\,|\,S,\theta)\,P(S\,|\theta)}{P(D\,|\,\theta)}$$

θ Represents the phylogenetic tree and the evolutionary model, S corresponds to the ancestral states and D is the experimental data. $P(D\,|\,S,\theta)$ Is the likelihood of the experimental data that could be computed, $P(S\,|\theta)$ refers to the prior probability of an ancestral node for a known tree and model and $P(D\,|\,\theta)$ is the probability of a known tree and model integrated for all possible ancestral states. This formula refers to the empirical Bayes approach. This estimates the probabilities of several ancestral nodes for a given tree and model of evolution [139].

2.3 Experimental Phylogenetic Analysis

Phylogenetic analysis only needs of informatics tools. Different webpages and Software's are used during the process. In this section we find a description of the process.

The reconstruction of the three ancestral lignocellulosic enzymes was carried out following the next steps: (1) Selection of the sequence of interest and its BLAST, (2) Alignment of the selected sequences, (3) Construction of the phylogenetic tree and (4) Reconstruction of the ancestral sequence.

Figure 2.3: Example of an alienation of selected sequences to construct the tree. On the right part the organisms name, ordered by taxonomy and aligned. From the alignment a phylogenetic tree is inferred with its corresponding nodes. One of these nodes will be the selected ancestral node from where the ancestral gene is inferred.

2.3.1 Select extant sequences: Uniprot and NCBI

The sequence of the protein of interest for the ancestral reconstruction must be in a genome or sequences databases and it is called QUERY. There are many different free online protein sequences databases but there are two that allow us to find and select other sequences having the query as a starting point: Uniprot UniProt and NCBI NCBI [140].In this thesis we searched for the query at Uniprot, because it is as a catalog of proteins, with all the information (M_w, structure, function, etc...). And then, we search for homologous sequences at NCBI (BLAST, because it allows us to search for specific taxa's or exclude them besides giving a value for the query cover.

2.3.1.1. BLAST

After the identification of a suitable Query at Uniprot we run a BLAST (Basic Local Alignment search Tool) in NCBI. This tool allows us to find regions of similarity between biological sequences. The program

compares protein sequences to sequence databases and calculate the statistical significance. Moreover, we can select different parameters to accurate our search (Table 2.1) after entering our query sequence.

Table 2.1: Possible settings and their meaning for a protein BLAST at NCBI

SETTING	MEANING
Database	NCBI Blast allows us to search sequences among different databases.
Search set	Allows to select specific taxon to only search for them or either exclude them.
Algorithm	BlastP compares a protein query to a protein database. A new accelerated version came up Quick BLASTP, is very fast and works best if the target percent identity is 50% or more.
Max tar. Seq.	Allows choosing the maximum number of aligned sequences to display.
Word size	The length of the seed that initiates an alignment. We choose the maximum.
Expected threshold	Expected number of chance matches in a random model. The bigger this value is, the more unlikely to be significant a match. For our sequences, that has between 20 and 54 KDa an E-theshold of 0.001works.
Matrix	Assigns a score for aligning pairs of residues, and determines overall alignment score. We choose BLOSUM [141] matrix that is based on the frequency in which that substitution is known to occur among consensus blocks within related proteins.

Running the Blast can take several minutes depending on the complexity or length of the query sequence or the parameters we set. When it is finished, the sequences appear ordered from the one with the best score to the one with the worse score. This way, it is possible to choose sequences with **identity ≥ 40%** and **query cover ≥ 90%** easily. We select an identity equal or bigger than 40% because below this identity we consider them as different. Moreover, the query cover must be equal or bigger than 90% to ensure the selected sequences with a certain identity has it in the appropriate regions with reference to the query.

All selected sequences can be downloaded in different formats. Regarding the following steps we download them in FASTA format (a text base format for representing nucleotide or amino acid sequences).

2.3.1 Multiple sequence alignment: MUSCLE

The file generated from the selected sequences of the BLAST, saved as FASTA format, can be opened in MEGA (Molecular Evolutionary a genetic Alignment), a very versatile tool for phylogenetic and molecular evolution. MEGA is software, with powerful visual tools, that can crate multiple sequence alignment, infer phylogenetic trees, estimate rates of molecular evolution, infer ancestral sequences and much more phylogeny related applications.

Over the last years, many algorithms have been developed for the multiple sequence alignment. ClustalW and MUSCLE [81], [142] are the most popular. Both, with a heuristic approach, use progressive alignment method where those with the best alignment score, are aligned first. Then progressively more distant groups of sequences are aligned until a global

alignment is obtained. Moreover, MUSCLE (integrated at MEGA software) is the fastest at aligning large sequences [143], [144] and the method we choose in this thesis for MSA.

When the FASTA format document with our sequences of interest is loaded, a non-aligned sequence map will appear at the interface. To run with MUSCLE algorithm, one should select all the sequences and click on the "align with MUSCLE" icon. A pop-up window will appear showing the different parameters that one can change before computing the algorithm (Table 2.2).

Table 2.2: Alignment parameters at MEGA program for running MUSCLE.

PARAMETER	MEANING
Gap opening penalty	By increasing this value, the gaps become less frequent in the alignment
Gap extension penalty	The bigger the value is, the shorter the gaps are in the alignment. Terminal gaps do not penalize.
Max memory in MB	The algorithm sets a computational memory upper limit by default set, the use of all the computer resource is avoided.
Max interaction	Sets the maximum number of permitted iterations.
Clustering method	Sets the clustering method used in the first two interactions.
Max. Diagonal Length	Maximum length of diagonal of the matrix made by aligning the sequences.

After running MUSCLE algorithm and all interactions are finished, we see an aligned sequences map on the interface. Sometimes regions of ambiguous alignment or with gaps in several sequences can be removed manually, directly at the alignment, or using GBLOCKS [145]. GBLOCKS is software that gets rid of poorly aligns positions and divergent regions of an alignment of DNA or protein sequences. Each residue shows a different color depending on its biochemical properties. If a residue is conserve in all sequences an asterisk is show on top of the residues. Finally, the alignment needs to be saved as nexus file for the following steps.

2.3.2 Computing a phylogenetic tree

To construct the phylogenetic tree from the alignment we choose software that uses Bayesian analysis of molecular sequences using MCMC (Markov Chain Monte Carlo): BEAST (Bayesian Evolutionary Analysis Sampling Trees). It uses MCMC to average over tree space, this way the tree is weighted proportional to its posterior probability. Moreover, it is orientated towards rooted, time-measured phylogenies inferred using strict or relaxed molecular clock models.

From Beast package we will use BEAUti, BEAST and TREE Anotator but we need to download BEAGLE library first in order to run the program, TRACER to track the phylogenetic tree and FIG tree to draw the tree.

2.3.2.1. BEAGLE

BEAGLE is a high-performance library able to run the core calculations at the heart of most Bayesian and Maximum Likelihood phylogenetic package. With BEAGLE, processors can be used in parallel, such as those in graphics processing units (GPUs) found in many PCs.

2.3.2.2. BEAUti

BEAUti (Bayesian Evolutionary analysis Utility) is a graphical user interface to generate a suitable input data for BEAST (xml). Here we can design the analysis by setting the evolutionary model or options for MCMC from the aligned sequences. It has a user-friendly interface with several tabs and allows modifying these parameters.

Table 2.3. parameters that can be changed at beauty to generate the desire best input data.

SETTING	MEANING
Partitions	Allows to load sequences that were not in the initial pool and to make partitions within this pool.
Taxa	Clusters selected taxa into subgroups. There is the possibility to force these subgroups to be monophyletic.
Tips	Permits data sampling of individual taxa.
Traits	Sets the phenotypic trait analysis.
Sites	Allows selecting the substitution model and the site heterogeneity.
Clocks	Permit to choose the clock model the different clock models use the mutation rate of biomolecules to estimate when they

diverged.

Trees	Set the tree prior.
States	Allows reconstruction the states of all the ancestors or only certain subgroups.
Priors	Sets the prior distribution for the subgroups, the gamma shape parameter, the proportion of invariant sites parameter (a parameter that assumes which residues do not mutate)
Operators	Switches on or off the some of the parameters set in previous tags.
MCMC	Sets the MCMC values for phylogeny computing

When we have set all the parameters, we push "Generate BEAST file" button and an XML command file is generated, ready to be run in BEAST.

2.3.2.2 BEAST

BEAST takes as input an XML command file and returns as output log files, files that record a sample of the states that the Markov chain encountered. Here we can analyze the evolutionary rates, divergence times, population sizes and tree topologies. An examination of this output is needed to determine whether the Markov chain has been run for long enough to obtain accurate estimates of the parameters. Tracer carries out this post-analysis. Running BEAST with a graphical user interface will open the following dialog box:
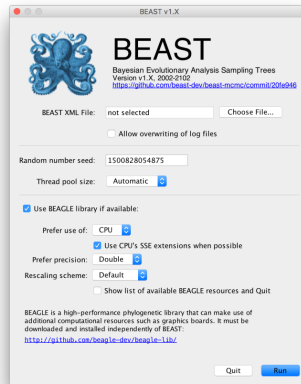
Figure 2.4. Graphical user interface (GUI) application of BEAST where .XML

file is loaded in order to infer the tree.

This graphical interface is user friendly and let you set different parameters. The following table shows all the parameters that can be set at BEAST

Table 2.4. Parameters at the Grafical User Interface (GUI) used by BEAST.

PARAMETER	MEANING
BEAST XML file	The name of the selected BEAST XML file.
Allow overwriting of log files	By default, BEAST does not overwrite existing log or trees files to protect the data. .
Random number seed	Like all computer software, BEAST uses a pseudo-random number generator. The random numbers are used to generate the random starting tree, to propose new states in the MCMC tree and to accept or reject these proposals.
Thread pool size	Specifies the number of threads that BEAST will use to divide up the computation amongst cores and processors.
Use BEAGLE	This option and the rest of the dialog box is used to tell BEAST

library   to use the BEAGLE library and control the options for its use.

Once al the parameters are set by pressing RUN button the analysis will start until completion.

2.3.2.3 Tracer

Tracer is graphical interface (Figure 2.5) that allows monitoring and analyzing the MCMC output carried out in BEAST. Once the log file that corresponds to the analysis of the phylogenetic computation is opened, the name of the log file lo the traces that it contains and many parameters related to MCMC analysis will appear on the left side of the interface. These parameters are pondered by their Effective Sample Sizes (ESSs). A low ESS means that the trace contained a lot of correlated samples and may not represent the posterior distribution well. It is advisable to run BEAST again until ESS reaches a value higher than 100.
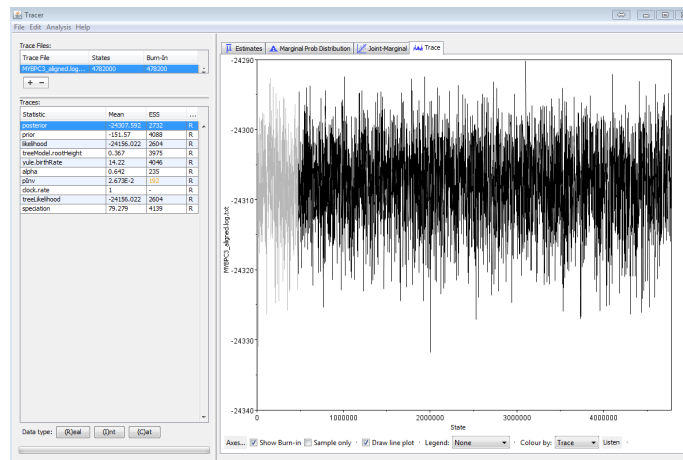
Figure 2.5. Tracer graphical interface. The evolution of posterior probability can be visualized in this picture. The data in grey is excluded for the calculation of average.

2.3.2.4 Treeanotator

This tool permits us to select a tree among all the trees that has been generated during the BEAST process. This tree contains the generated posterior probabilities of the nodes in the target tree. The program includes the options in Table 2.5.

Tabla 2.5: table explaining the different options for setting in TreeAnotator and their meaning.

PARAMETER	MEANING
Burnin	Number of samples that will be discarded at the start of the run.
Posterior probability limit	Specifying a limit as starting point in PAUP. They will only be calculated for the nodes that in the target tree have a bigger posterior probability than the specified.

Target tree type	In "maximum clade credibility" the node height and rate statistics will be merge on the tree posterior probabilities on its (n-2) internal nodes. The "user target tree" option summarizes the tree statistics on a user-specified tree.
Node heights	To set the way the node heights are summarized on the target tree.

2.3.2.5 FigTree

After the selection made by TreeAnotator among all the trees generated by Beast Fig tree gives an easy and intuitive interface to visualize and modify the phylogenetic tree. We can select between different options at the left side of the screen such as time scale, node labels, node bars, scale bar, scale axis or legend.

2.3.2.6 TreeGraph

With TreeGraph is possible to both visualize and edit a phylogenetic tree. Besides, it supports several (visible or invisible) annotations (e.g. support values) for every branch or node. These annotations can be imported from Nexus tree files or text files containing tabulated data (e.g. exported from a spreadsheet program).

2.3.1 Reconstruction of ancestral sequences

A consensus phylogenetic tree has been obtained following the previous steps. Now we have to infer the ancestral sequence from the tree. Although there are different software for this purpose, in this thesis we

use PAML. As we mention previously, the ancestral sequences are inferred using maximum likelihood algorithm.

2.3.3.1. PAML

Phylogenetic Analysis by Maximum Likelihood (PAML) is a package of programs for phylogenetic analysis. It can analyse DNA or protein sequences using maximum likelihood. To reconstruct an ancestral sequence in this thesis we use **codml** programme. To run **codml** we need a sequence data file, a tree file (in Newick format), a matrix file (usually Jones matrix) and the control file before running the programme.

We have to make sure that all the input files of the alienation (.fas) and the tree file (.tre) have the exact same sequences and names of sequences. Moreover, we have to make sure that the model we choose at Beauty to construct the tree is the same that we input in codml, in this case Jones model. Once everything is set up, we run the executable file and the program will start computing the algorithm to calculate the ancestral sequences.

When the process is over, the program creates an output file (.rst) with all the information of the process, posterior probabilities and the joint and marginal protein reconstructions. For this thesis we choose marginal

Figure 2.6. Screenshot of control file of condml in PAML. There we change our input .Tree and .Fas files in order to reconstruct nodes sequences.

Figure 2.7. PAML program running the calculations for the ancestral reconstruction of our protein.

Chapter 3: Experimental methods

This chapter explains the experimental procedures followed during this thesis. The experimental works of this thesis are divided in three different sections. The first section explains the molecular biology techniques and procedures applied for obtaining the different proteins used during this work. The second section, following the results from the first section, shows the experiments realized in order to characterize each enzyme. Finally, the third section explains the production and characterization of the nanomaterial obtained with our proteins.

It is important to note that all the experiments carried out during this work were applied equal and at the same time to the ancestral enzyme and its homologous extant one (Query).

3.1 Molecular biology

To carry out this research it has been necessary to design, express and purify the ancestral sequences obtained from the ancestral reconstruction and their Query. During this section all the procedures followed, from the genetic engineering to the purification, are explained.

3.1.1 Genetic engineering procedures

The genes obtained from the ancestral reconstruction step are purchase from Life Technologies, which are provided inside of the commercial plasmid and with the sequences of the genes codon-optimized for their expression in *Escherichia coli*. Besides, we need to order all the genes with specific restriction sites, that later will allow to introduce our gene in a desire expression to produce the enzyme.

3.1.1.1 Commercial plasmid amplification and digestion

The first step is to amplify the commercial plasmid containing the gene of interest. We use *E. coli* XL1-Blue competent cells (Agilent technologies) with approximately 50 ng of DNA following the protocol of Agilent Technologies. The transformation takes places in ice, in order to maintain the competent character of the cells. Then, the bacteria's are grown in SOC media (Invitrogen) for 1 h at 37Cº and 250 rpm to later culture them over plates made of LB, Agar, Carbenicillin (Fisher) at 37ºC. The commercial plasmid is resistant to Carbenicillin antibiotic. After letting them grow overnight, single columns are picked and grown in LB containing 0.1 µg/mL of Carbenicillin for 12-16 h at 37Cº and 250

rpm. Cells were then pelleted by centrifuging them to 4000 rpm for 10 min at 4Cº and plasmidic DNA is extracted from them with QIAprep© miniprep kit (QIAGEN). Finally, plasmidic DNA is eluted and stored in nuclease-free water (Fisher) and concentration is determined using the Nanodrop 200L spectrophotometer (Thermo Scientific).

When the commercial plasmid, containing our genes, is amplified the digestion is done to bind our gene sequence to the desire plasmid. All our genes have a *BamH*I restriction site in their 5´end and a *Kpn*I site in their 3´end. These sites can be open using restriction enzymes purchase in Thermo Scientific and following the manufacturer´s Fast Digest protocol. The final digestion volume is adjusted to 50 µL and incubated at 37Cº for one hour. The digestion products are screened in a DNA agarose gel (1%) in TAE buffer. All DNA agarose gels are run using Bio Rad agarose electrophoresis equipment for approximately 90 min. using UV light we can observe two bands in the gel, one corresponding to the commercial plasmid and another one corresponding to our protein gene. For each gene we will have different sizes. Finally, the gene is purified with a DNA extraction kit from *Thermo Scientific* following the usual protocol. Concentrations are calculated using *Nanodrop 2000L* system.

3.1.1.2 Expression plasmid-protein sequence constructs ligation

When the digestion is successfully done, the genes encoding our ancestral and query enzymes have to be ligated to a high-efficiency bacterial expression vector with compatible cohesive ends. The restriction enzymes produce cohesive ends so the insertion of the

purified gene into the plasmid was bidirectional and could produce the chimeric protein of interest. In this work, we use two different plasmids:

PQE80 L

This plasmid was a kind gift from Professor Julio Fernandez´s lab at Columbia University. It contains ampicillin resistance gen for selection and the protein expression is induced with Isopropyl β-D-1 thiogalactopyranoside (IPTG). IPTG is a chemical compound that inhibits the suppression of the lac inhibitor in the lac operon that controls the expression of the protein of interest. The plasmid also has an N-terminal *his-tag* that attaches a 6-histidine chain to the recombinant proteins and helps the partial or total protein purification by using metal affinity resins.
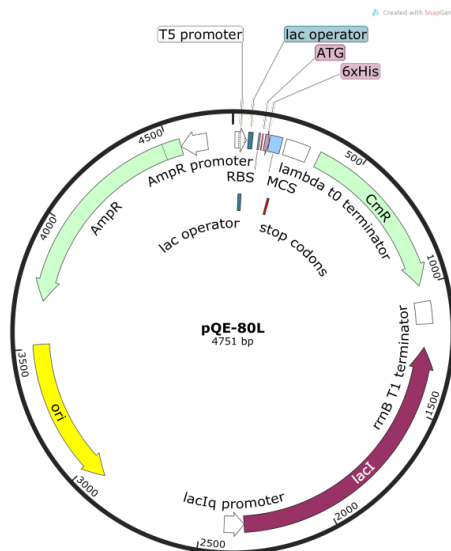
Figure 3.1. pQE80 expression plasmid. This plasmid contains carbenicillin resistance gen for selection and a lac operator for controlling protein overexpression by IPTG.

71

pETT28a

plasmid contains kanamycin resistance gene for selection and the protein expression is induced with IPTG. This plasmid has the *his-tag* that attaches a 6-histidine chain to the recombinant proteins in the C-terminal.

Invitrogen´s T4-DNA ligase protocol is used for the ligation process between the plasmids and the genes. The mol ratio between the amount of plasmid vector and gene inserts is 3:1. Ligations were incubated overnight at room temperature. Diluting the ligation with 5 times deionized water stops the process.
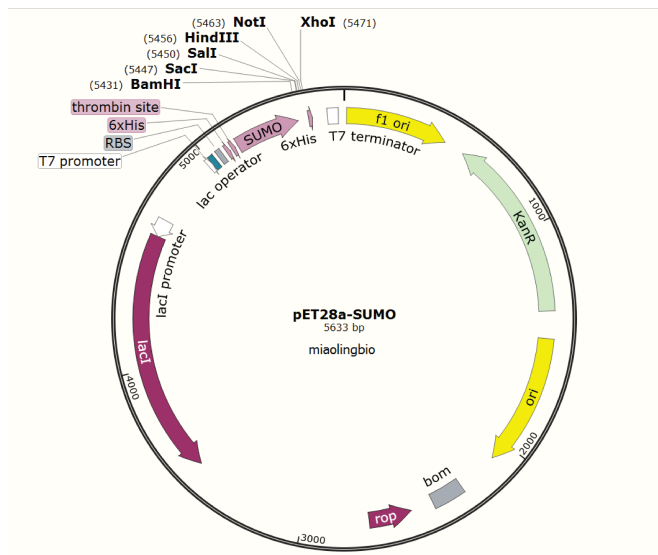
Figure

3.2.PET28a+ expression plasmid. This plasmid contains kanamycin resistance gen for selection and a lac operator for controlling protein overexpression by IPTG.

3.1.1.3 Amplification of recombinant plasmid

We transformed *E.coli* XL1-Blue competent cells with 50 ng of DNA using the manufacturer protocol [146], the bacteria were grown in super optimal broth medium (Invitrogen) for 1 hour at 37ºC and 250 rpm of agitation. The cells were plated onto LB agar plates with carbenicillin (for xylanases and endoglucanases) or kanamycin (for LPMOs) and were grown overnight at 37ºC.

We took individual colonies to grow them in 5 mL of LB media with the corresponding antibiotic for 16 hours at 37ºC and 250 rpm of agitation. After incubation time, cells were harvested by centrifugation at 14000 rpm, 10 minutes and at 4ºC. We discarded the supernatant and plasmids were extracted from the pellets using the commercial QIAprep® miniprep kit (QIAGEN). The plasmid was eluted using miliQ water and we measured the DNA concentration using Nanodrop 200L spectrophotometer (Thermo Scientific).

3.1.1.4 Expression test of recombinant proteins

Once we have our gene of interest, we transformed the construct into BL21 cells to test the protein expression, like with XL1 Blue, around 50 ng of the plasmid were used. We selected several colonies from the LB agar plates, with their corresponding antibiotic, that were grown overnight. Grew each colony in 10 mL LB media with carbenicillin or kanamycin until 0.6 O.D. Then, divide the 10 mL in two of 5 mL and to one of it we added 1 mM IPTG for protein over expression induction and the other one as a control. We let them grow overnight at 37ºC and

agitation of 250 rpm. The next day we took 1 mL of both tubes to pellet the cells by centrifugation (4000 rpm, 4ºC and 10 min). The supernatant was discarded, and the cells resuspended in 20 µL of extraction buffer (50 mM sodium phosphate an 300 mM NaCl at pH 7) and 20 µL of 2xSDS page sample buffer solution for the denaturalization and charging of the protein in acrylamide electrophoresis gel separation. Samples were again centrifuged for 30 minutes at 14000 rpm for separating the soluble proteins to the debris. The samples were incubated at 95ºC for 5 minutes to finally run the SDS-PAGE gel to separate proteins by size. We loaded 20 µL of the samples into the polyacrylamide gels (stacking 4%, resolving 8%) and the electrophoresis was run for 1hour and 30 minutes in a Bio-Rad electrophoresis system, at 120 V in running buffer solution (Bio-Rad). After the gels run to the end, they were cleaned in deionized water for 30 minutes. Then, the proteins are stained with Bradford solution (Thermo Scientific) for 30 minutes and washed with washing solution (10% ethanol, 10% acetic acid and 80 % miliQ water). We took an image of the gels in the Kodak image Station 4000R.

In the gel we could observe the protein overexpression in the IPTG treated samples, as an intense band. The colonies with the best expression were used for large-scale protein production.

3.1.2 Protein production

The colony that showed more expression was selected for bigger production and purification. First, 1 mL of the selected inoculum without IPTG is added to 1L of LB media with 0.1% of the corresponding antibiotic. To maintain the ability of overexpression of the bacterial pLys

system we add 0.05 % of chloramphenicol. We incubated the culture at 37ºC until the OD of the culture reached 0.6, in this moment we added 100 mg/mL IPTG to the culture for protein overexpression induction. We add 0.5 mM Cu^{2+} to LPMO culture, as LPMO has a conjugated Cu^{2+} atom [147]. We let bacteria produce the protein overnight at the specific temperature for each protein and 250 rpm. We have seen that all, ancestral and extant, proteins are expressed at the same conditions except the LPMO from *S.viridosporus* and the reconstructed ancestral one needed to be expressed at 20ºC in order to have an optimum expression.

After the expression of the protein from our host bacteria we removed the medium from the bacteria by centrifugation (4000 rpm at 4ºC for 30 minutes) and we collected the pellets. The pellets were resuspended in 16 mL of extraction buffer (50 mM sodium phosphate and 300 mM NaCl at pH 7), we added 160 µL of protease inhibitor (Merk Milipore) and 100 mg/mL of lysozyme (Thermo Scientific) incubating the mixture 30 minutes at 4ºC with gentle agitation. We followed the lysis of the bacteria with a chemical lysis by adding 1.6 mL of 10% Triton solution (Sigma Aldrich) for the destabilization of the bacterial membrane, 80 µL of 11 mg/mL DNAse I (Invitrogen) and 80 µL of 1 mg/mL RNAse A (Ambion) for DNA and RNA enzymatic degradation respectively, and 160 µL of MgCl2 1M (Sigma Aldrich) to improve the DNAse and RNAse catalytic activities. The mixture was incubated for 10 min at 4ºC with gentle agitation again. We followed the chemical lysis with a mechanical cell lysis using French press (G. Heinemann HTU DIGI-F Press). The mixture was introduced in the press chamber and lysed at 18000 psi during approximately 30 minutes. We obtained a darker solution that we centrifuged in an ultracentrigation system (Beckman

Coulter Avanti J-26 XOI) at 33000 G at 4ºC for 90 minutes to separate the soluble proteins from the debris.

The pellet was discarded and the soluble proteins in the supernatant were mixed with His Trap column cobalt or nickel affinity resins (thermos Scientific) for 1 hour at 4ºC with gentle agitation. The pQE-80L and pET28 a+ plasmids have a His Tag that gives to the expressed protein a 6 histidine tag that has affinity to His Trap columns, making our protein to bind them. After the incubation of our protein of interest with the resin we washed the resin with the attached protein 3 times with extraction buffer. We place the resin in a proper column with a filter to appropriately elude our protein from the resin. For that, we use an elution buffer containing imidazole (50 mM sodium phosphate an 300 mM NaCl at pH 7 and 150 mM Imidazole).

3.1.3 Protein purification

All the proteins were purify using a similar protocol, with small changes regarding the stability of each protein.

3.1.3.1 Xylanase purification

Both xylanases, the ancestral and the query, were expressed as explained in section 3.1.2. The Protein expression at the host bacteria was made at 37ºC overnight. After the expression of the protein and lysis of the bacteria we bind our protein to a His trap cobalt column. Both xylanases contain a His tag composed of 6 consecutive histidines in the N terminus of the construct, which poses the ability to specifically bind to the cobalt

affinity column. This binding was later eluded with elution buffer containing imidazole (50 mM sodium phosphate and 300 mM NaCl at pH 7 and 150 mM Imidazole). We eluted in fractions of 0.5 mL and measure the protein concentration on the Nanodrop at A280, we eluted until the protein concentration decrease considerably. Fractions containing the mayor protein concentration were mixed and the others discarded.

We followed the purification by using a ÄKTA pure fast protein liquid chromatography (FPLC) system (GE Healthcare) with a Superdex 200 column of 30 cm in diameter (GE Healthcare). In this step we separate proteins in terms of sizes and using the most appropriate buffer for our proteins. In the case of xylanase we used phosphate buffer 50 mM at pH 7. Fractions are collected in terms of their UV absorption, above 100 mAU. The fractions of the peak appearing

We followed the purification by using a ÄKTA pure fast protein liquid chromatography (FPLC) system (GE Healthcare) with a Superdex 200 column of 30 cm in diameter (GE Healthcare). In this step we separate proteins in terms of sizes and using the most appropriate buffer for our proteins. In the case of xylanase we used phosphate buffer 50 mM at pH 7. Fractions are collected in terms of their UV absorption, above 100 mAU. The fractions of the peak appearing after 80 minutes of elution are the ones containing pure xylanase.
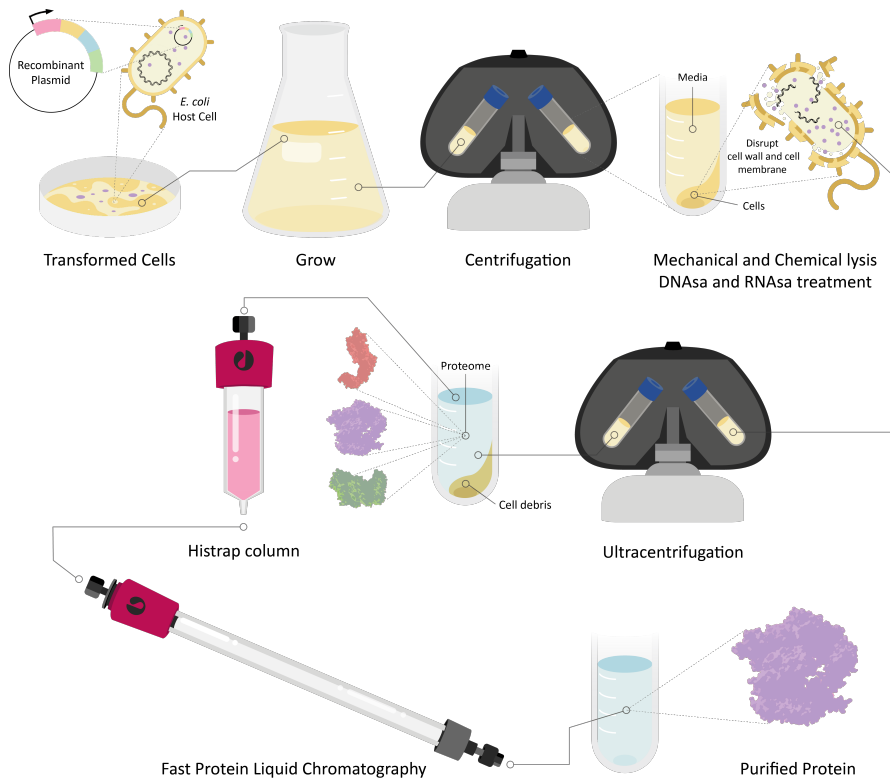
Figure 3.3. Overview of protein purification protocol in our laboratory, the expression plasmids pQE80 or pT28A+ with the cloned genes, were transformed in the competent E.coli BL21 bacteria. The bacteria were grown in 1L of LB media overnight at 37ºC and 250 rpm agitation. The bacteria were separated from the culture medium using centrifugation and then pellets were resuspended in extraction buffer. Then, bacterial cells were put down to chemical and mechanical lysis to disrupt the bacterial membrane to liberate the proteins. Soluble proteins were separated by ultracentrifugation and were mixed to a trap column as first purification step. To conclude the purification we used a size exclusion chromatography step to obtain purified proteins.

3.1.3.2 C- LPMO purification

Both LPMOs, the ancestral and the query, were expressed as explained in section 3.1.2. The expression of the protein form the host bacterium was made at 37 ºC overnight. After the expression of the protein and lysis of it form the bacteria we bind our protein to a His trap Nickel column. Both LPMOs contain a His tag composed of 6 consecutive histidines in the C terminus of the construct, which poses the ability to specifically bind to the cobalt affinity column. Elution is made as specified in section 3.1.3.2.

Purification by using an ÄKTA pure fast protein liquid chromatography (FPLC) system (GE Healthcare) with a Superdex 200 column of 30 cm in diameter (GE Healthcare) was done using Acetate buffer 50 mM at pH 4. Fractions are collected in terms of their UV absorption, above 100 mAU. The fractions of the peak appearing after 85 minutes of elution are the ones containing pure LPMOs.

3.1.3.3 Ch-LPMO purification

Both LPMOs, the ancestral and the query, were expressed as explained in section 3.1.2. But this time the expression of the protein form the host bacterium was made at 20 ºC overnight. After the expression of the protein and lysis of the bacteria we bind our protein to a His trap Nickel column. Both LPMOs contain a His tag composed of 6 consecutive histidines in the C terminus of the construct, which poses the ability to specifically bind to the cobalt affinity column. Elution is made as specified in section 3.1.3.2.

Purification by using an ÄKTA pure fast protein liquid chromatography (FPLC) system (GE Healthcare) with a Superdex 200 column of 30 cm in diameter (GE Healthcare) was done using acetate buffer 50 mM at pH 4. Fractions are collected in terms of their UV absorption, above 100 mAU. The fractions of the peak appearing after 85 minutes of elution are the ones containing pure LPMOs.

3.1.3.3 Endoglucanase purification

We used two Endoglucanases (EG); a *B.subtillis* EG and Ancestral EG (Anc EG), reconstructed in our laboratory. The purification of endoglucase was carried out first by temperature. We incubated the supernatant after the centrifugation of the last step of the lysis. The incubation is made at 50 ºC for 30 minutes. In this step, proteins that are not thermostable precipitate. These proteins are discarded as pellet after centrifugation (4000 rpm, 4ºC and 20 minutes). The remaining purification steps are made as specified in section 3.1.3.2

3.2 Enzymatic Characterization assays

Each protein has its own specific activity, in this part of the chapter I will explain the activity they perform and the biocatalytic characterization assays we did in our laboratory. In one hand we characterized all proteins separately (each ancestor with their query) and in the other hand we have characterized their ability to work in combination: the synergy between our reconstructed enzymes. In the first step we characterize the proteins that have been reconstructed in this work whereas in the second part we use the reconstructed enzymes in this thesis and a reconstructed endoglucanase on a previous thesis in our laboratory.

3.2.1 Xylanase characterization

Xylanases (EC.3.2.1.8) is a class of enzymes that degrade the linear polysaccharide Xylan into Xylose. So, Endo β-1, 4- xylanase are mainly responsible for the hydrolysis of β-1, 4- bonds in plant Xylan, the main component of hemicellulose [148]–[150] (Figure. 3.4).
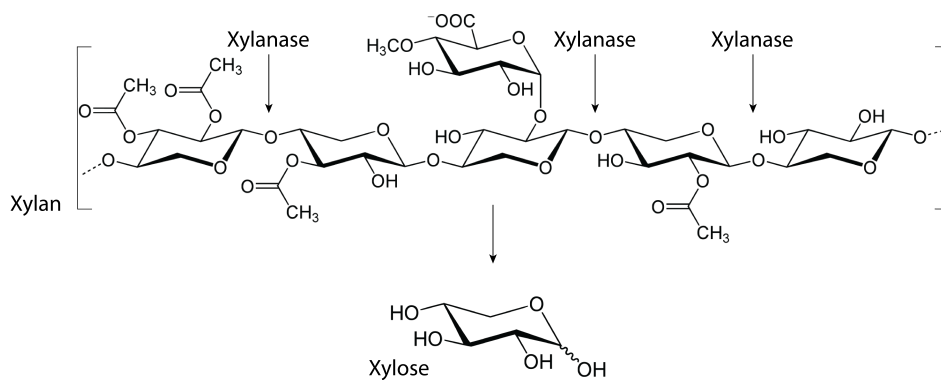
Figure 3.4. β-1,4- xylanase cleaving reaction to xylose. This enzyme cleaves β-1,4- bonds randomly, resulting in reduced xyloses.

The most commonly used method for the measurement of xylanase activity is commercial enzyme preparation is the 3,5-dinitrosalicylic acid (DNS) reducing sugar method [151] using Wheat Arabinoxylan (Megazymes) as substrate. Besides, in order to proof the promiscuity of our reconstructed xylanase we measure the kinetics of the ancestral xylanase and the *B.subtillis* xylanase with carboxymethylcellulose (CMC).

3.2.1.1 Wheat Arabinoxylan assays

We use commercial Wheat Arabinoxylan (Megazymes) to characterize Anc Xln and *Bt* Xln. Enzymatic hydrolysis with the substrate produces reduced xyloses. These reduced xyloses react with 3,5-dinitrosalicylic acid (DNS) that turns colour from pale yellow to brown. This colorimetric change allows us to determine the quantity of reducing sugars cleaved during the hydrolysis reaction. In this assay we prepared the reactants and substrate following commercial instructions, but modify the assay performance in order to characterize different features.

1% arabinoxylose solution was made adding 1 g of commercial wheat arabinoxylan to 90 mL of 100 mM sodium acetate buffer (pH 4.5) and dissolved by stirring at approximately 50ºC for 10 minutes on a magnetic stirrer hot plate. Volume was adjusted to 100 mL using 100 mM sodium acetate buffer (pH 4.5). Solution was stored in a well-sealed Duran bottle at room temperature.

Different characterization assays were performed following the same steps and concentrations, unless the assay had specific requirements. In order to have a significant colorimetric signal the enzyme substrate ratio used in the reactions was 5 mg of xylanase for 1 g of substrate.

The hydrolysis were carried out using 450 µL of 1% arabinoxylan, which made 0.0045 g of substrate, and 50 µL of enzyme solution, with a total xylanase concentration of 0.0225 mg, to maintain the enzyme substrate ratio of 5 mg enzyme per gram of substrate. The substrate dilutions together with the enzyme are incubated at the desired conditions for 15 minutes prior 750 µL of DNS were added to stop reaction. After DNS

reactant was added and the solutions were placed in a boiling water bath and boiled for 15 minutes. DNS reaction was stopped placing the boiled tubes in ice for at least 5 minutes and finally tubes colour intensity was measured at 540 nm using Nanodrop 200L spectrophotometer (Thermo Scientific). A reagent blank with the substrate and an enzyme blank were made for each condition and experiment.

Calibration curve was made using commercial xylose (Sigma Aldrich) standards at different concentrations in 500 µL total volume and following the same procedure as for the blank and samples. Using this method we have characterize different features of the Xylanase.

3.2.1.1.1 Temperature assay

Thermo stability of the xylanases with arabinoxylan was measured by first preincubating the enzyme to the desire temperature for 5 minutes and then followed the hydrolysis reaction and reducing sugars measurement as explained in section 3.2.1.1.

3.2.1.1.2 pH stability assay

To characterize the pH stability of the xylanases we prepared arabinoxylan substrate 1% with different pH buffers ranging from pH 4 to pH 10. The hydrolysis reactions were carried out as explained in section 3.2.1.1. A substrate blank at each pH was made in order to avoid substrate degradation interference because of different pH buffers.

3.2.1.1.3 T_{60} Assay

Enzymes were preincubated at 60 ºC 0,5,10,15,20 and 25 minutes. After preincubation time they were place in ice for 10 minutes and 10 minutes

at room temperature. Preincubated xylanases were characterized by hydrolysis reaction as explained in section 3.2.1.1

3.2.1.1.4 T$_{50}$ Assay

Enzymes were preincubated at different temperatures from 40ºC until activity decreased. After preincubation time they were place in ice for 10 minutes and 10 minutes at room temperature. Preincubated xylanases were characterized by hydrolysis reaction as explained in section 3.2.1.1.

3.2.1.1.5 Kinetics

Different arabinoxylose concentration samples were prepared from 0.4 mM xylose to 40 mM xylose. 450 µL of each solution were used to measure the kinetic parameters. Reactions were performed as explained in section 3.2.1.1.

3.2.1.2 CMC assays

This assay is specific for measuring endoglucanases activity we wanted to see weather our enzyme could cleave randomly beta 1-4 glucosidics bonds in substrates that are not their own specific ones. Water-soluble derivatives of cellulose such CMC and hydroxyethylcellulose (HEC) are commonly used to characterize beta 1,4-glucosidics bonds cleavage [152], [153]. The reaction of hydrolysis can be determined in different ways determining the reducing sugars: using colorimetric measurements or measuring the viscosity. We determined the colorimetric change in order to characterized reducing sugar changes. This method requires 0.5 mg of absolute glucose released under the reaction condition [154]. The reducing sugars concentrations are finally measured by the DNS method.

An enzyme and reactive with substrate blanks are made together with the samples.

3.2.1.1.5 Kinetics

We prepared different substrate concentrations by diluting CMC in 50 mM and pH 5 citrate buffer. Ranging concentrations from 0.005 mg/mL to 40 mg/mL, 250 µL of CMC at different concentration were hydrolysed using 50 µL of enzyme dilution containing a final concentration of 5 mg. The reaction was placed for 90 minutes and stopped by adding 750 µL of DNS reagent. The mixture of the samples together with the DNS was placed in a boiling water bath and boiled for 15 minutes. DNS reaction was stopped placing the boiled tubes in ice for at least 5 minutes and finally tubes colour intensity was measured at 540 nm using Nanodrop 200L spectrophotometer (Thermo Scientific). A reagent blank with the substrate at all the used concentrations and an enzyme blank were made. A calibration curve was made by using different D-glucose concentrations and following the same conditions as the samples.

3.2.2 LPMO characterization

C-LPMO and its Query Sv LPMO and Ch-LPMO) and its query Bt LPMO were characterized using a colorimetric assay to rapidly determine the activity.This assay is based on the recently discovered peroxidase activity of LPMO using 2,6 dimethoxyphenol (2,6-DMP) as chromogenic substrate. The enzymatic assay consists of 1 mM 2,6-DMP as chromogenic substrate, 100 µM H2O2 as co-substrate and an adequate
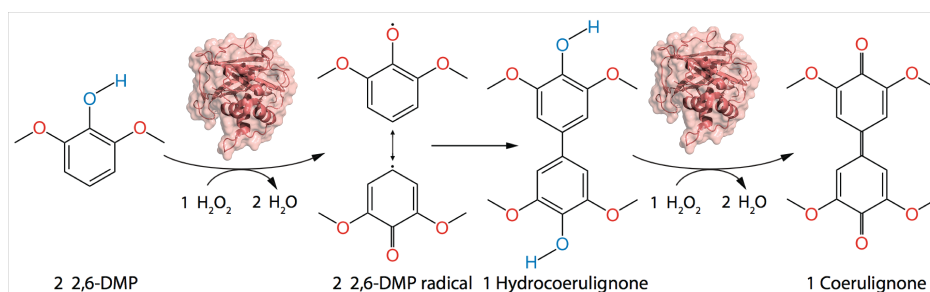
Figure 3.5. 2,6-DMP oxidation: LPMO catalyses the oxidation of 2,6-DMP to the corresponding phenoxy radical at the presence of H_2O_2 as co-substrate. 2,6 that generates the 2,6-DMP radical reduces the active site Cu (II). Two formed 2,6-DMP radicals dimerize rapidly to form the molecule call hydrocoerulignone, which by action of the LPMO is converted to coerulignone.

concentration of LPMO in a suitable buffer. The high molar absorption coefficient of the formed product coerulignone ($\varepsilon 469 = 53{,}200$ M-1 cm-1) makes the assay sensitive allowing performing reliable activity measurements of LPMO [155].

The reaction of the LPMO with 2,6-DMP has to be in presence of H_2O_2 as cosubstrate, which outperforms oxygen as cosubstrate. This reaction depends on the initial reduction of the active site copper from its Cu (II) state to Cu (I). 2,6-DMP and H_2O_2 are converted to 2,6-phenoxy radicals and water. Next, two 2,6-DMP phenoxy radicals dimerize and form hydrocoerulinone, which is again oxidized by LPMO forming the final Chromogenic product coerulignone (Figure 3.6).

Figure 3.6. Spectrophotometer EPCCH 2 plate reader from BioTek.

Unless otherwise specified, all the reactions were performed in 96- well plates mixing 166 µL of 100 mM phosphate buffer pH 8, 20 µL of 10 mM DMP (final concentration of 1mM) and 4 µL of 5mM H2O2 (final concentration of 0,1 mM). Finally we add 10 µL of enzyme dilution at a suitable concentration to each well, in a final volume of 200 µL per well. Plates were orbital mixed and reaction was follow at 50ºC. Activity was calculated measuring the increase of absorbance at 469 nm for 5 minutes using the molar absorption coefficient of coerulignone (ε_{469}=53,200 M^{-1} cm^{-1}) to calculate the peroxidase activity of LPMO.

3.2.2.1 DMP Kinetic assay

The Michaelis-Menten constant (K_m) was calculated changing DMP concentration in each well from 0.05 to 4 mM. The catalytic constant (K_{cat}) was determined from V_{max} and enzyme concentration. Reaction was carried out as explained in section 3.2.2. The enzymatic

concentration for all the substrate concentration was 0.15 mg per each well (Figure 3.7).

Figure 3.7. Example of a kinetic assay of DMP. The less substrate concentration the lighter orange we see. The experiment was carried out from 0.05 mM to 4 mM DMP.

3.2.2.2 DMP temperature assay

To study the resistance of the LPMO to the temperature we designed a new experiment based on the DMP assay (section 3.2.2). As DMP is a substrate sensible to high temperatures we could not carry out the reaction at high temperatures. We preincubated the LPMO enzymes to the targeted temperatures for 5 minutes in a 1.5 mL eppendorf in a thermo mixer and then place them in ice for 5 minutes. The reaction was carried out at a 96 well plate (thermo fisher). 169,5 µL 100 mM phosphate buffer pH 8, 20 µL 10 mM DMP as a chromogenic substrate that makes a final concentration of 1mM, 0.5 µL of 5 mM H_2O_2 and 10 µL of enzyme dissolution with adequate enzyme concentration making a final volume of 200 µL. Absorbance at 469 nm was measured during 30 minutes at Epoch 2 spectrophotometer (Figure 3.7).

3.2.2.3 DMP pH assay

pH profile of the LPMO´s was performed using DMP assay, as DMP uses H_2O_2 as a cofactor. pH profile for the activity of the all LPMO´s was made at a 96 well plate (thermo fisher) following the steps in 3.2.2.2. The pH of the reaction was the pH of the buffer (169,5 µL out of 200 µL). Different pH buffer dissolutions were prepared from pH 3 to pH 10: 100 mM citric acid for pH 3-5, 100 mM phosphate buffer for pH 6-7, 100 mM carbonate buffer for pH 8 and 100 mM Boric acid for 9-10.

3.2.3 Reconstructed enzymes synergy

To test the synergy between our ancestral enzymes comparing with the nowadays enzymes we designed an assay in were we studied the activity of the enzymes in lignocellulosic complex substrates. For that, we used 0,4 grams of pulps kindly provided from ENCE: One Unbleached Kraft pulp (UBKP) and other Bleached Kraft pulp (BKP) (Figure 3.8). Before the enzymatic hydrolysis we milled the substrates with 1mm grinder to facility the enzyme activity.

These two substrates are what we called non-ideal substrates. We used them as reference of lignocellulosic materials that we can found in lignocellulosic waste. UBKP contains lignin, hemicellulose and cellulose, and BKP contains hemicellulose and cellulose.
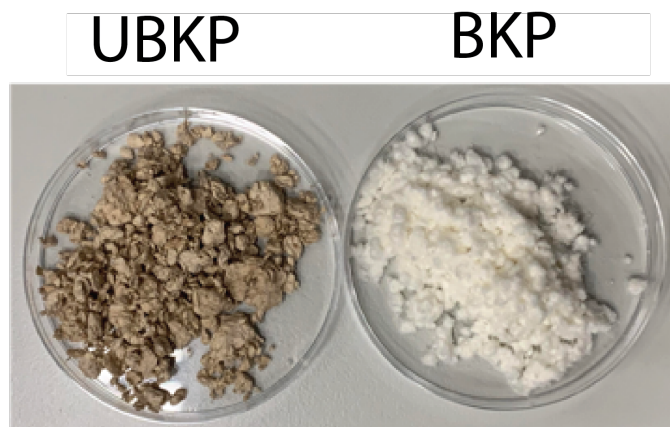
Figure 3.8: Non-ideal lignocellulosic substrates used for synergy test in our laboratory. On the left side, Unbleached Kraft Pulp (UBKP) a paper pulp that contains the main three components of lignocellulosic biomass (lignin, hemicellulose and cellulose). On the right side, Bleached Kraft Pulp (BKP) a pulp after a bleaching process to remove lignin, it is mainly composed of hemicellulose and cellulose.

For the comparison of different enzyme contribution in this hydrolysis we used different enzymatic ratios always maintaining the ratio of 5 total enzyme mg per gram of substrate. The ratios we used were: EG+CBM alone, EG+CBM + Xln (ratio 4:1), EG+CBM + LPMO (ratio 3:1), EG+CBM + Xln+ LPMO (ratio 3:1:1). As a control we used the queries in the same raitios. When C-LPMO was in the cocktail we added 2 mM of ascorbic acid (Sigma Aldrich) as electron donor. The hydrolysis was then incubated at 50ºC in agitation and water for 24 hours. We stopped the reaction by placing the Erlenmeyer in ice. Then we took a sample to calculate reducing sugars yield. EG reacts with cellulose, Xln with hemicellulose and C-LPMO with both.

Reducing sugars yield was measured by 3,5- dinitrosalicilic acid (DNS) [B 163,164], similar to section 3.2.1.1 but with some changes. Again, we used this method to quantify the reducing sugars produced during the enzymatic hydrolysis with a colorimetric measurement. We took 1 mL from each reaction and we added 3 mL of DNS reagent. Then, we boiled the solution for 5 minutes in a boiling water bath and cooled the boiled samples in ice for 10 minutes. Finally, we added 20 mL of miliQ water and measured the absorbance at 540 nm in a Nanodrop 200L spectrophotometer (Thermo Scientific). We used a glucose standard curve to calculate the reducing sugars concentrations in the different hydrolysis samples s (mg/mL). We prepared standards from 0.05 mg/mL glucose to 2 mg/mL and follow the same procedure that with the samples.

We prepared a substrate blank and all the enzyme ratios blanks, with ascorbic acid when needed, at each experiment. The blank were treated same to the samples and their absorbance was rested to the samples. Once we calculated the reducing sugar concentration we used the equation [1] or [2] to calculate the total reducing sugar conversion. The formulas are the same, but the concentration of the products sensible to the hydrolysed was different depending on the substrate and the enzymes we used.

$$[1] \ (\%) = \frac{Reducing \ sugars \ (mg/mL) \, x \, 0,9}{Initial \ cellulose \ concentration} \ x \ 100$$

$$[2] \ (\%) = \frac{Reducing \ sugars \ (mg/mL) \, x \, 0,9}{Initial \ cellulose \ and \ hemicellulose \ concentration} \ x \ 100$$

Where the factor 0,9 was to transform polysaccharides to monosaccharide's considering the water uptake during the hydrolysis.

3.2.4 CNC isolation and Characterization from BKP and UBKP

0,4 grams of BKP and UBKP substrates were milled with a 1 mm grinder previous hydrolysis to facility the enzyme activity. For this hydrolysis we used enzymatic treatments with 5 mg of total protein per gram of substrate: we used ANC EG+CBM (Anc EG to abbreviate) alone and ANC EG+ C-LPMO (ratio 4:1). As a control for the ancestral enzymes we used the queries enzymes from the reconstruction trees, the EG (Q EG) and *Sv* LPMO from *S. viridiosporus*. we added 2 mM ascorbic acid (Sigma Aldrich) as electron donor fro LPMO activity. The hydrolysis was incubated at 50 ºC in agitation and water for 24 hours. We stopped the reaction by placing the hydrolysis on ice and sonication the mixtures with a micro tip sonicator UPH100H Ultrasonic Processor (Hielscher) for 25 min at 75%. The nanocellulose was isolated by a gradual centrifugation steps and concentrated by ultracentrifugation at 33000 G for 1 hour. The pellets were resuspended in water and could be lyophilized in a Telstar Lyoquest for physicochemical characterization by freeze-drying for 24 hours. This process was optimized in a previous work in our laboratory [89].

3.2.4.1 Atomic force microscopy of CNC

The morphology of the nanoparticles obtained with enzymatic hydrolysis was studied with atomic forces microscopy (AFM) (Figure 3.9). This is a type of scanning probe microscopy (SPM), with demonstrated resolution

on the order of fractions and nanometre. It has more than 1000 times better diffraction limit than optical microscopes. A tip that is attached to a cantilever scans our sample. Information is gathered by the interaction between the tip and the sample, these interactions are attractive repulsion forces between them. Each time that there is an attractive or repulsive force between the tip and the sample the cantilever deflects, these phenomena allows us to map the deflections of the cantilever in each point of the sample. Images were captured at room temperature, in taping mode, using Nanoscope V scanning probe microscope (Multimode 8 Bruker Digital instruments) with an integrated force generated by cantilever/silicon probes. The applied resonance frequency was 320 kHz. The cantilever had a tip radius of 5-10 m and was 125 µm long. Samples were prepared by spin –coating (Spincoater P6700) at 2000 rpm for 130 seconds by casting a droplet of nanocellulose suspension on mica substrate. AFM height and phase mages were collected at the same time at all the samples.

Image frame were of 3x5 µm for all samples, in there fibbers from different sizes could be distinguished. The length of 100 nanoparticles was measured to calculate the average length. The diameter and the aspect ratio (Length/Diameter) .
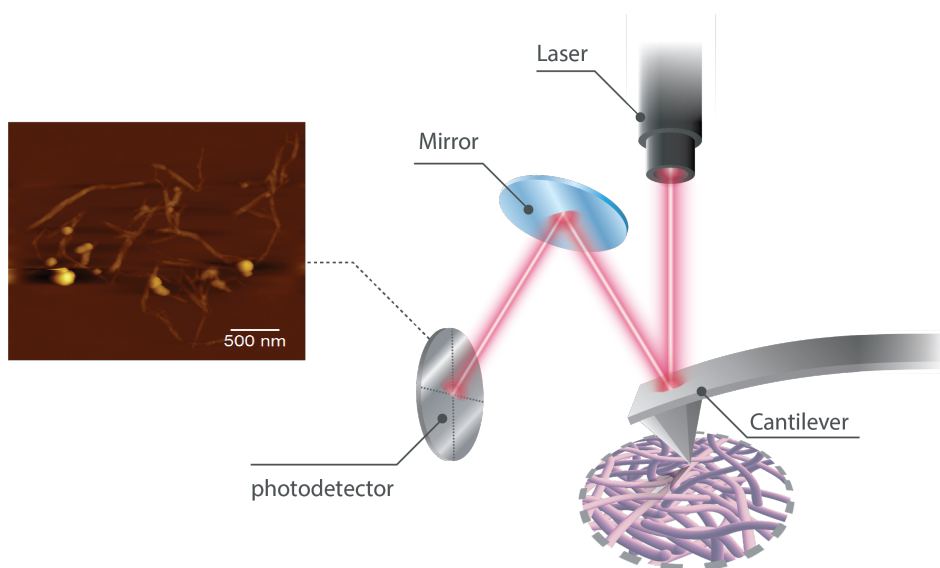
Figure 3.9. Atomic force microscopy (AFM). This technique uses the interaction between the tip and the sample attractive-repulsion forces to create a deflection in the tip and permits to infer the images by mapping the deflections in each point of the sample.

3.2.4.2 Fourier-transform infrared spectroscopy of CNC

This technique was used to analyse the characteristic functional groups of cellulose and nanocellulose. This technique allows us to measure the changes that happen in a reflected IR beam when it contacts the sample. The IR we used in our measurements is a Nicolet Nexus spectrometer provided with a MKII Golden Gate accessory (Specac) with a diamond crystal at a nominal incidence angle of 45º and ZnSe lens. Spectra were measured in attenuated reflection (ATR) mode between 4000 and 650 cm^{-1}, with averaging 32 scans and a resolution of 4 $cm^{-1}\,s^{-1}$.

3.3 New Nano biomaterial isolation

In this section I will explain both how we isolated nanochitin (CNCh) from commercial α-chitin (Sigma Aldrich) and how we characterized it. For the first time, we show how it is possible to isolate CNCh from recalcitrant commercial alpha chitin by using only enzymes. More accurately, Ch-LPMO .

3.3.1 Enzymatic nanochitin isolation

We used 1.6 grams of commercial chitin (Sigma Aldrich) for the isolation of nanochitin. As they have flake shapes there was no need to chopped or pre-treated them. We made the reaction in 200 mL of water by using 10 mg of C-LPMO and Sv LPMO per gram of substrate. The hydrolysis reaction was incubated at 50ºC in agitation for different times (12, 24, 48 and 72 hours). We stopped the reaction placing the reaction on ice. Later, and in order to avoid nanochitin aggregation, the mixtures were sonicated with a micro tip sonicator UPH100H Ultrasonic Processor (Hielscher) for 25 minutes at 75%. The nanochitin was isolated by gradual centrifugation steps and concentrated by ultracentrifugation at 33000 G for 1 hour. The pellets were resuspended in water and could be lyophilized in a Telstar Lyoquest for physicochemical characterization by freeze-drying for 24 hours.

3.3.2 Nanochitin yields

The total nanochitin conversion was calculated by weighting the freeze-dried nanochitin from each hydrolysis and we calculated the yield by terms of equation 3.

$$[3] \text{ Nanochitin yield (\%)} = \frac{nanochitin\ mass}{initial\ chitin\ mass}\ x\ 100$$

3.3 New Nano biomaterial characterization: CNCh

Nanochitin isolated by enzymatic treatment (En CNCh) with both ancestral and nowadays enzymes and nanochitin produced by acid hydrolysis from University of the Basque country (Ac CNCh) were analysed with different physicochemical techniques.

3.3.2 Atomic force microscopy

To study the morphology of the nanoparticles obtained with enzymatic hydrolysis and acid processes we used atomic force microscopy (AFM). We follow the same protocol that in section 3.2.4.1.

3.3.2 Fourier-transform infrared spectroscopy

This technique was used to analyse the characteristic functional groups of Chitin and nanochitin. We follow the same protocol that in section 3.2.4.2. In Figure 3.10 we show the characteristic peaks of α-Chitin.
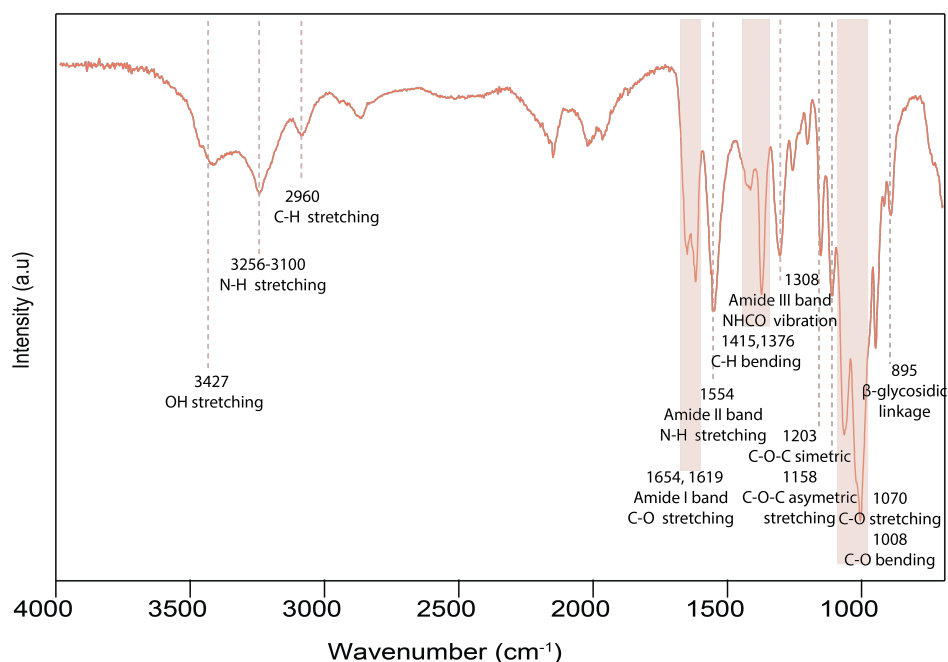
Figure 3.10. FTIR spectrum from α-chitin. Typical spectrum from a chitin sample with the characteristics chitin functional groups measured in a range from 4000 to 650 cm^{-1}.

3.3.3 X-Ray diffraction

In order to analyse the crystalline structure and crystalinity of CNCh, we use X-ray diffraction (XRD). This technique is based on the constructive interference of monochromatic X-rays and crystalline structures on the samples. X-ray powder diffraction patterns were collected using a Philips X´pert PRO automatic diffractometer operating at 40 kV and 40 mA, in theta-theta configuration, a secondary monochromator with CuK α radiation (λ=1.5418 Å) and PIXcel solid-state detector (active length in 2θ 3.347º). Data were collected from 5 to 80º 2θ(step size 0.026 and time per step 80 seconds) at room temperature. A fixed divergence and

antiscaterring slit giving constant volume of sample illumination were used.
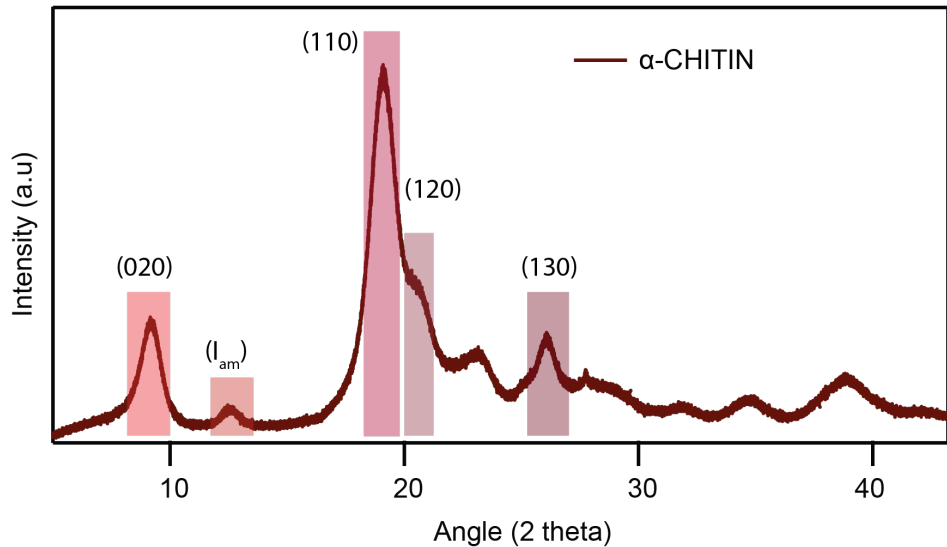
Figure 3.11. XRD spectrum from α-chitin. Typical spectrum from chitin crystalline structure peaks. We note the crystalline planes that they represent.

The data collected was both analysed as raw data and a diffractogram. Crystallinity index (CI%) was calculated using Segal equation (Equation 4):

[4] Crystallinity index (%)$=\frac{I200-Iam}{I200} x\ 100$

Where I_{200} is the intensity of the cellulose crystalline peak an I_{am} is the intensity of the amorphous peak. In Figure 3.11 we show an X-ray spectrogram of α-Chitin.

3.3.4 Solid-state cross-polarization magic angle spinning 13C nuclear magnetic resonance

The chitin functional groups were verified with solid-state cross polarization magic angle spinning ^{13}C nuclear magnetic resonance CP/MAS ^{13}C isotropic chemical shift, the contributions of aromatic and aliphatic groups. These groups contribute in different ranges: Aliphatic groups are placed in the range from 0 to 90 ppm and aromatic groups from 110 to 160 ppm. In the range of 160 to 185 we can found peaks corresponding to carboxylic acid derivatives. ^{13}C CP/MAS NMR spectra were recorded on a 400 MHz BRUKER system equipped with a 4mm MASDVT TRIPLE resonance HYX MAS probe. Larmor frequencies were between 400,17 mHz and 100,63 MHz for ^1H and ^{13}C nuclei, respectively. Chemical shifts were reported relative to the signals of ^{13}C nuclei in glycine. Relaxation delay of the sample was 5 seconds and rotation frequency was 12 KHz with 2K scans. Polarization transfer was achieves with RAMP cross-polarization (ramp on proton channel) with 5 ms contact time. High-power SPINAL 64 heteronuclear proton decoupling was applied during acquisition.

From the chitin NMR spectrum peaks (Figure 3.12) corresponding to each carbon from chitin polymer are observed. Peaks from 0 to 40 ppm correspond to alkyl groups, in this case to the CH_3. The peak around 55 ppm correspond to C_2, peaks between 60 and 70 correspond to C_6 of the primary alcohol group. The peak between 70 and 80 ppm are assigned to carbons C_3, $C_{5\ and}$ C_4. Between 100 and 110 ppm we find the anomeric C_1 carbon peak, and finally around 170 and 180 ppm we have the carboxyl group C
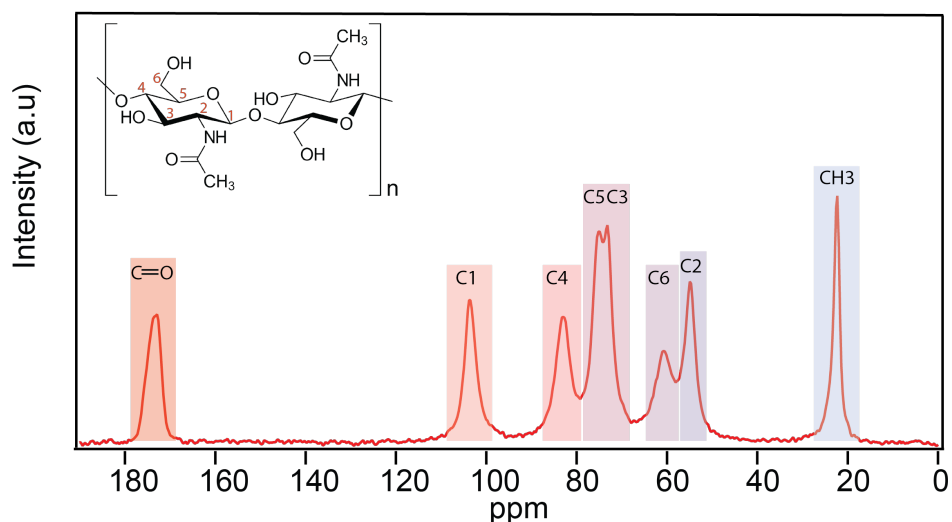
Figure 3.12.CP/MAS ^{13}C NMR spectra of α-chitin. At the NMR spectra different carbons from the chitin polymer can be observe. Peaks in the range of 0-40ppm correspond to primary, secondary or tertiary alkyl. Peaks in the range of 60 and 70 ppm usually correspond to C6 of the primary alcohol group. Peaks in the range of 70 and 80 ppm are assigned to C_2, C_3 and C_5. Peaks from 80 to 95 ppm are the ones corresponding to C4 and between 100 and 110 ppm to the C1 anomeric carbon. Finally the peak between 175 and 180 ppm correspond to the carboxyl C in the acetyl group

.

3.3.5 Thermogravimetric analysis

To determine the thermal stability of chitin samples we used thermogravimetric analysis (TGA). This technique measures the weight of the samples in a microbalance during a heating scan. This way, we follow the degradation process of the material. The analysis was carried

put using TGA/ADTA 85 Mettler Toledo equipment. 10 mg of all the samples were used and heated from 30ºC to 800ºC in a nitrogen atmosphere at scanning rate of 10ºC/minute.
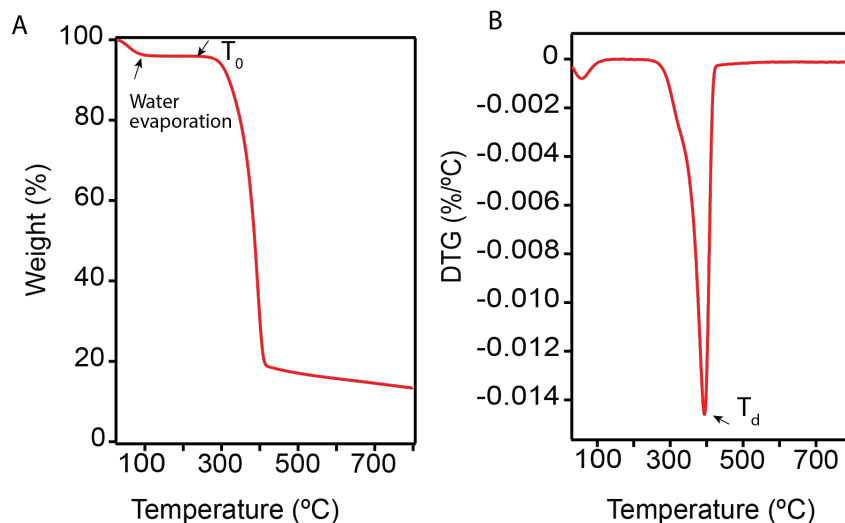
Figure 3.12: (A) Weigh loss with temperature (B) DTG curves from commercial alpha chitin. The early weight loss, indicated in figure (A) is due to water evaporation absorbed by the chitin. The T_0 is the loss of the 5 % of the weight and the Td is the minimum of the peak in the derivative curve.

The initial degradation temperature (T_0) is described as the loss of 5% of the weight of the total sample and the maximum degradation temperature (Td) is the minimum of the degradation peak in the derivative of thermogravimetric curves (DTG).

PART III

Capítulos 4, 5, 6, 7 y 8 sujetos a confidencialidad por la autora

PART IV

Chapter 9:Discussion

The aim of this thesis was to reconstruct ancestral enzymes with improved activity ranges, in terms of pH and temperature, for nanomaterial obtaining from biomass. First, we reconstructed a bacterial endo 1,4-β-xylanase and a bacterial LPMO to help an Anc EG, reconstructed in our laboratory; these enzymes are able to produces nanocellulose from lignocellulosic biomass. Moreover, to consolidate the capability of ancestral enzymes for the obtention of nanomaterials we reconstructed another bacterial LPMO to hydrolyze chitin and compare the activities of the reconstructed enzymes with their homologous extant ones.

Reconstructed enzymes (table 9.1) showed identities equal or higher than 60% with their Queries, this means that they have enough mutations to act differently but still having the same catalytic activity.

Table 9.1:Ancestral enzymes reconstructed in this work, their geological time, their Query, identity with the query and number of aminoacid mutated.

Enzyme	Age (My)	Query	Identity	Mutations
Anc Xln	2.97	*Bacillus subtillis*	88%	25
C-LPMO	3.077	*Bacillus thurigiensis*	60%	71
Ch-LPMO	3.077	*Streptomyces viridisporus*	63%	69

The catalytic pocket is conserved in all reconstructed nodes, which means that the changes in activity we observed could be due to enzyme-substrate interaction. However, an ancestral node of Ch-LPMO tree showed an interesting mutation in its catalytic pocket were, instead of having a Valine (V), characteristic amino acid present in chitin hydrolyzing LPMOs, has a Leucine (L) (Figure 9.1 A), characteristic aminoacid present in cellulose hydrolyzing LPMOs [212], and it shares more common amino acids with cellulose hydrolyzing LPMO. The sequence belongs to Last Actinobacteria Common ancestor (LACA) node, the youngest bacterial filum in Ch-LPMO tree, which may indicate that LPMO evolved from degrading chitin to cellulose. There are some fossil evidences of chitin present in the first branched organisms 500 million years ago [213] and in multicellular organic-walled microfifrils around 900 million years ago [214], before cellulose appear in plant cell walls (cita), suggesting that our 900 mya reconstructed LACA Ch-LPMO could be on its transition to live with chitin structured organisms

to cellulose structures organisms, right before Cambrian explosion, when modern marine ecosystem started [212].

Figure 9.1. Ch-LPMO aligment A) Ancestral Ch-LPMO sequences from different nodes and Bt LPMO sequence aliniation, frmae node 54 (LACA) mutations B) Aliniation of reconstructed ancestral C-LPMO and St LPMO and reconstructed ancestral Ch-LPMO and Bt LPMO.

Reconstructed enzymes were from the Hadean and the Archeon eons where ocean temperature has been estimated to be around 60-70ºC [71], [73], [215], [216], so bacteria living in that environment were extremophiles. Ancestral enzymes displayed this thermophile phenotype by showing increased activity in a range of temperatures and better thermal stability. Moreover, ancestral enzymes enhanced other properties such as pH, chemical promiscuity or higher expression rates. Several strategies have been proposed as protein engineering to improve enzymes activity for industrial applications, but none of them is capable of improving activity in different aspects at the same time. This proves

that ASR is an effective method to improve enzymes for industrial applications.

Reconstructed enzymes in this work improved several features regarding their Queries. Thermal stability was better in the three reconstructed enzymes: Ancestral xylanase showed more than 50% activity between 30-62ºC whereas query xylanase from 40-60ºC; Anc C-LPMO and Q C-LPMO had their optimal temperature at 70ºC, but Anc C-LPMO showed 2.5 times higher activity; same for Anc Ch-LPMO that showed almost 2 times better activity than Q Ch-LPMO at their optimal temperature, 70ºC. Reconstructed enzymes had same pH profiles than their extant homologous but showing higher activities in all pH-s for Anc Xln, and in basic pHs for the LPMOs. With REAL substrates, not laboratory substrates, they showed better activities and synergy than the queries. From these real subtracted we can obtain valuable nanomaterials, and here we proved that our reconstructed enzymes were more efficient in the obtention of nanomaterials from biomass than their queries.

Material science is predominantly an engineering discipline; where building blocks are intentionally assemble to achieve materials with desire properties, on the contrary, enzymes have been studied by biochemistry discipline, that provide exceptional features such as substrate specificity, rate acceleration, region-, chemo, and stereo selectivity within catalyzed reactions, so enzymes in industry are used to obtain chemically valuable products. **Here we combined the unique properties of both, materials and enzymes in a synergistic way leading to advance materials production and modification,** this integration of material and enzyme science is used in many fields such as

grafting coating of material, biosynthesis of materials, biodegradation of materials, immobilization of enzymes, biosensor... [217] (Figure 9.2) .

Ancestral enzymes effectively obtained nanomaterials from biomass. CNC and CNCh obtention from biomass is a promising application to these enzymes. These nanomaterials are natural, biocompatible, and biodegradable and can be obtained from waste. Besides their natural characteristics, their nano size give raise to great potential applications in biomedicine.

Our group successfully used an ancestral EG to obtain nanocellulose from cellulose and in this work, we have confirmed that the use of ancestral enzymes in material science develops unique and functional materials. Ancestral enzymes produced higher yields of CNC and CNCh than their modern homologues, but the most novel contribution of this work is not only that we obtain nanocellulose and nanochitin enzymatically from non ideal substrates, but **we are able, for the first time, to enzymatically oxidized these CNC and CNCh with a single enzymatic treatment: LPMO**. Reconstructed LPMOs in this work have, C-LPMO and Ch-LPMO, hydrolyze in an oxidative cleavage cellulose and chitin respectively.

Oxidation of these two nanoparticles surface is very important in material engineering to selectively modify these nanoparticles surfaces. Carboxyl groups are very reactive, and a functional group can be attached or substitute to the carboxyl group placing the desire functional group in these surfaces

Biosynthesis of materials

Biodegradation of materials

Grafting/coating of materials

Biocompatibility of materials

Inmobilization of enzymes

Bio-responsive materials

Biosensors
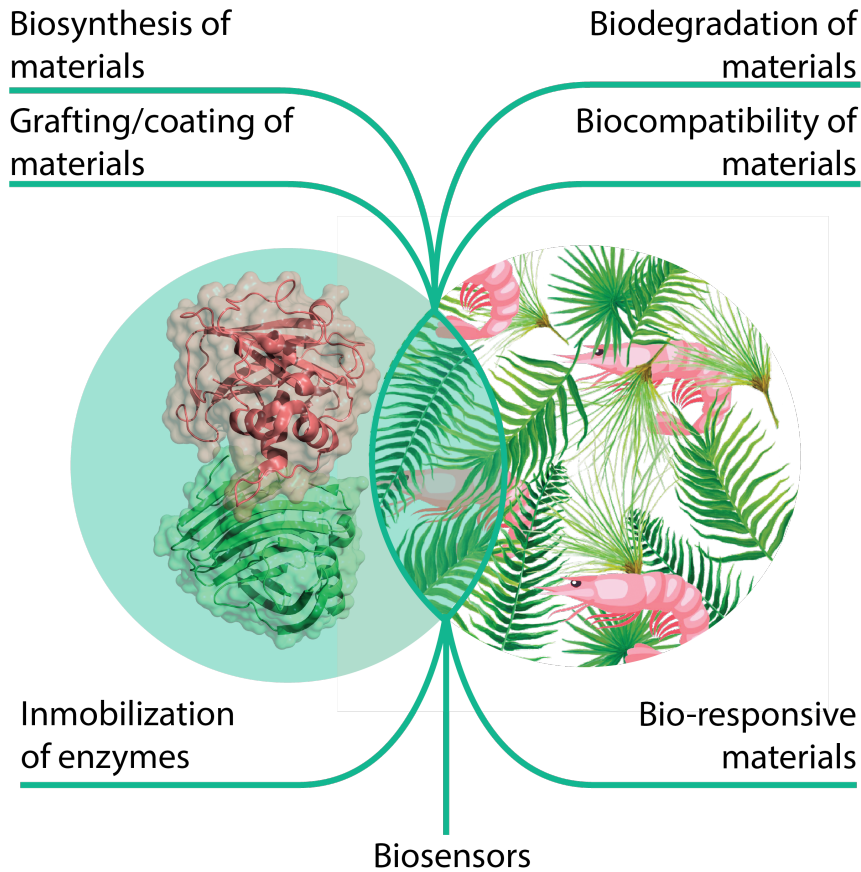
Figure 9.2: Fields of enzymes application in material science. In the Figure as an example the two most abundantly row materials on earth: lignocellulosic biomass and chitin.

.The enzymatic functionalization of these nanomaterials opens a new range of applications, as the biocompatibility and non-toxicity of the material is directly linked to its production protocol. For example, in pharmacology or biomedicine, where target drugs can be placed in these ends; or wrap treatments in a biocompatible CNC or CNCh films. We demonstrate that LPMO hydrolysis oxidizes nanocellulose and nanochitin as TEMPO, a chemical used in industry to oxidized materials,

do it [106]. Although cellulose C_6 carbon is oxidized like TEMPO, chitin, that has also an acetyl group in C_2, showed two oxidized carbons. This has not been reported with TEMPO or other chemical mediation. Regarding RMN spectra (Figure 9.3), we think that LPMO acts oxidizing both C_6 and C_2 from nanochitin, demonstrating ancestral LPMO promiscuity towards different carbon oxidation from a polysaccharide chain, in comparison with the extant homologous.
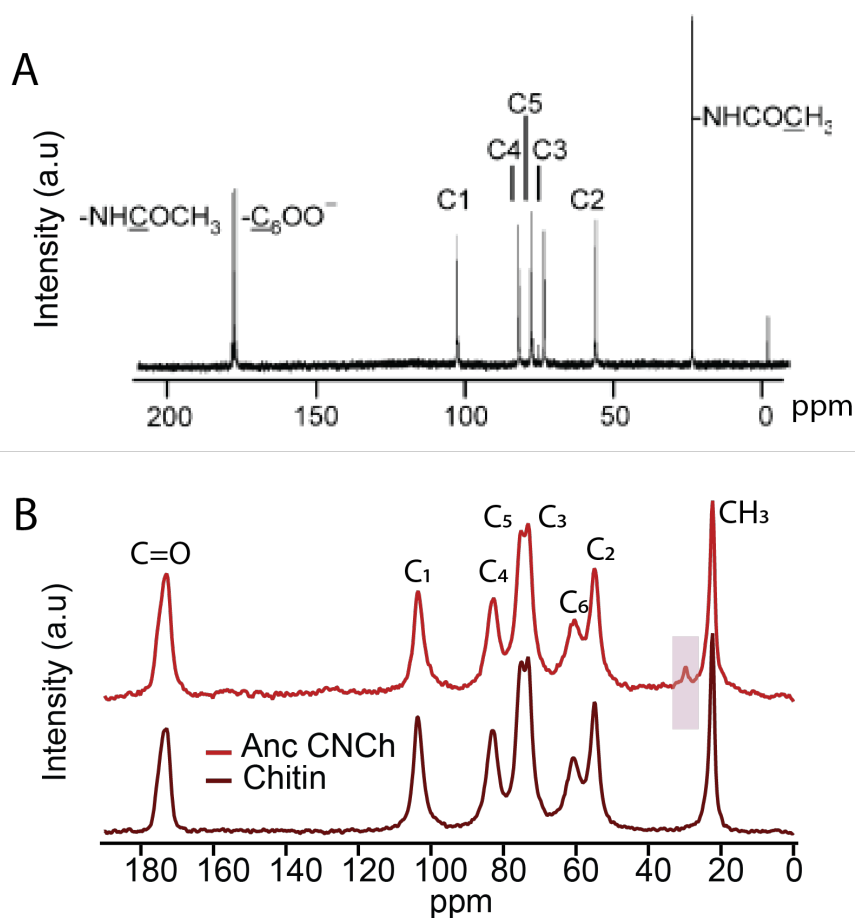
Figure 9.3 CP/MAS ^{13}C NMR A) TEMPO oxidized CNCh from literature [206] and, B) α-chitin (marron) and Ch-LPMO oxidized CNCh (red).

Our enzymatic treatment of the materials is made with water, avoiding further dialysis processes to purify the nanomaterial and avoiding salts from the buffer to be attached to the nanomaterial surface. In addition, enzymes, as biomolecules, are biocompatible so the biocompatibility of the enzymatically engineered nanomaterials is preserved after treatment.

Reconstructed enzymes with not an oxidative cleavage are able of isolating nanomaterials from biomass preserving the native structure of the material. This was showed in a previous work in our group were Anc EG+CBM produced CNC with native cellulose polymorph from pure cellulose filter paper. Here we reconstructed an Endo 1,4-beta xylanase to degrade hemicellulose and help Anc EG+CBM isolate CNC. Ancestral enzymes showed better synergic work than modern enzymes, by obtaining higher reducing sugars and CNC conversion yields. Obtained CNC with EG+CBM and Anc Xln preserved cellulose native structure, the addition of the C-LPMO to the enzymatic cocktail boost the enzymatic reducing sugars and CNC conversion yields and produced an oxidized CNCs. It is therefore reasonable to think that native structure of chitin can be preserved using a chitinase to produce CNCh. In this work, we tried with a commercial chitinase that did not react with the recalcitrant chitin; in consequence it would be interesting to reconstruct an ancestral chitinase as we have seen that ancestral enzymes are more active with recalcitrant substrates. There are also enzymes for the deacetylation of chitin called chitin deacetylase, these combined with chitinases could produces for instance nanochitosan crystals that due to the lack of acetyl group would be CNC like but hydrophobic.

Table 9.2: This table show the enzyme combination have been used in our laboratory, with which substrate and after the hydrolysis the structure of the nanomaterial.

ENZYME	SUBSTRATE	CNC/CNCh STRUCTURE
EG	Cellulose	Native
EG+Xln	Cellulose	Native
EG+Xln+C-LPMO	Cellulose	Oxidized
Ch-LPMO	Chitin	Oxidized

In this work we show that different combination of enzymes produces nanomaterials with different structures (Table 9.2) so **we can have CNC and CNCh "a la carte" with the only use of enzymes.** This opens a new world to engineer nanomaterials with different features and biocompatible. According to our data, reconstructed enzymes are fundamental in these processes as they nanomaterial productions yields (table 9.3).

Table 9.3: conversion yield in % of CNC from two UBKP and BKP substrates for enzymatic cocktail (CK) of EG+Xln+LPMO. We compare the yield difference for ancestral and modern proteins.

SUBSTRATE	Conversion yield of enzymatic CK	
	Ancestral	Modern
UBKP	15%	11.5%
BKP	16%	12%

Nanochitin conversion is also more efficient with ancestral Ch-LPMO than with modern Bt LPMO obtaining 5 mg of CNCh and 2 mg of CNCh respectively. This means that ancestral LPMO is two folds efficient with recalcitrant chitin as substrate.

In this work we have successfully enhanced activities in terms of pH, T, expression levels and promiscuity of three enzymes using Ancestral Reconstruction method, a method that commonly is used in paleoenzymology studies. Moreover, these biomass degrading enzymes isolate CNC and CNCh from biomass more effectively than their extant homologous. On top of that, C-LPMO enzyme oxidized CNC and CNCh, opening a fully sustainable and biocompatible path to engineer nanobiomaterials for further biomedical applications.

Bibliography

[1]   K. Kostarelos, A. Bianco, and M. Prato, "Promises, facts and challenges for carbon nanotubes in imaging and therapeutics," *Nat. Nanotechnol.*, vol. 4, no. 10, pp. 627–633, 2009.

[2]   K. F. Mak, J. Shan, and D. C. Ralph, "Probing and controlling magnetic states in 2D layered magnetic materials," *Nat. Rev. Phys.*, 2019.

[3]   M. N. Mustafa, S. Shafie, M. H. Wahid, and Y. Sulaiman, "Light scattering effect of polyvinyl- alcohol / titanium dioxide nanofibers in the dye-sensitized solar cell," *Sci. Rep.*, pp. 1–8, 2019.

[4]   M. R. Begley, D. S. Gianola, and T. R. Ray, "Bridging functional nanocomposites to robust macroscale devices," *Science (80-.).*, vol. 364, no. 6447, 2019.

[5]   C. Hong, T. T. Tang, C. Y. Hung, R. P. Pan, and W. Fang, "Liquid crystal alignment in nanoporous anodic aluminum oxide layer for LCD panel applications," *Nanotechnology*, vol. 21, no. 28, 2010.

[6]   I. Dierking, "Nanoscience with liquid crystals," *Liq. Cryst. Today*, vol. 26, no. 3, pp. 63–65, 2017.

[7]   A. Kamyshny and S. Magdassi, "Conductive nanomaterials for printed electronics," *Small*, vol. 10, no. 17, pp. 3515–3535, 2014.

[8]   M. Joshi and U. Chatterjee, *Polymer nanocomposite*. Elsevier Ltd, 2016.

[9]   J. Kumirska *et al.*, "Application of spectroscopic methods for structural analysis of chitin and chitosan," *Mar. Drugs*, vol. 8, no.

5, pp. 1567–1636, 2010.

[10] C. Santhosh, V. Velmurugan, G. Jacob, S. K. Jeong, A. N. Grace, and A. Bhatnagar, "Role of nanomaterials in water treatment applications: A review," *Chem. Eng. J.*, vol. 306, pp. 1116–1137, 2016.

[11] K. Gajanan and S. N. Tijare, "Applications of nanomaterials," *Mater. Today Proc.*, vol. 5, no. 1, pp. 1093–1096, 2018.

[12] N. Lin and A. Dufresne, "Nanocellulose in biomedicine: Current status and future prospect," *Eur. Polym. J.*, vol. 59, no. July, pp. 302–325, 2014.

[13] A. Dufresne, "Nanocellulose: A new ageless bionanomaterial," *Mater. Today*, vol. 16, no. 6, pp. 220–227, 2013.

[14] N. Angulakshmi, S. Thomas, K. S. Nahm, M. M. Stephan, and N. N. Elizabeth, "Electrochemical and mechanical properties of nanochitin-incorporated PVDF-HFP-based polymer electrolytes for lithium batteries," *Ionics (Kiel).*, vol. 17, no. 5, pp. 407–414, 2011.

[15] J. G. Torres-Rendon *et al.*, "Bioactive gyroid scaffolds formed by sacrificial templating of nanocellulose and nanochitin hydrogels as instructive platforms for biomimetic tissue engineering," *Adv. Mater.*, vol. 27, no. 19, pp. 2989–2995, 2015.

[16] K. Kerman, M. Saito, E. Tamiya, S. Yamamura, and Y. Takamura, "Nanomaterial-based electrochemical biosensors for medical applications," *TrAC - Trends Anal. Chem.*, vol. 27, no. 7, pp. 585–592, 2008.

[17] M. Keerthi, G. Boopathy, S.-M. Chen, T.-W. Chen, and B.-S. Lou, "A core-shell molybdenum nanoparticles entrapped f-MWCNTs hybrid nanostructured material based non-enzymatic biosensor for electrochemical detection of dopamine neurotransmitter in biological samples," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.

[18]   Y. S. Zhang and A. Khademhosseini, "Advances in engineering hydrogels," *Science (80-.).*, vol. 356, no. 6337, 2017.

[19]   G. A. A. Saracino, D. Cigognini, D. Silva, A. Caprini, and F. Gelain, "Nanomaterials design and tests for neural tissue engineering," *Chem. Soc. Rev.*, vol. 42, no. 1, pp. 225–262, 2013.

[20]   G. Wunner *et al.*, "Certain Critical Fi Eld Strength, the 3 Σ," vol. 337, no. July, pp. 303–305, 2012.

[21]   R. Aversa, R. V. V. Petrescu, R. Sorrentino, F. I. T. Petrescu, and A. Apicella, "Hybrid ceramo-polymeric nanocomposite for biomimetic scaffolds design and preparation," *Am. J. Eng. Appl. Sci.*, vol. 9, no. 4, pp. 1096–1105, 2016.

[22]   R. A. Hule and D. J. Pochan, "Polymer nanocomposites for biomedical applications," *MRS Bull.*, vol. 32, no. 4, pp. 354–358, 2007.

[23]   B. J. Biol *et al.*, ", B. J. Biol. 14. BENDICH, A ., RUSSELL," no. 101, 1953.

[24]   R. A. Khan *et al.*, "Production and properties of nanocellulose-reinforced methylcellulose-based biodegradable films," *J. Agric. Food Chem.*, vol. 58, no. 13, pp. 7878–7885, 2010.

[25]   S. P. Mishra, A. S. Manent, B. Chabot, and C. Daneault, "Production of nanocellulose from native cellulose - Various options utilizing ultrasound," *BioResources*, vol. 7, no. 1, pp. 422–435, 2012.

[26]   Y. Zhou, T. Saito, L. Bergström, and A. Isogai, "Acid-Free Preparation of Cellulose Nanocrystals by TEMPO Oxidation and Subsequent Cavitation," *Biomacromolecules*, vol. 19, no. 2, pp. 633–639, 2018.

[27]   M. Martelli-Tosi, M. D. S. Torricillas, M. A. Martins, O. B. G. De Assis, and D. R. Tapia-Blácido, "Using Commercial Enzymes to Produce Cellulose Nanofibers from Soybean Straw," *J.*

Nanomater., vol. 2016, 2016.

[28]   J. Y. Zhu, R. Sabo, and X. Luo, "Integrated production of nano-fibrillated cellulose and cellulosic biofuel (ethanol) by enzymatic fractionation of wood fibers," *Green Chem.*, vol. 13, no. 5, pp. 1339–1344, 2011.

[29]   Z. Karim, S. Afrin, Q. Husain, and R. Danish, "Necessity of enzymatic hydrolysis for production and functionalization of nanocelluloses," *Crit. Rev. Biotechnol.*, vol. 37, no. 3, pp. 355–370, 2017.

[30]   P. K. Busk and L. Lange, "Classification of fungal and bacterial lytic polysaccharide monooxygenases," *BMC Genomics*, vol. 16, no. 1, pp. 1–13, 2015.

[31]   E. Chiu *et al.*, "Structural basis for the enhancement of virulence by viral spindles and their in vivo crystallization," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 13, pp. 3973–3978, 2015.

[32]   G. Vaaje-Kolstad *et al.*, "An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides," *Science (80-.).*, vol. 330, no. 6001, pp. 219–222, 2010.

[33]   R. J. Quinlan *et al.*, "Source : Proceedings of the National Academy of Sciences of the United States of America , Vol .," 2015.

[34]   R. J. Quinlan *et al.*, "Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 37, pp. 15079–15084, 2011.

[35]   L. Lo Leggio *et al.*, "Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase," *Nat. Commun.*, vol. 6, 2015.

[36]   C. M. Phillips, W. T. Beeson, J. H. Cate, and M. A. Marletta, "Cellobiose dehydrogenase and a copper-dependent

polysaccharide monooxygenase potentiate cellulose degradation by Neurospora crassa," *ACS Chem. Biol.*, vol. 6, no. 12, pp. 1399–1406, 2011.

[37]   W. T. Beeson, C. M. Phillips, J. H. D. Cate, and M. A. Marletta, "Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases," *J. Am. Chem. Soc.*, vol. 134, no. 2, pp. 890–892, 2012.

[38]   Y. Arfi, M. Shamshoum, I. Rogachev, Y. Peleg, and E. A. Bayer, "Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 25, pp. 9109–9114, 2014.

[39]   G. Müller, P. Chylenski, B. Bissaro, V. G. H. Eijsink, and S. J. Horn, "The impact of hydrogen peroxide supply on LPMO activity and overall saccharification efficiency of a commercial cellulase cocktail," *Biotechnol. Biofuels*, vol. 11, no. 1, pp. 1–17, 2018.

[40]   David Cannella, Chia-wen C Hsieh, Claus Felby, and Henning Jørgensen, "Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content," *Biotechnol. Biofuels*, vol. 5, no. 26, pp. 1–10, 2012.

[41]   S. P. S. Chundawat, G. T. Beckham, M. E. Himmel, and B. E. Dale, "Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals," *Annu. Rev. Chem. Biomol. Eng.*, vol. 2, no. 1, pp. 121–145, 2011.

[42]   N. C. Carpita and D. M. Gibeaut, "Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth," *Plant J.*, vol. 3, no. 1, pp. 1–30, 1993.

[43]   H. Ishizaki and K. Hasumi, *Ethanol Production from Biomass*. Elsevier, 2013.

[44]   N. Xu *et al.*, "Biomethane Production From Lignocellulose: Biomass Recalcitrance and Its Impacts on Anaerobic Digestion," *Front. Bioeng. Biotechnol.*, vol. 7, no. August, pp. 1–12, 2019.

[45]   L. Viikari, M. Alapuranen, T. Puranen, J. Vehmaanperä, and M. Siika-Aho, "Thermostable enzymes in lignocellulose hydrolysis," *Adv. Biochem. Eng. Biotechnol.*, vol. 108, no. June, pp. 121–145, 2007.

[46]   P. M. D. Jaramillo, H. A. R. Gomes, A. V Monclaro, C. O. G. Silva, and E. X. F. Filho, "Section 2 Production of recombinant peptides Chapter 6 Lignocellulose   degrading enzymes : An overview of the global market."

[47]   D. M. Reimer, "The fine line of patient consent.," *J. Am. Assoc. Nurse Anesth.*, vol. 43, no. 1, pp. 57–58, 1975.

[48]   Y. S. Cheng, Y. Zheng, C. W. Yu, T. M. Dooley, B. M. Jenkins, and J. S. Vandergheynst, "Evaluation of high solids alkaline pretreatment of rice straw," *Appl. Biochem. Biotechnol.*, vol. 162, no. 6, pp. 1768–1784, 2010.

[49]   M. W. Easson *et al.*, "The application of ultrasound in the enzymatic hydrolysis of switchgrass," *Appl. Biochem. Biotechnol.*, vol. 165, no. 5–6, pp. 1322–1331, 2011.

[50]   K. K. Wong, L. U. Tan, and J. N. Saddler, "Multiplicity of beta-1,4-xylanase in microorganisms: functions and applications.," *Microbiol. Rev.*, vol. 52, no. 3, pp. 305–317, 1988.

[51]   A. M. Madlala, S. Bissoon, S. Singh, and L. Christov, "Xylanase-induced reduction of chlorine dioxide consumption during elemental chlorine-free bleaching of different pulp types," *Biotechnol. Lett.*, vol. 23, no. 5, pp. 345–351, 2001.

[52]   B. Battan, J. Sharma, S. S. Dhiman, and R. C. Kuhad, "Enhanced production of cellulase-free thermostable xylanase by Bacillus pumilus ASH and its potential application in paper industry," *Enzyme Microb. Technol.*, vol. 41, no. 6–7, pp. 733–739, 2007.

[53] L. Laureano-Perez, F. Teymouri, H. Alizadeh, and B. E. Dale, "Understanding Factors that Limit Enzymatic Hydrolysis of Biomass," *Twenty-Sixth Symp. Biotechnol. Fuels Chem.*, vol. 121, pp. 1081–1099, 2009.

[54] N. Sathitsuksanoh *et al.*, "Lignin fate and characterization during ionic liquid biomass pretreatment for renewable chemicals and fuels production," *Green Chem.*, vol. 16, no. 3, pp. 1236–1247, 2014.

[55] K. Öhgren, R. Bura, J. Saddler, and G. Zacchi, "Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover," *Bioresour. Technol.*, vol. 98, no. 13, pp. 2503–2510, 2007.

[56] M. Heinzkill, L. Bech, T. Halkier, P. Schneider, and T. Anke, "Characterization of laccases and peroxidases from wood-rotting fungi (family Coprinaceae)," *Appl. Environ. Microbiol.*, vol. 64, no. 5, pp. 1601–1606, 1998.

[57] M. E. Arias, M. Arenas, J. Rodríguez, J. Soliveri, A. S. Ball, and M. Hernández, "Kraft pulp biobleaching and mediated oxidation of a nonphenolic substrate by laccase from Streptomyces cyaneus CECT 3335," *Appl. Environ. Microbiol.*, vol. 69, no. 4, pp. 1953–1958, 2003.

[58] X. F. Zhang, G. Y. Yang, Y. Zhang, Y. Xie, S. G. Withers, and Y. Feng, "A general and efficient strategy for generating the stable enzymes," *Sci. Rep.*, vol. 6, no. May, pp. 1–12, 2016.

[59] S. C. Dodani, G. Kiss, J. K. B. Cahn, Y. Su, V. S. Pande, and F. H. Arnold, "Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models," *Nat. Chem.*, vol. 8, no. 5, pp. 419–425, 2016.

[60] A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady, and D. N. Woolfson, "Installing hydrolytic activity into a completely de novo protein framework," *Nat. Chem.*, vol. 8, no. 9, pp. 837–844,

2016.

[61]   M. R. Smith, E. Khera, and F. Wen, "Engineering novel and improved biocatalysts by cell surface display," *Ind. Eng. Chem. Res.*, vol. 54, no. 16, pp. 4021–4032, 2015.

[62]   X. Duan, J. Chen, and J. Wu, "Improving the thermostability and catalytic efficiency of Bacillus deramificans pullulanase by site-directed mutagenesis," *Appl. Environ. Microbiol.*, vol. 79, no. 13, pp. 4072–4077, 2013.

[63]   E. G. Hibbert *et al.*, "Directed evolution of biocatalytic processes," *Biomol. Eng.*, vol. 22, no. 1–3, pp. 11–19, 2005.

[64]   P. A. Dalby, "Optimising enzyme function by directed evolution," *Curr. Opin. Struct. Biol.*, vol. 13, no. 4, pp. 500–505, 2003.

[65]   M. Anbar, O. Gul, R. Lamed, U. O. Sezerman, and E. A. Bayer, "Improved thermostability of Clostridium thermocellum endoglucanase Cel8A by using consensus-guided mutagenesis," *Appl. Environ. Microbiol.*, vol. 78, no. 9, pp. 3458–3464, 2012.

[66]   P. Molina-Espeja, J. Viña-Gonzalez, B. J. Gomez-Fernandez, J. Martin-Diaz, E. Garcia-Ruiz, and M. Alcalde, "Beyond the outer limits of nature by directed evolution," *Biotechnol. Adv.*, vol. 34, no. 5, pp. 754–767, 2016.

[67]   O. F. Brandenberg, K. Chen, and F. H. Arnold, "Directed Evolution of a Cytochrome P450 Carbene Transferase for Selective Functionalization of Cyclic Compounds," *J. Am. Chem. Soc.*, vol. 141, no. 22, pp. 8989–8995, 2019.

[68]   B. K. Tischer, G. Smith, and N. Osterrieder, "En Passant: A Two Step Markerless REd Recombination System," vol. 634, no. 7, pp. 421–430, 2010.

[69]   S. Lutz, "Beyond directed evolution-semi-rational protein engineering and design," *Curr. Opin. Biotechnol.*, vol. 21, no. 6, pp. 734–743, 2010.

[70] P. A. Dalby, "Strategy and success for the directed evolution of enzymes," *Curr. Opin. Struct. Biol.*, vol. 21, no. 4, pp. 473–480, 2011.

[71] E. A. Gaucher, S. Govindarajan, and O. K. Ganesh, "Palaeotemperature trend for Precambrian life inferred from resurrected proteins," *Nature*, vol. 451, no. 7179, pp. 704–707, 2008.

[72] M. J. Harms and J. W. Thornton, "Historical contingency and its biophysical basis in glucocorticoid receptor evolution," *Nature*, vol. 512, no. 7513, pp. 203–207, 2014.

[73] V. Nguyen *et al.*, "10.1126@Science.Aah3717," *Science (80-.).*, vol. 3717, no. December, 2016.

[74] R. Perez-Jimenez *et al.*, "Single-molecule paleoenzymology probes the chemistry of resurrected enzymes," *Nat. Struct. Mol. Biol.*, vol. 18, no. 5, pp. 592–596, 2011.

[75] V. A. Risso, J. A. Gavira, D. F. Mejia-Carmona, E. A. Gaucher, and J. M. Sanchez-Ruiz, "Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β-lactamases," *J. Am. Chem. Soc.*, vol. 135, no. 8, pp. 2899–2902, 2013.

[76] M. G. Plach, B. Reisinger, R. Sterner, and R. Merkl, "Long-Term Persistence of Bi-functionality Contributes to the Robustness of Microbial Life through Exaptation," *PLoS Genet.*, vol. 12, no. 1, pp. 1–14, 2016.

[77] G. N. Eick, J. K. Colucci, M. J. Harms, E. A. Ortlund, and J. W. Thornton, "Evolution of Minimal Specificity and Promiscuity in Steroid Hormone Receptors," *PLoS Genet.*, vol. 8, no. 11, 2012.

[78] T. Devamani *et al.*, "Catalytic Promiscuity of Ancestral Esterases and Hydroxynitrile Lyases," *J. Am. Chem. Soc.*, vol. 138, no. 3, pp. 1046–1056, 2016.

[79] M. Alcalde, "When directed evolution met ancestral enzyme

resurrection," *Microb. Biotechnol.*, vol. 10, no. 1, pp. 22–24, 2017.

[80]   B. G. Hall, "Simple and accurate estimation of ancestral protein sequences," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 14, pp. 5431–5436, 2006.

[81]   R. Merkl and R. Sterner, "Ancestral protein reconstruction: Techniques and applications," *Biol. Chem.*, vol. 397, no. 1, pp. 1–21, 2016.

[82]   J. W. Thornton, "Resurrecting ancient genes: Experimental analysis of extinct molecules," *Nat. Rev. Genet.*, vol. 5, no. 5, pp. 366–375, 2004.

[83]   I. N. Shindyalov, N. A. Kolchanov, and C. Sander, "Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations?," *Protein Eng. Des. Sel.*, vol. 7, no. 3, pp. 349–358, 1994.

[84]   P. M. Zakas *et al.*, "Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction," *Nat. Biotechnol.*, vol. 35, no. 1, pp. 35–37, 2017.

[85]   N. Barruetabeña *et al.*, "Resurrection of efficient Precambrian endoglucanases for lignocellulosic biomass hydrolysis," *Commun. Chem.*, vol. 2, no. 1, p. 76, 2019.

[86]   F. Robert and M. Chaussidon, "A palaeotemperature curve for the Precambrian oceans based on silicon isotopes in cherts," *Nature*, vol. 443, no. 7114, pp. 969–972, 2006.

[87]   M. Richter, C. Schulenburg, D. Jankowska, T. Heck, and G. Faccio, "Novel materials through Nature's catalysts," *Mater. Today*, vol. 18, no. 8, pp. 459–467, 2015.

[88]   O. Kirk, T. V. Borchert, and C. C. Fuglsang, "Industrial enzyme applications," *Curr. Opin. Biotechnol.*, vol. 13, no. 4, pp. 345–351, 2002.

[89] B. A. Lerma, "Isolation , characterization and applications of nanocellulose produced by ancestral enzymes," 2019.

[90] X. J. Shen, P. L. Huang, J. H. Chen, Y. Y. Wu, Q. Y. Liu, and R. C. Sun, "Comparison of acid-hydrolyzed and TEMPO-oxidized nanocellulose for reinforcing alginate fibers," *BioResources*, vol. 12, no. 4, pp. 8180–8198, 2017.

[91] T. Saito and A. Isogai, "TEMPO-mediated oxidation of native cellulose. The effect of oxidation conditions on chemical and crystal structures of the water-insoluble fractions," *Biomacromolecules*, vol. 5, no. 5, pp. 1983–1989, 2004.

[92] A. Rees *et al.*, "3D bioprinting of carboxymethylated-periodate oxidized nanocellulose constructs for wound dressing applications," *Biomed Res. Int.*, vol. 2015, 2015.

[93] H. J. BLUMENTHAL and S. ROSEMAN, "Quantitative estimation of chitin in fungi.," *J. Bacteriol.*, vol. 74, no. 2, pp. 222–224, 1957.

[94] M. Rinaudo, "Chitin and chitosan: Properties and applications," *Prog. Polym. Sci.*, vol. 31, no. 7, pp. 603–632, 2006.

[95] F. Khoushab and M. Yamabhai, "Chitin research revisited," *Mar. Drugs*, vol. 8, no. 7, pp. 1988–2012, 2010.

[96] F. Shahidi, J. K. V. Arachchi, and Y. J. Jeon, "Food applications of chitin and chitosans," *Trends Food Sci. Technol.*, vol. 10, no. 2, pp. 37–51, 1999.

[97] K. V. Harish Prashanth and R. N. Tharanathan, "Chitin/chitosan: modifications and their unlimited application potential-an overview," *Trends Food Sci. Technol.*, vol. 18, no. 3, pp. 117–131, 2007.

[98] R. Sharp, "A Review of the Applications of Chitin and Its Derivatives in Agriculture to Modify Plant-Microbial Interactions and Improve Crop Yields," *Agronomy*, vol. 3, no. 4, pp. 757–793,

2013.

[99] A. Bhatnagar and M. Sillanpää, "Applications of chitin- and chitosan-derivatives for the detoxification of water and wastewater - A short review," *Adv. Colloid Interface Sci.*, vol. 152, no. 1–2, pp. 26–38, 2009.

[100] T. Hebrew, "The molecular biology of chitin digestion Rachel Cohen-Kupiec and Ilan Chet," *Curr. Opin. Biotechnol.*, pp. 270–277, 1998.

[101] C. R. Berger, M. E. Roloff, and D. R. Roskos-ewoldsen, "Communication Science Edited by," vol. II.

[102] S. Nikolov *et al.*, "Robustness and optimal use of design principles of arthropod exoskeletons studied by ab initio-based multiscale simulations," *J. Mech. Behav. Biomed. Mater.*, vol. 4, no. 2, pp. 129–145, 2011.

[103] A. Ilnicka and J. P. Lukaszewicz, "Discussion remarks on the role of wood and chitin constituents during carbonization," *Front. Mater.*, vol. 2, no. March, pp. 1–6, 2015.

[104] D. Raabe *et al.*, "Microstructure and crystallographic texture of the chitin-protein network in the biological composite material of the exoskeleton of the lobster Homarus americanus," *Mater. Sci. Eng. A*, vol. 421, no. 1–2, pp. 143–153, 2006.

[105] J. Blackwell, "Physical methods for the determination of chitin structure and conformation," *Methods Enzymol.*, vol. 161, no. C, pp. 435–442, 1988.

[106] Y. Fan, T. Saito, and A. Isogai, "Chitin nanocrystals prepared by TEMPO-mediated oxidation of α-chitin," *Biomacromolecules*, vol. 9, no. 1, pp. 192–198, 2008.

[107] G. N. Ramachandran and C. Ramakrishnan, "The structure of chitin," *BBA - Biochim. Biophys. Acta*, vol. 63, no. 2, pp. 307–309, 1962.

[108] Y. Lu, L. Weng, and L. Zhang, "Morphology and properties of soy protein isolate thermoplastics reinforced with chitin whiskers," *Biomacromolecules*, vol. 5, no. 3, pp. 1046–1051, 2004.

[109] M. V. Tzoumaki, T. Moschakis, V. Kiosseoglou, and C. G. Biliaderis, "Oil-in-water emulsions stabilized by chitin nanocrystal particles," *Food Hydrocoll.*, vol. 25, no. 6, pp. 1521–1529, 2011.

[110] J. F. Revol and R. H. Marchessault, "In vitro chiral nematic ordering of chitin crystallites," *Int. J. Biol. Macromol.*, vol. 15, no. 6, pp. 329–335, 1993.

[111] M. V. Tzoumaki, T. Moschakis, and C. G. Biliaderis, "Metastability of nematic gels made of aqueous chitin nanocrystal dispersions," *Biomacromolecules*, vol. 11, no. 1, pp. 175–181, 2010.

[112] H. P. Zhao, X. Q. Feng, and H. Gao, "Ultrasonic technique for extracting nanofibers from nature materials," *Appl. Phys. Lett.*, vol. 90, no. 7, pp. 97–99, 2007.

[113] B. M. Min *et al.*, "Chitin and chitosan nanofibers: Electrospinning of chitin and deacetylation of chitin nanofibers," *Polymer (Guildf).*, vol. 45, no. 21, pp. 7137–7142, 2004.

[114] S. Ifuku *et al.*, "Preparation of chitin nanofibers with a uniform width as α-chitin from crab shells," *Biomacromolecules*, vol. 10, no. 6, pp. 1584–1588, 2009.

[115] M. Mincea, A. Negrulescu, and V. Ostafe, "Preparation, modification, and applications of chitin nanowhiskers: A review," *Rev. Adv. Mater. Sci.*, vol. 30, no. 3, pp. 225–242, 2012.

[116] J. B. Zeng, Y. S. He, S. L. Li, and Y. Z. Wang, "Chitin whiskers: An overview," *Biomacromolecules*, vol. 13, no. 1, pp. 1–11, 2012.

[117] A. S. Sahai and M. S. Manocha, "Chitinases of fungi and plants:

their involvement in morphogenesis and host-parasite interaction," *FEMS Microbiol. Rev.*, vol. 11, no. 4, pp. 317–338, 1993.

[118] T. Oku and K. Ishikawa, "Analysis of the hyperthermophilic chitinase from Pyrococcus furiosus: Activity toward crystalline chitin," *Biosci. Biotechnol. Biochem.*, vol. 70, no. 7, pp. 1696–1701, 2006.

[119] T. Tanaka, T. Fukui, and T. Imanaka, "Different Cleavage Specificities of the Dual Catalytic Domains in Chitinase from the Hyperthermophilic Archaeon Thermococcus kodakaraensis KOD1," *J. Biol. Chem.*, vol. 276, no. 38, pp. 35629–35635, 2001.

[120] K. G. Nair, A. Dufresne, A. Gandini, and M. N. Belgacem, "Crab shell chitin whiskers reinforced natural rubber nanocomposites. 3. Effect of Chemical Modification of chitin whiskers," *Biomacromolecules*, vol. 4, no. 6, pp. 1835–1842, 2003.

[121] Z. Li, R. Yang, F. Yang, M. Zhang, and B. Wang, "Structure and properties of chitin whisker reinforced papers for food packaging application," *BioResources*, vol. 10, no. 2, pp. 2995–3004, 2015.

[122] Y. Feng *et al.*, "Novel nanofibrillated cellulose/chitin whisker hybrid nanocomposites and their use for mechanical performance enhancements," *BioResources*, vol. 13, no. 2, pp. 3030–3044, 2018.

[123] N. I. Platnick and H. D. Cameron, "Cladistic Methods in Textual, Linguistic, and Phylogenetic Analysis," *Syst. Zool.*, vol. 26, no. 4, p. 380, 2006.

[124] J. J. Tehrani, "The phylogeny of little red riding hood," *PLoS One*, vol. 8, no. 11, 2013.

[125] R. S. Walker, K. R. Hill, M. V. Flinn, and R. M. Ellsworth, "Evolutionary history of hunter-gatherer marriage practices," *PLoS One*, vol. 6, no. 4, pp. 2–7, 2011.

[126] V. Hanson-Smith, B. Kolaczkowski, and J. W. Thornton,

"Robustness of ancestral sequence reconstruction to phylogenetic uncertainty," *Mol. Biol. Evol.*, vol. 27, no. 9, pp. 1988–1999, 2010.

[127] J. A. Eisen, "Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis," *Genome Res.*, vol. 8, no. 3, pp. 163–167, 1998.

[128] L. Pauling and L.a.E.Z, "Chemical Paleogenetics." .

[129] G. J. Morgan, E. Zuckerlandl, and L. Pauling, "and the molecular evolutionary clock," *J Hist Biol*, vol. 31, no. 2, pp. 155–178, 1998.

[130] W. M. Fitch and W. M. Fitch, "Society of Systematic Biologists Toward Defining the Course of Evolution : Minimum Change for a Specific Tree Topology TOWARD DEFINING THE COURSE OF EVOLUTION : MINIMUM CHANGE FOR A SPECIFIC TREE T ' OPOLOGY," vol. 20, no. 4, pp. 406–416, 2013.

[131] A. goleman, daniel; boyatzis, Richard; Mckee, "济无No Title No Title," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2019.

[132] Z. Yang, S. Kumar, and M. Nei, "A new method of inference of ancestral nucleotide and amino acid sequences," *Genetics*, vol. 141, no. 4, pp. 1641–1650, 1995.

[133] J. M. Koshi and R. A. Goldstein, "Probabilistic reconstruction of ancestral protein sequences," *J. Mol. Evol.*, vol. 42, no. 2, pp. 313–320, 1996.

[134] T. R. Schultz, R. B. Cocroft, and G. A. Churchill, "The Reconstruction of Ancestral Character States," *Evolution (N. Y).*, vol. 50, no. 2, p. 504, 1996.

[135] J. P. Huelsenbeck, B. Larget, and D. Swofford, "A compound Poisson process for relaxing the molecular clock," *Genetics*, vol. 154, no. 4, pp. 1879–1892, 2000.

[136] A. Eyre-Walker, "Problems with parsimony in sequences of biased base composition," *J. Mol. Evol.*, vol. 47, no. 6, pp. 686–

690, 1998.

[137] P. D. Williams, D. D. Pollock, B. P. Blackburne, and R. A. Goldstein, "Assessing the accuracy of ancestral protein reconstruction methods," *PLoS Comput. Biol.*, vol. 2, no. 6, pp. 0598–0605, 2006.

[138] J. Felsenstein, "Evolutionary trees from DNA sequences: A maximum likelihood approach," *J. Mol. Evol.*, vol. 17, no. 6, pp. 368–376, 1981.

[139] S. Biologists, "Society of Systematic Biologists Empirical and Hierarchical Bayesian Estimation of Ancestral States Author (s): John P . Huelsenbeck and Jonathan P . Bollback Published by : Oxford University Press for the Society of Systematic Biologists Stable URL : h," vol. 50, no. 3, pp. 351–366, 2018.

[140] K. Lai, *Chapter 10 WheatGenome . info : A Resource for Wheat Genomics*, vol. 1374. 2016.

[141] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 22, pp. 10915–10919, 1992.

[142] M. Manuel, "A new semi-subterranean diving beetle of the Hydroporus normandi-complex from south-eastern France, with notes on other taxa of the complex (Coleoptera: Dytiscidae)," *Zootaxa*, vol. 3652, no. 4, pp. 453–474, 2013.

[143] R. V. Engeset *et al.*, "Improving runoff simulations using satellite-observed time-series of snow covered area," *Nord. Hydrol.*, vol. 34, no. 4, pp. 281–294, 2003.

[144] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, "MEGA6: Molecular evolutionary genetics analysis version 6.0," *Mol. Biol. Evol.*, vol. 30, no. 12, pp. 2725–2729, 2013.

[145] J. Castresana, "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis," *Mol. Biol.*

Evol., vol. 17, no. 4, pp. 540–552, 2000.

[146] Y. K. Edward, M. J. Christopher, and T. E. Danielle, "Engineering transcriptional regulation to control Pdu microcompartment formation," *PLoS One*, vol. 9, no. 11, pp. 1–14, 2014.

[147] B. Wang, P. H. Walton, and C. Rovira, "Molecular Mechanisms of Oxygen Activation and Hydrogen Peroxide Formation in Lytic Polysaccharide Monooxygenases," *ACS Catal.*, vol. 9, pp. 4958–4969, 2019.

[148] C. Ouephanit, N. Boonvitthya, M. Theerachat, S. Bozonnet, and W. Chulalaksananukul, "Efficient expression and secretion of endo-1,4-β-xylanase from Penicillium citrinum in non-conventional yeast Yarrowia lipolytica directed by the native and the preproLIP2 signal peptides," *Protein Expr. Purif.*, vol. 160, no. March, pp. 1–6, 2019.

[149] A. Törrönen, A. Harkki, and J. Rouvinen, "Three-dimensional structure of endo-1,4-beta-xylanase II from Trichoderma reesei: two conformational states in the active site.," *EMBO J.*, vol. 13, no. 11, pp. 2493–2501, 1994.

[150] M. G. Godoy, G. M. Amorim, M. S. Barreto, and D. M. G. Freire, *Agricultural Residues as Animal Feed*. Elsevier B.V., 2018.

[151] B. V. McCleary and P. McGeough, "A Comparison of Polysaccharide Substrates and Reducing Sugar Methods for the Measurement of endo-1,4-β-Xylanase," *Appl. Biochem. Biotechnol.*, vol. 177, no. 5, pp. 1152–1163, 2015.

[152] Y. H. P. Zhang and L. R. Lynd, "Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems," *Biotechnol. Bioeng.*, vol. 88, no. 7, pp. 797–824, 2004.

[153] J. Hong, X. Ye, and Y. H. P. Zhang, "Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its

applications," *Langmuir*, vol. 23, no. 25, pp. 12535–12540, 2007.

[154] T. K. Ghose, "Measurement of cellulase activities," *Pure Appl. Chem.*, vol. 59, no. 2, pp. 257–268, 1987.

[155] E. Breslmayr *et al.*, "A fast and sensitive activity assay for lytic polysaccharide monooxygenase," *Biotechnol. Biofuels*, vol. 11, no. 1, pp. 1–13, 2018.

[156] G. Janusz, A. Pawlik, J. Sulej, U. Świderska-Burek, A. Jarosz-Wilkolazka, and A. Paszczyński, "Lignin degradation: Microorganisms, enzymes involved, genomes analysis and evolution," *FEMS Microbiol. Rev.*, vol. 41, no. 6, pp. 941–962, 2017.

[157] A. J. Drummond and A. Rambaut, "BMC Evolutionary Biology BEAST: Bayesian evolutionary analysis by sampling trees," *BioMed Cent.*, vol. 7, no. 214, pp. 1–8, 2007.

[158] D. L. Ayres *et al.*, "BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics," *Syst. Biol.*, vol. 61, no. 1, pp. 170–173, 2012.

[159] A. Rambaut, A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard, "Posterior summarization in Bayesian phylogenetics using Tracer 1.7," *Syst. Biol.*, vol. 67, no. 5, pp. 901–904, 2018.

[160] Z. Yang, "PAML 4: Phylogenetic analysis by maximum likelihood," *Mol. Biol. Evol.*, vol. 24, no. 8, pp. 1586–1591, 2007.

[161] J. Adachi and M. Hasegawa, "Model of amino acid substitution in proteins encoded by mitochondrial DNA," *J. Mol. Evol.*, vol. 42, no. 4, pp. 459–468, 1996.

[162] A. Filipski, O. Murillo, A. Freydenzon, K. Tamura, and S. Kumar, "Prospects for building large timetrees using molecular data with incomplete gene coverage among species," *Mol. Biol. Evol.*, vol. 31, no. 9, pp. 2542–2550, 2014.

[163] Y. Kato and D. J. Nevins, " Enzymic Dissociation of Zea Shoot

Cell Wall Polysaccharides ," *Plant Physiol.*, vol. 75, no. 3, pp. 759–765, 1984.

[164] P. Sá-Pereira, M. Costa-Ferreira, and M. R. Aires-Barros, "Enzymatic properties of a neutral endo-1,3(4)-β-xylanase Xyl II from Bacillus subtilis," *J. Biotechnol.*, vol. 94, no. 3, pp. 265–275, 2002.

[165] R. Bourbonnais and P. Claire, "Hicrobielngy," pp. 201–206, 1986.

[166] T. Pupko, I. Pe'er, R. Shamir, and D. Graur, "A fast algorithm for joint reconstruction of ancestral amino acid sequences," *Mol. Biol. Evol.*, vol. 17, no. 6, pp. 890–896, 2000.

[167] M. Galla, K. Wicke, and M. Fischer, "On the statistical inconsistency of Maximum Parsimony for k -tuple-site data," vol. 0, no. 0, pp. 1–23, 2013.

[168] M. J. Bailey, P. Biely, and K. Poutanen, "Interlaboratory testing of methods for assay of xylanase activity," *J. Biotechnol.*, vol. 23, no. 3, pp. 257–270, 1992.

[169] M. Devarapalli and H. K. Atiyeh, "A review of conversion processes for bioethanol production with a focus on syngas fermentation," *Biofuel Res. J.*, vol. 2, no. 3, pp. 268–280, 2015.

[170] J. Hu, V. Arantes, and J. N. Saddler, "The enhancement of enzymatic hydrolysis of lignocellulosic substrates by the addition of accessory enzymes such as xylanase: Is it an additive or synergistic effect?," *Biotechnol. Biofuels*, vol. 4, no. 1, p. 36, 2011.

[171] R. Kumar and C. E. Wyman, "Effect of xylanase supplementation of cellulase on digestion of corn stover solids prepared by leading pretreatment technologies," *Bioresour. Technol.*, vol. 100, no. 18, pp. 4203–4213, 2009.

[172] A. Requejo, A. Rodríguez, J. L. Colodette, J. L. Gomide, and L. Jiménez, "TCF bleaching sequence in kraft pulping of olive tree

pruning residues," *Bioresour. Technol.*, vol. 117, pp. 117–123, 2012.

[173] C. Characterization, O. F. Pulp, C. In, U. Softwood, K. Fibers, and R. With, "CHEMICAL CHARACTERIZATION OF PULP COMPONENTS IN," vol. 2010, no. 2006, pp. 616–629, 2010.

[174] E. Lindfors and S. Forest, "Structural Changes in Lignin During Kraft Pulping," vol. 38, pp. 151–158, 1984.

[175] J. Hu, D. Tian, S. Renneckar, and J. N. Saddler, "Enzyme mediated nanofibrillation of cellulose by the synergistic actions of an endoglucanase , lytic polysaccharide monooxygenase (LPMO) and xylanase," *Sci. Rep.*, no. October 2017, pp. 4–11, 2018.

[176] A. Peciulyte *et al.*, "Redox processes acidify and decarboxylate steam-pretreated lignocellulosic biomass and are modulated by LPMO and catalase," *Biotechnol. Biofuels*, vol. 11, no. 1, pp. 1–16, 2018.

[177] F. Sabbadin *et al.*, "An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion," *Nat. Commun.*, vol. 9, no. 1, 2018.

[178] U. F. Rodríguez-Zúñiga, D. Cannella, R. D. C. Giordano, R. D. L. C. Giordano, H. Jørgensen, and C. Felby, "Lignocellulose pretreatment technologies affect the level of enzymatic cellulose oxidation by LPMO," *Green Chem.*, vol. 17, no. 5, pp. 2896–2903, 2015.

[179] Y. Gumulya *et al.*, "Engineering highly functional thermostable proteins using ancestral sequence reconstruction," *Nat. Catal.*, vol. 1, no. 11, pp. 878–888, 2018.

[180] E. A. Gaucher, J. M. Thomson, M. F. Burgan, and S. A. Benner, "Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins," *Nature*, vol. 425, no. 6955, pp. 285–288, 2003.

[181] J. Bouchard, N. Abatzoglou, E. Chornet, and R. P. Overend, "Characterization of depolymerized cellulosic residues - Part 1: Residues obtained by acid hydrolysis processes," *Wood Sci. Technol.*, vol. 23, no. 4, pp. 343–355, 1989.

[182] M. Åkerholm, B. Hinterstoisser, and L. Salmén, "Characterization of the crystalline structure of cellulose using static and dynamic FT-IR spectroscopy," *Carbohydr. Res.*, vol. 339, no. 3, pp. 569–578, 2004.

[183] Y. Cao and H. Tan, "Study on crystal structures of enzyme-hydrolyzed cellulosic materials by X-ray diffraction," *Enzyme Microb. Technol.*, vol. 36, no. 2–3, pp. 314–317, 2005.

[184] S. Fujisawa, Y. Okita, H. Fukuzumi, T. Saito, and A. Isogai, "Preparation and characterization of TEMPO-oxidized cellulose nanofibril films with free carboxyl groups," *Carbohydr. Polym.*, vol. 84, no. 1, pp. 579–583, 2011.

[185] W. L. Teng, E. Khor, T. K. Tan, L. Y. Lim, and S. C. Tan, "Concurrent production of chitin from shrimp shells and fungi," *Carbohydr. Res.*, vol. 332, no. 3, pp. 305–316, 2001.

[186] F. Boßelmann, P. Romano, H. Fabritius, D. Raabe, and M. Epple, "The composition of the exoskeleton of two crustacea: The American lobster Homarus americanus and the edible crab Cancer pagurus," *Thermochim. Acta*, vol. 463, no. 1–2, pp. 65–68, 2007.

[187] D. Raabe, C. Sachs, and P. Romano, "The crustacean exoskeleton as an example of a structurally and mechanically graded biological nanocomposite material," *Acta Mater.*, vol. 53, no. 15, pp. 4281–4292, 2005.

[188] B. Bissaro, I. Isaksen, G. Vaaje-Kolstad, V. G. H. Eijsink, and Å. K. Røhr, "How a Lytic Polysaccharide Monooxygenase Binds Crystalline Chitin," *Biochemistry*, vol. 57, no. 12, pp. 1893–1906, 2018.

[189] T. Fukazimo and K. J. Kramer, "Mechanism of Chitin Hydrolysis

By the Binary Chitinase Retention Time (Min),” *Insect Biochem.*, vol. 15, no. 2, pp. 141–145, 1985.

[190] M. S. Reid, S. A. Kedzior, M. Villalobos, and E. D. Cranston, “Effect of Ionic Strength and Surface Charge Density on the Kinetics of Cellulose Nanocrystal Thin Film Swelling,” *Langmuir*, vol. 33, no. 30, pp. 7403–7411, 2017.

[191] Y. Tang, S. Yang, N. Zhang, and J. Zhang, “Preparation and characterization of nanocrystalline cellulose via low-intensity ultrasonic-assisted sulfuric acid hydrolysis,” *Cellulose*, vol. 21, no. 1, pp. 335–346, 2014.

[192] L. Liu *et al.*, “High Axial Ratio Nanochitins for Ultrastrong and Shape-Recoverable Hydrogels and Cryogels via Ice Templating,” *ACS Nano*, vol. 13, no. 3, pp. 2927–2935, 2019.

[193] M. Lewin, “Oxidation and aging of cellulose,” *Macromol. Symp.*, vol. 118, pp. 715–724, 1997.

[194] K. Ahn *et al.*, “Yellowing and brightness reversion of celluloses: CO or COOH, who is the culprit?,” *Cellulose*, vol. 26, no. 1, pp. 429–444, 2019.

[195] W. T. Goodrich, J. D.; Winter, “Alpha-chitin nanocrystals prepared from shrimp shells and their specific properties,” *Biomacromolecules*, no. 8, p. 252− 257., 2007.

[196] M. M. De Souza Lima and R. Borsali, “Rodlike cellulose microcrystals: Structure, properties, and applications,” *Macromol. Rapid Commun.*, vol. 25, no. 7, pp. 771–787, 2004.

[197] Y. Fan, H. Fukuzumi, T. Saito, and A. Isogai, “Comparative characterization of aqueous dispersions and cast films of different chitin nanowhiskers/nanofibers,” *Int. J. Biol. Macromol.*, vol. 50, no. 1, pp. 69–76, 2012.

[198] J. D. Goodrich and W. T. Winter, “α-Chitin nanocrystals prepared from shrimp shells and their specific surface area measurement,”

Biomacromolecules, vol. 8, no. 1, pp. 252–257, 2007.

[199] B. Focher, A. Naggi, G. Torri, A. Cosani, and M. Terbojevich, "Chitosans from Euphausia superba. 2: Characterization of solid state structure," *Carbohydr. Polym.*, vol. 18, no. 1, pp. 43–49, 1992.

[200] "27- 270," vol. 31, no. 2006, p. 38041, 2007.

[201] M. Kaya *et al.*, "Differentiations of Chitin content and surface morphologies of chitins extracted from male and female grasshopper species," *PLoS One*, vol. 10, no. 1, 2015.

[202] M. Poletto, H. L. Ornaghi Júnior, and A. J. Zattera, "Native cellulose: Structure, characterization and thermal properties," *Materials (Basel).*, vol. 7, no. 9, pp. 6105–6119, 2014.

[203] Q. Wang, X. Yan, Y. Chang, L. Ren, and J. Zhou, "Fabrication and characterization of chitin nanofibers through esterification and ultrasound treatment," *Carbohydr. Polym.*, vol. 180, pp. 81–87, 2018.

[204] W. C. Tsai, S. T. Wang, K. L. B. Chang, and M. L. Tsai, "Enhancing saltiness perception using chitin nanomaterials," *Polymers (Basel).*, vol. 11, no. 4, 2019.

[205] J. Majtán, K. Bíliková, O. Markovič, J. Gróf, G. Kogan, and J. Šimúth, "Isolation and characterization of chitin from bumblebee (Bombus terrestris)," *Int. J. Biol. Macromol.*, vol. 40, no. 3, pp. 237–241, 2007.

[206] Y. Kato, J. Kaminaga, R. Matsuo, and A. Isogai, "TEMPO-mediated oxidation of chitin, regenerated chitin and N-acetylated chitosan," *Carbohydr. Polym.*, vol. 58, no. 4, pp. 421–426, 2004.

[207] E. Szymańska and K. Winnicka, "Stability of chitosan - A challenge for pharmaceutical and biomedical applications," *Mar. Drugs*, vol. 13, no. 4, pp. 1819–1846, 2015.

[208] H. Moussout, H. Ahlafi, M. Aazza, and M. Bourakhouadar,

"Kinetics and mechanism of the thermal degradation of biopolymers chitin and chitosan using thermogravimetric analysis," *Polym. Degrad. Stab.*, vol. 130, pp. 1–9, 2016.

[209] M. Ziegler-Borowska, D. Chełminiak, and H. Kaczmarek, "Thermal stability of magnetic nanoparticles coated by blends of modified chitosan and poly(quaternary ammonium) salt," *J. Therm. Anal. Calorim.*, vol. 119, no. 1, pp. 499–506, 2015.

[210] A. A. Oun and J. W. Rhim, "Effect of oxidized chitin nanocrystals isolated by ammonium persulfate method on the properties of carboxymethyl cellulose-based films," *Carbohydr. Polym.*, vol. 175, pp. 712–720, 2017.

[211] Y. Zhang, C. Xue, Y. Xue, R. Gao, and X. Zhang, "Determination of the degree of deacetylation of chitin and chitosan by X-ray powder diffraction," *Carbohydr. Res.*, vol. 340, no. 11, pp. 1914–1917, 2005.

[212] V. Moses, R. Hatherley, and Ö. Tastan Bishop, "Bioinformatic characterization of type-specific sequence and structural features in auxiliary activity family 9 proteins," *Biotechnol. Biofuels*, vol. 9, no. 1, pp. 1–17, 2016.

[213] H. Ehrlich *et al.*, "Discovery of 505-million-year old chitin in the basal demosponge Vauxia gracilenta," *Sci. Rep.*, vol. 3, pp. 17–20, 2013.

[214] C. C. Loron, C. François, R. H. Rainbird, E. C. Turner, S. Borensztajn, and E. J. Javaux, "Early fungi from the Proterozoic era in Arctic Canada," *Nature*, vol. 570, no. 7760, pp. 232–235, 2019.

[215] M. L. Romero-Romero, V. A. Risso, S. Martinez-Rodriguez, E. A. Gaucher, B. Ibarra-Molero, and J. M. Sanchez-Ruiz, "Selection for protein kinetic stability connects denaturation temperatures to organismal temperatures and provides clues to archaean life," *PLoS One*, vol. 11, no. 6, pp. 1–38, 2016.

[216] A. K. Garcia, J. W. Schopf, S. I. Yokobori, S. Akanuma, and A. Yamagishi, "Reconstructed ancestral enzymes suggest long-Term cooling of Earth's photic zone since the Archean," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 18, pp. 4619–4624, 2017.

[217] M. Richter, C. Schulenburg, D. Jankowska, T. Heck, and G. Faccio, "Novel materials through Nature's catalysts," *Mater. Today*, vol. 18, no. 8, pp. 459–467, 2015.

Appendix

In appendix I I show all the sequences that were used in the alienation to construct the phylogenetic trees. Underline I show the NCBI code to find the sequences on their internet database, on its right the organisms name of the corresponding protein. Below the amino acid sequence. In appendix II I show the ancestral sequences of each node and their corresponding query sequences.

APPENDIX I

List of Endo-1,4-beta-Xln A

proteins from the species used in the construction of the phylogenetic tree

<u>P18429</u> *Bacillus subtillis (Query)*
KNFLVGLSAALMSISLFSATASAASTDYWQNWTDGGGIVNAVNGSGG
NYSVNWSNTGNFVVGKGWTTGSPFRTINYNAGVWAPNGNGYLTLYG
WTRSPLIEYYVVDSWGTYRPTGTYKGTVKSDGGTYDIYTTTRYNAPSI
DGDRTTFTQYWSVRQSKRPTGSNATITFSNHVNAWKSHGMNLGSNWA
YQVMATEGYQSSGSSNVTVW
<u>ACT21830.1</u> *Bacillus megaterium*
KNFLVGLTAAFMSISMFSATASAAGTDYWQNWTDGGGTVNAVNGSG
GNYSVNWSNTGNFVVGKGWTTGSPFRTINYNAGVWAPNGNGYLTLY
GWTRSPLIEYYVVDSWGTYRPTGTYKGTVKSDGGTYDIYTTTRYNAPS
IDGDRTTFTQYWSVRQTKRPTGSNATITFSNHVNAWKSHGMNLGSNW
AYQVMATEGYQSSGSSNVTVW
<u>AAM08360.1</u> *Bacillus circulans*
KNFLVGLSAALMRIILFSATASAASTDYWQNWTDGGGIVNAVNGSGG
NYSVNWSNTGNFVVGKGWTTGSPFRTINYNAGVWAPNGNGYLTLYG
WPRSPLIEYYVVDSWGTYRPTGTYKGTVKSDGGTYDIYTTTRYNAPSID
GDRTTFTQYWSVRQTKRPTGSNATITFSNHVNAWKSHGMNLGSNWAY
QVMATGGYQSSGSSNVTVW
<u>WP_075747623.1</u> *Bacillus licheniformis*
KKFLVGLTAALMSISLFSATASAAGTDYWQNWTDGGGVVNAVNGAG
GNYSVNWSNTGNFVVGKGWTTGSPFRTINYNAGVWAPNGNGYLTLY
GWTRSPLIEYYVVDSWGTYRPTGTYKGTVKSDGGTYDIYTTTRYNAPS
IDGDNTTFTQYWSVRQTKRPTGSNATITFSNHVNAWKSHGMNLGSNW
SYQVLATEGYQSSGSSNVTVW
<u>WP_046226114.1</u> *Paenibacillus dauci*
KKFLTTILAVSMSSSVFAATSSAATDYWQNWTDGGGTVNSVNGTGGN
YSVNWQNTGNFVVGKGWTTGSPNRVINYNAGLWAPSGNGYLTLYGW
TRNSLIEYYVVDSWGTYRPTGTYKGTVTSDGGVYDIYTTTRYDAPSIEG
DHKTFTQYWSVRQTKRPTGSNATITFSNHVKAWASHGMKLGSTWSYQ
VMATEGYQSSGSSNVTVW

<u>WP_060532780.1</u> *Paenibacillus bovis*
KKFLTTILAISMSSSVFAATSSAATDYWQNWTDGGGTVNSVNGNGGN
YSVNWQNTGNFVVGKGWTTGSPNRVINYNAGLWAPSGNGYLTLYGW

TRNSLIEYYVVDSWGTYRPTGTYKGTVTSDGGVYDIYTTTRYDAPSIEG
DHKTFTQYWSVRQTKRPTGSNATITFSNHVKAWASHGMNLGSTWSYQ
VLATEGYQSSGSSNVTVW

WP_046216042.1 *Paenibacillus wulumuqiensis*
KKFLTTILAISMSSSVFAATSSAATDYWQNWTDGGGTVNAVNGAGGN
YSVNWQNTGNFVVGKGWTTGSPNRVVNYNAGQWAPSGNGYLTLYG
WTRNSLIEYYVVDSWGTYRPTGTYKGTVTSDGGVYDIYTTTRYDAPSI
EGDHKTFTQYWSVRQTKRPTGSNATITFSNHVKAWASHGMNLGSTWS
YQVMATEGYKSSGSSNVTVW

CAJ87325.1 *Thermobacillus xylanilyticus*
YWQYWTDGIGYVNATNGQGGNYSVSWSNSGNFVIGKGWQYGAHNR
VVNYNAGAWQPNGNAYLTLYGWTRNPLIEYYVVDSWGSYRPTGDYR
GSVYSDGAWYDLYHSWRYNAPSIDGTQTFQQYWSVRQQKRPTGSNVS
ITFENHVNAWGAAGMPMGSSWSYQVLATEGYYSSGYSNVTVW

WP_015253740.1 *Thermobacillus composti*
KLMAVLVAVAMVSALFAVNGASYWQYWTDGIGYVNATNGQGGNYS
VSWSNSGNFVIGKGWQYGAHNRVVNYNAGAWQPNGNAYLTLYGWT
RNPLIEYYVVDSWGSYRPTGDYRGSVYSDGAWYDLYHSWRYNAPSID
GTQTFQQYWSVRQQKRPTGSNVSITFENHVNAWGAAGMPMGSSWSY
QVLATEGYYSSGYSNVTVW

WP_089855761.1 *Halolactibacillus miurensis*
MFFLAIVFCFALILPTVIVTSNETGTHDYEYWKDSGGSGIMTLNSGGTFS
AEWSNVNNILFRKGKKFDETQTSINY-
GANYQPNGNSYLTVYGWAVDPLVEFYIVDSWGSYRPTGTHKGTINVD
GGTYDIYETTRTNQPSIKG-TATFKQYWSVRTSKRTSGTISV---
SEHFKAWENLGMQMGNMY-EVALTVEGYQSSGSADVYSN

WP_090795741.1 *Pelagirhabdus alkalitolerans*
KTKYALMAICFISLLATVTSNELGTHDYEFWKDSGGSGSMTLGNGGTF
SAEWSNVNNILFRKGQRFDETRTSIDYGVNYQPNGNSYLTVYGWSVDP
LVEYYIVDSWGTWRPPGANSGTVQVDGGTYDIYETTRVNQPSIRGTQT
FQQYWSVRRDKRTSGTISVSEHFRAWEQHGMPLGNLY-
EAALTIEGYQSSGQANVYRN

WP_080062516.1 *Clostridium hungatei*
KKLFGLFLAVSMSLTSLFALPSYAATDYWQNWTDGGGTVRATNGSGG
NFSVTWQNCGNFVVGKGWGSGNASRVCNYNAGQFAPSGNGYLTFYG
WTRNALIEYYIVDSWGTYRPTGTYKGSVSSDGGTYDIYTTQRVNAPSID
G-
TKTFTQYWSVRQSKRATGSNVAITFSNHVNAWKSKGMNLGSSWSYQA
LCVEGYQSSGSANVTVW

WP_020814600.1 *Clostridium papyrosolvens*
KKVLGLIFAISMSSMFAGTTYAYWQNWTDGGGTVNVTNGSGGNFSVN
WYNCGNFVVGKGWNYGTANRVCNYNAGVFSPSGNGYLSFYGWTRNS
LIEYYIVDTWGTYRPTGTYKGSVTSDGGTYDIYTTQRYNAPSIDGTQSF
TQFWSVRQSKRPTGQNVQINFANHVNAWKSKGMNLGSSWSYQALCV
EGYQSSGSANVTVW

WP_054739261.1 *Cellulosilyticum ruminicola*
KKLLTVFCIATMALATNVQATDYWQNWTDGGGYVNAVNGQGGNYS
VSWSNCGNFVVGKGWNYGTPNRVVGYNAGAFQPSGNGYLSFYGWTR
NSLIEYYIVDTWGTYRPTGTYKGSVYSDGGNYDIYTTMRYNQPSIDG-
TQTFQQYWSVRQSKRPTGTNSQITFQNHVNAWQRCGMYLGSNWSYQ
ALCVEGYQSSGYANVTVW
WP_103201672.1 *Herbinix hemicellulosilytica*
RKAIALIFAASLCSTMLAGNAYYWQYWTDGGGYVNAVNGSGGNFSV
NWYNCGNFVVGKGWATGSPNRVINYNAGVWAPSGNGYLTLYGWTR
NALIEYYVVDSWGNWRPPGSNSGTVYSDGGTYDIYRTMRYNAPSIDG-
TQTFAQYWSVRTSKRPTGSNVSITFANHVNAWASRGLYLGSSWSYQV
MAVEGYQSSGSANVTVW
WP_078766814.1 *Eubacterium uniforme*
AGVLCAVVALVPPLTATKVYDSKTGKIDFSLWKDYGNTSMKLNGGGA
FECQWSNIGNALFRTGKTLDSSKRSVNYSANYQPNGNSYLCVYGWSR
NPLVEYYIVDSWGSWRPTGQYKGQISVDGGTYDLYQDTRYNAPSIDGN
TTFTQFWSVRTSKRTSGTISVSKHFDAWAQHGMTLGNLT-
EVALTVEGYKSSGYANVTNA
WP_015906728.1 *Caldicellulosiruptor bescii*
KVLIAVLMCFMLGNPFYALTSNASGTYYYELWKDSGNTTMTVDTGGR
FSCQWSNINLFRTGKKFSTAWGTVKITYSATYNPNGNSYLCIYGWSRNP
LVEFYIVESWGTWRPPGATSGTVTIDGGTYDIYKTTRVNQPSIEGTTTFD
QYWSVRTSKRTSGTVTVTDHFKAWAAKGLNLGTID-
QITLCVEGYQSSGSANITQN
WP_013411078.1 *Caldicellulosiruptor owensensis*
KLLLAVLMCVVLGNPFYALTSNASGTYYYELWKDSGNTTMTVDTGGR
FSCQWSNINLFRTGKKFSTAWGTVKITYSATYNPNGNSYLCIYGWSRNP
LVEFYIVESWGSWRPPGATSGTVTIDGGTYDIYKTTRVNQPSIEGTTTFD
QYWSVRTSKRTSGTVTVTDHFKAWAAKGLNLGTID-
QITLCVEGYQSSGSANITQN

SLC61209.1 *Mycobacterium abscessus*
KKFLVGLTAAFMSISMFSATASAAGTDYWQNWTDGGGTVNAVNGSG
GNYSVNWSNTGNFVVGKGWTTGSPFRTINYNAGVWAPNGNGYLTLY
GWTRAPLIEYYVVDSWGTYRPTGTYKGTVKSDGGTYDIYTTTRYNAPS
IDGDNTTFTQYWSVRQSKRPTGSNAAITFSNHVNAWKSHGMNLGSNW
AYQVLATEGYKSSGSSNVTVW
WP_033332249.1 *Actinomadura madurae*
IGAIGAIILTAAMALMPAITSNQTGTHFYSFWTDSSGTVSMELGSGGNY
STSWRNTGNFVAGKGWSTGGRRSVNYSGSFNPSGNAYLTLYGWTRNP
LIEYYIVDNWGTYRPTGTYKGTVTSDGGTYDIYETTRYNAPSIEGTRTF
QQYWSVRQSKRTGGTISAGNHFDAWAGHGMNLGSHD-
YQILATEGYQSSGNSNITIG
WP_015772628.1 *Jonesia denitrificans*

ATLASVVALVAGTLAMATINDNTQGTHFYTHWTDAPGTASITLGSGGN
YSTRWSNTGNFVSGKGWATGSRRTVNYSAQFNPSGNGYLTLYGWTRG
PLVEYYIIESWGTYRPTGEFKGTVTTDGGTYDIYKTTRVNKPSIESDYST
FDQYWSVRQQKRTSGSITVANHFDAWARHGMNLGTHD-
YQVMATEGYQSSGSSNVTIS
WP_095567270.1 *Plantactinospora*
RLLTGGACAVALVAAILPITSNQTGTNFFSFWTDGGGSVSMDLGSGGN
YSTTWSNVGNFVAGKGWKPGTRQTVTYSGSFNPSGNAYLALYGWTR
NPLVEYYIVDNWGTYRPTGTYKGTVTSDGGTYDIYETTRYNAPSIEGN
QTFNQYWSVRQSKKTGGTITTGNHFDAWASHGMNLGSHD-
YQILATEGYQSSGNSNITVG
WP_012877523.1 *Xylanimonas cellulosilytica*
TIAAAGASLAMLVAGVAAISTNGTGNSFYSFWTDSPGTVSMELGSGGN
YSTSWSNTGNFVAGKGWSTGSARTITYSGSFNPSGNAYLTVYGWSRNP
LVEYYIVDSWGTYRPTGTHMGTVTSDGGTYDIYKTTRTNSPSIEGDSST
FDQFWSVRQSKRVGGTVTTANHFNAWASKGMNLGSHD-
YQIVATEGYQSSGSSNITVG
SFD76457.1 *Streptomyces aidingensis*
ACAALVGAALLPITQNSTGTDFYSFWTDAPGTVTMNTGSGGNYSTQW
SNTGNFVAGKGWSTGSSRTVTYSGTFNPSGNAYLTLYGWSRNPLVEYY
IVDNWGTYRPTGTYKGTVNSDGGTYDIYETTRNNAPSIEGTQTFKQFW
SVRQQKRTGGTITTGNHFNAWAGHGMHLGSHD-
YQIMATEGYQSSGSSNITVG
WP_100203418.1 *Streptomyces*
MVGAALLPITQNSTGTDFYSFWTDAPGTVTMNTGSGGNYSTQWSNTG
NFVAGKGWSTGSSRTVTYSGTFNPSGNAYLTLYGWSRNPLVEYYIVDN
WGTYRPTGTYKGTVNSDGGTYDIYETTRNNAPSIEGTQTFKQFWSVRQ
QKRTGGTITTGNHFNAWAGHGMHLGSHD-
YQIMATEGYQSSGSSNITVG
CAJ57849.1 *Cellulomonas flavigena*
KQLVAGVAVAGLVGLAIAVTSNQTGNSYYSFWTDSQGTVSMELGSGG
NYSTSWRTTGNFVAGKGWQTGSGRACSYSGQFTPSGNAYLTLYGWTT
GPLIEYYIVDNWGTYRPTGTYMGTVNSDGGTYDIYRTQRVNQPSIQGT
ATFYQYWSVRQQKRTGGTITAANHFNAWASKGMNLGSHN-
YQILATEGYQSSGSSNITVS
WP_061296570.1 *Herbidospora cretacea*
GLAAAGIWATVIAMVLMSVTTNQTGTNWFSFWTDSQGTVSMELGSGG
NYSTSWRNTGNFVAGKGWQTGGRRSVSYSGSFNPSGNAYLTLYGWTR
NPLIEYYIVDNWGTYRPTGSYKGTVTSDGGTYDIYETTRTNAPSIEGTRT
FKQFWSVRQQKKTGGTITSGNHFDAWSRAGMQLGNHD-
YMILATEGYQSSGSSNITIG
WP_033216338.1 *Kitasatospora phosalacinea*
ARVLPLALAAVVLLPGTAVTTNQTGNDFYSFWSDGNGSSSMSLNGGG
NYGTSWSNVGNFVAGKGWSTGSRNPVTYSGTWSTNGNAYLSLYGWT
TNPLVEYYVVENYGSYKPTGSYKGTVTTDGGTYDIYETTRYNAPSIEG-

220

NKTFNQYWSVRQSKRTGGTITVGNHFDAWARYGMNLGTMN-
YEILATEGYQSSGNSNITLG

WP_077689185.1 *Nocardiopsis alkaliphila*
LTGAGALAVAAPSLLLPGLTENQTGTHFYSFWTDGGGSVSMTLGSGGS
YSTQWRDTGNFVCGKGWHNGGYRNVNYSGSFHPSGNGYLCLYGWTS
NPLVEYYIVEAFGTYRPEGDYRGTVYSDGGTYDMYHTMRYNAPSVEG
DSRTFPQFWSVRQSQRTGGTINTGNHFNAWGGAGMQLGSFDHYMILA
TEGYQSSGSSNLSVS

WP_051235018.1 *Marinimicrobium agarilyticum*
KSLSFALLTGATALGVSTICDNQTGNNYYTHWTDGGGSACMTLGSEG
NYSYEWSNTGNFVGGKGWSTGASNRIIGYNAGNYSPSGNSYLTLYGW
TTGPLVEYYVVDSWGSYRPGGTPAGTVTTDGGTYDLYRTQRVNKPSIE
G-NTTFYQYWSVRTSKRPQGNNNTITFQNHVDAWARQGWNLGSHN-
YQVLATEGWQSSGSSNVSVW

WP_052481306.1 *Gilvimarinus agarilyticus*
KKSLTAFIGAAAALCITALTSNQTGTHYYSFWTDAPGTVSMTLGSGGN
YTSQWSNTGNWVGGKGWNPGGRRTVSYSGTFNPSGNGYLTLYGWTT
SPLIEYYIVDNWGSYRPSGTYYGTVNTDGGTYDIYRTQRVNQPSIQGTA
TFYQYWSVRQQKRTGGTITTGNHFDAWASHGLNLGNHD-
YMVMATEGYQSSGNSNITLG

WP_020210149.1 *Gilvimarinus chinensis*
VKKIASLVAATTALCFTSLTSNQTGTHYYSFWTDAPGTVSMTLGSGGN
YSSQWSNTGNWVGGKGWNPGGRRTVNYSGTFNPSGNGYLTLYGWTR
SPLIEYYIVDNWGTYRPSGTYYGTVYSDGGTYDIYRTQRVNQPSIEGTA
TFYQYWSVRQQKRTGGTITTGNHFDAWASHGLNLGNHD-
YMVMATEGYQSSGNSNITLG

AAT81215.1 *Microbulbifer hydrolyticus*
RSTLSVFALGVIASLGFSLTSNQTGNHFYSFWTDAPGTVSMTLGAGGN
YSSSWSNTGNWVGGKGWNPGGRRTVNYSGTFNPSGNGYLTLYGWTR
NPLVEYYIVDNWGSYRPSGTYYGTVNTDGGTYDIYRTQRVNKPSIEGD
SSTFYQYWSVRQQKRTGGTITAGNHGNHFDAWASHGMNLGTHD-
YMVMATEGYQSSGNSNITLG

BAB79287.1 *Pseudomonas*
KKSLAILTTAAAVLSVAALTSNQTGTHYYSFWTDAPGTVSMTLGSGGN
YSSQWSNTGNWVGGKGWNPGGRRTVSYSGTFNPSGNGYLTLYGWTT
SPLIEYYIVDNWGSYRPSGTYYGTVNTDGGTYDIYRTQRVNQPSIQGTA
TFYQYWSVRQQKRTGGTITTGNHFDAWASHGLNLGNHD-
YMVMATEGYQSSGNSNITLG

WP_077337594.1 *Vibrio ruber*
KKFLTILTVTAAALSATVLTSNETGIHYYSFWTDAPGTVSMTLGDGGN
YSSQWHNTGNWVGGKGWNPGGRRTVNYSGTFNPSGNGYLALYGWTT
NPLVEYYIIDNWGSYRPNGTYYGTVNTDGGTYDIYRTLRVDKPSIEGNK
STFYQYWSVRQQKRTGGTITTGNHFDAWASKGLNLGTFN-
YMVMATEGYQSSGSSNITLG

WP_077316466.1 *Vibrio gazogenes*
KKFLTILTVTAGVLSATALTSNETGTHYYSFWTDAPGTVSMTLGNGGS
YSSQWRNTGNWVGGKGWNPGGRRSVQYSGTFSPSGNGYLALYGWTT
NPLVEYYIIENWGSYRPNGTYYGTVNTDGGTYDIYRTLRVDKPSIEGDR
STFYQYWSVRQQKRTGGTITTGNHFDAWASKGLNLGKFN-
YMVMATEGYQSSGSSNITLG
WP_084717381.1 *Vibrio rhizosphaerae*
KKLLTILTVTAAALSATVLTSNETGTHYYSFWTDAPGTVSMTLGNGGN
YSSQWRNTGNWVGGKGWNPGGRRSVNYSGTFNPSGNGYLALYGWTT
NPLIEYYIVDNWGSYRPGGTYYGTVNTDGGTYDIYRTLRVDKPSIEGNK
STFYQYWSVRQQKRTGGTITTGNHFDAWASKGLNLGKFN-
YMIMATEGYQSSGSSNIVLG
WP_078045385.1 *Cellvibrio*
KNIFTAFIGAAALCATLSSNSTGTHYYSFWKDSGNATMTLQAGGRYTS
QWNSGNWVGGKGWNPGNSTRVINYSGSYSNSQNSYLALYGWTRSPLI
EYYVIESYGSYNPGGTDYGSFQSDGATYNVRRCQRVNQPSIDGNQTFY
QYFSVRNPKKISGTITFANHVNFWASKGLNLGNHN-
YQVLATEGYQSTGSSDITVS

WP_073319678.1 *Aestuariibacter aggregatus*
KRLAMGLSLLTCSLGVTALYENETGTHYYSFWKDSGNANFTLQPGGR
YQSQWNSGNWVGGKGWNPGGRKTVNYSGYYDSSQNSYLALYGWSQ
NPLVEYYIIESYGSYDPGGVDYGTFQSDGATYKIRRCQRVQQPSIEGTQ
TFYQYFSVRQPIKISGTITTGNHFDAWASVGLNLGTFN-
YMVMATEGYQSSGSSDITVS
WP_015817458.1 *Teredinibacter turnerae*
NSALAAAVLAATATVFTALTSNSTGTHYYSFWKDSGDASFTLHNGGR
YASEWNSTNNWVGGKGWNPGGAKVVNYEGYYSNSQNSYLALYGWT
KNPLIEYYIIESYGSYNPGGTNYGSFQSDGATYNVRRCQRVQQPSIEGTA
TFYQYFSVRSPKKISGTINVGNHFNYWASQGLNLGSHD-
YMVLATEGYRSTGSSDISV

C-LPMO

proteins from the species used in the construction of the phylogenetic tree

AAD27623.1 Streptomyces viridosporus (query)
HGVAMVPGSRTYLCQLDAITGTGALNPTNPACRDALNKSGSSALYNW
FAVLDSRAAGRGPGYVPDGTLCSAGDRSPYDFSAYNAARADWPRTHL
TSGATVKVQYSNWAAHPGDFRVYLTKPGWSPTSPLGWNDLELIQTVT

NPPQQGSPGTNGGHYYWDLKLPSGRSGDRLIFMQWVRSDSQENFFSCS
DI

WP_037928552.1 *Streptomyces toyocaensis*
HGVAMMPGSRTYLCQLDAVTGTGALDPSNPACRDALNESGATALYN
WFAVLDSNAGGRGAGYVPDGTLCSAGDRSPYDFSAYNAARADWPRT
HLTSGATIKVQYSNWAAHPGDFRVYLTKPGWSPTSQLGWDDLELIQTV
TNPPQQGSPGTNGGHYYWNLNLPSGRSGDALIFMQWVRSDSQENFFSC
SDV

WP_055629450.1 *Streptomyces hirsutus*
HGVAMMPGSRTYLCQLDATTGTGALDPTNPACRAALNQSGASALYN
WFAVLDSQAAGRGPGYVPDGKLCSAGDRSPYNFSAYNAARSDWPRTH
LTSGATIKAQYSNWAAHPGDFRVYITKPGWSPTSQLGWNDLELIQTVT
NPPQQGSPGTNGGHYYWNLKLPSGRSGDALIFMQWVRSDSQENFFSCS
DV

WP_044379538.1 *Streptomyces cyaneogriseus*
HGVAMMPGSRTYLCQLDAVSGTGALNPTNPACKDALSKSGATALYN
WFAVLDSNAGGRGAGYVPDGKLCSAGDRSPYDFSAYNAARADWPRT
HLTSGATIKVQYSNWAAHPGDFRVYVTKPGWSPTSPLGWSDLELIQTV
TNPPQQGGAGSNGGHYYWDLKLPSGRSGDALIFMQWVRSDSQENFFS
CSDI

WP_055607049.1 *Streptomyces prasinus*
HGVAMMPGSRTYLCQLDAITGTGSLDPTNPACRAALNQSGASALYNW
FAVLDSRAAGRGQGYVPDGTLCSAGDRSPYDFSAYNAARSDWPRTHL
TSGATIKAQYSNWAAHPGDFRVYITKPGWSPTSQLGWNDLELIQTVTD
PPQQGSPGANGGHYYWNLKLPSGRSGDAVIFMQWVRSDSQENFFSCSD
V

KMS77933.1 *Streptomyces leeuwenhoekii*
HGVAMMPGSRTYLCQLDAISGTGALNPTNPACKDALSKSGSTALYNW
FAVLDSNAGGRGAGYVPDGKLCSAGDRSPYDFSAYNAARADWPRTHL
TSGATIKVQYSNWAAHPGDFRVYVTKPGWSPTSQLGWSDLELIQTVTN
PPQQGGAGSNGGHYYWDLKLPSGRSGDALIFMQWVRSDSQENFFSCS
DI

WP_067284436.1 *Streptomyces jeddahensis*
HGVAMMPGSRTYLCQLDAITETGALDPTNPACQDALNESGANALYNW
YAVLDAKAGGRGPGYVPDGTLCSAGDRSPYDFSAYNAARADWPRTHL
TSGATMRVQYSNWAAHPGDFRVYLSKPGWSPTSTLGWDDLELIQTVT
DPPQQGSPGNDGGHYYWDLTLPSGRSGDALIFIQWVRSDSQENFFSCSD
V

WP_102928211.1 *Streptomyces diastaticus*
HGVAMMPGSRTYLCQLDAKTGTGALDPTNPACQAALDQSGATALYN
WFAVLDSNAGGRGAGYVPDGTLCSAGDRSPYDFSAYNAARGDWPRT
HLTSGATIPVQYSNWAAHPGDFRVYLTKPGWSPTSELGWDDLELIQTV
TNPPQQGSPGTDGGHYYWDLALPSGRSGDALIFMQWVRSDSQENFFSC
SDV

WP_030181860.1 *Streptomyces violaceorubidus*

HGVAMMPGSRTYLCQLDAKTGTGALDPTNPACRSALDRSGATALYN
WFAVLDSNAGGRGAGYVPDGTLCSAGDRSPYDFSAYNAARSDWPRTH
LTSGATIPVQYSNWAAHPGDFRVYLTKPGWSPTSELGWDDLELIQTVT
NPPQQGSPGTDGGHYYWNLALPSGRSGDALIFMQWVRSDSQENFFSCS
DV
WP_012879227.1 *Xylanimonas cellulosilytica*
HGAEVFPGSRQYLCWVDGQSASGALDPSNPACASALAVSGSGAYYNW
FGNLDSNGAGRTEGYIPDGTICSGGDRGPYDFSAFNAPRTDWPTTHLTA
GATYEFQHNNWAQHPGRFDVYVTRAGFDPTKPLGWSDLELIDSVTNPP
DTGGPGSD-NYYYWDVTLPADRTGRHIVFTHWVRSDSTENFYSCSDV

WP_043615549.1 *Nonomuraea candida*
HGVSMFPGSRTFLCWQDGLRDNGQILPYNPACAAAVNQSGTTPLYNW
FAVLRSDGAGRTTGFIPDGQICSGGTGGPYDFSAYNAVRNDWPLTHLTS
GATIQMRHSNWAEHPGSFKYYVTKNGWNPNGPLKWSDLEQFGSVTNP
PKSGGAGGL-NYYYWNAQLPSGKSGRHIIFTHWIRSDSNENFYSCSDV
WP_084959626.1 *Thermoactinospora rubra*
HGVTMFPGSRTFLCWQDGLRENGQIIPYNPACAAAVAQSGTTPLYNWF
AVLRSDGGGRTTGFIPDGQICSGGTGGPYDFTPYNAVRNDWPLTHLTA
GRTIQVRHSNWAAHPGSFVVYITKNGWNPNAPLRWADLESIGSVANPP
QSGGPGEL-NYYYWNLQLPSGKTGRHIIFIHWIRSDSQENFYSCSDV
WP_103960203.1 *Nonomuraea solani*
HGVSMFPGSRTFLCWQDGLRDNGQIIPYNPACGAAVNQSGATPLYNWF
AVLRSDGAGRTTGFIPDGQICSGGTGGPYDFSAYNAVRSDWPLTHLTSG
ATIQMRHSNWAEHPGSFNYYVTKNGWNPNGPLKWSDLEPFGSVTNPP
KTGGAGAL-NYYYWNATLPSGKSGRHIIFTHWIRSDSNENFYSCSDV
WP_083929395.1 *Catelliglobosispora koreensis*
HGATMFPGSRTWLCYQDGLRPNGAIEAYNPACAAAIAQNGTTPLYNW
FAVLRSDAAGRTVGFIPDGQLCSAGTGGPYNFSAYNAARTDWPLTHLT
TGANIQFRHSNWAHHPGTFYLAITRQGWSPTSPLAWSDLQEFASVTNPP
ANGGPGAL-NYYYWNAQLPTGRSGRAIIYIRWVRSDSNENFFSCSDV
WP_088645651.1 *Micromonospora wenchangensis*
HGTPIVPGSRTYLCWQDGLTTTGEIRPNNPACSAAVAQSGTNSLYNWFS
VLRSDAGGRTVGFIPDGKLCSGGNTGFAGYDLARTDWPLTHLTSGARL
DFKYSNWAHHPGTFYFYVTKDSWSPTRPLAWSDLEPFLTVTNPPQQGA
SGTNDGHYYFSGNLPSGKSGRHIIYSRWVRSDSQENFFGCSDV
WP_018216208.1 *Salinispora pacifica*
HGAAMVPGSRTFLCWRDGLSPTGEIQPYNPACSAAVDQSGANSLYNW
FSVLRSDAAGRTVGFIPDGQLCSGGNSGFLGYDLARTDWPLTHLTAGQ
SIEFRYSNWAHHPGTFYFYITKDSWSPTRPLAWNDLEPFLTVTNPPQRG
GPGTNDGHYYFTGTLPADKSGRHLIYSRWVRSDSPENFFGCSDV
WP_098463482.1 *Isoptericola jiangsuensis*
HGAQMFPGSRQYFCWVDGTSAGGDISPTNPACADAVAASGKTPLYNW
FGNLDSNGAGRTVGYVPDGRICDGGGRGPYDFEPYNAPRTDWPTTHLT

AGDTYEFRHNNWAAHPGRFDVYLTTEGFDPTAPLAWDDLELVDSVTD
PPASGGPGDL-
NHYSWDVTLPADRTGRHVVFTHWVRSDSAENFYSCSDV
WP_067315554.1 *Micromonospora rifamycinica*
HGTPIVPGSRTYLCWQDGLSPTGEIRPNNPACSAAVAQSGTNSLYNWFS
VLRSDAGGRTVGFIPDGKLCSGGNTGFAGYDLARTDWPLTHLTSGARL
DFKYSNWAHHPGTFYFYVTKDSWSPTRPLAWSDLEPFLTVTNPPQRGA
SGTNEGHYYFSGNLPSGKSGRHIIYSRWVRSDSQENFFGCSDV
WP_023362433.1 *Actinoplanes friuliensis*
HGAMMAPGSRTYFCWKDGLSGNGSIVPKNPACAAAVSQSGTTPLYNW
FAVLRSDGGGRTSGFIPDGQLCSGGTGGPYDFTGYNLARNDWPQTHLT
AGASMRFDYNNWAKHPGTFSLYITKDSWSPTRPLAWSDLEPFSQVVNP
TSVGGPGTNDGRYSWTANLPSGKSGKHLIYSVWSRSDSQETFYGCSDV
WP_043522496.1 *Actinoplanes utahensis*
HGAIQVPGSRTWFCYQDGRNPTGAIEPKNPACAAAVAQSGVTSLYNWF
AVLRSDGAGRVDGFIPDGQLCSGGTGGPYNFSGFNLARNDWPTTHLTA
GANVQFKYNNWAKHPGTFSLYITKDGYDPTKPLAWSDLEPFDQVTNPP
ANGGPGTDDGHYYWNGKLPAGKTGKHLIYSVWSRSDSQETFYGCSDV
WP_051759284.1 *Herbidospora cretacea*
HGAMMMPGSRTYLCWKDGLNSSGAIQPNNPACAAAVAQSGTNSLYN
WFATLRSDGAGRMSGFIPDGSLCSGGAVVYNFSGFDLARNDWPVTHLT
SGANIQIRYNKWAAHPGTFRLYITKNTWSPTRPLAWSDLEPFDSSTNPP
SVGNPGNENAYYYWNANLPSGKSGRHIIYSVWQRSDSNETFYGCSDV
WP_013131295.1 *Thermobispora bispora*
HGAMMMPGSRTYLCWKDGLTPQGNIVPNNPACAAAVAQSGTNALYN
WFAVLRSDGAGRTRGYIPDGKLCSADAKVYDFSGFDLARDDWPVTHL
TAGATIQIRYNMWAHHPGTFRLYVTKDSWDPNRPLSWDDLEPFSEVTD
PPSVGSPGNEDAYYYWNAKLPENKSGRHIIYSIWQRSDSQETFYNCSDV
WP_081476174.1 *Verrucosispora maris*
HGAAMTPGARTYLCWRDGLSPTGEIRPQNPACSAAVAQSGANSLYNW
FSVLRSDAGGRTTGFIPDGQLCSGGATGFRGFDLARDDWPLTHLTAGR
TMEFRYSNWAHHPGTFYFYVTKNSWSPNRALAWSDLEPFLTVTNPPQR
GAVGTNDGHYYFTGTLPANKSGRHIIYSRWVRSDSQENFFGCSDV
PRY54227.1 *Glycomyces artemisiae*
HGATVFPGSRQYLCYFDALGGNGALDPYNEACQSALDQVGPTPFYNW
FGNLDSNGAGKTVGYIPDGQICDGGGRGPYDFSAFNDPGNWPKTHVTA
GDTVQWRYNNWAHHPGRFDLYITKDGWDPSQPLAWDDLEKFQTFTN
PASNGGPGSDDHYYYANLTLPQ-
KSGYHMVFTHWVRSDSNENFYACSDV
SDZ55585.1 *Asanoa ishikariensis*
HGTPMSPGSRTFLCWKDGLSQQGDIQPHNPACTAAVQQGGTNPLYNW
FSVLRSDAAGRTVGFIQDGQLCSGGNSTFAAYNQARNDWPLTHLTAGS
RFDFKYSNWAHHPGTFYFYVTRDSWSPTRALAWNDLEPFLTVTNPPQN
GGPGTNEGHYYFSGNLPSNKSGRHLIYSRWVRADSNENFFGCSDV
WP_068755309.1 *Thermobifida cellulosilytica*

HGAMTYPATRSYQCYRDGDNGGGGLAPTNPVCQNLLAENGNYPFYN
WFGNLISNAGGRHREIIPDGKLCGPHPQFSGLNAVSEHWPTTTLQAGSTI
TFQYNAWAPHPGTWYLYVTKDGWDPNSPLGWDDLEPFDQVTNPPIRQ
G-GPEGPEYYWESTLPN-KSGRHIIYSIWQRSDSPEAFYDCTDV
WP_061783670.1 *Thermobifida fusca*
HGAMTYPPTRSYICYVNGEGGGGNIAPTNPACQNLLAENGNYPFYNWF
GNLISDAGGRHREIIPDGQLCGPHPQFSGLNLVSEHWPTTTLVAGSTITF
QYNAWAPHPGTWYLYVTKDGWDPNSPLGWDDLEPFHTVTDPPIRPG-
GPEGPEYYWDATLPN-KSGRHIIYSIWQRSDSPEAFYDCTDV
WP_036015456.1 *Lentzea albidocapillata*
HGSMVWPGSRTYLCYEDGAGGGGDLQPTNPACAAAVAQGGKQPLWD
WFGNLISNAAGRHREIIPDGNLCGPTTKYSAYNLTRADWPVTQVQANS
TVTLRYNAWAPHPGTWEQYVTKDGFDPTMPLKWSDLEPFDKVTNPPL
ANGEYAWQARLPN-KSGRHIIYSLWQRSDSPEAFYSCSDV
WP_090096265.1 *Lentzea jiangxiensis*
HGSMVWPGSRTYLCYEDGVGGGGDLQPTNPACAAAVAQGGKQPLWD
WFGNLISNAAGRHREIVPDGSLCGPTTKFAAYNLPRADWPTTQVTANS
TVTFRYNAWAPHPGTWEQYVTKDGFDATKPLKWSDLEPFDKVTNPPL
ANGEYAWQARLPN-KTGRHIIYSLWQRSDSPEAFYSCSDV
AEN11025.1 *Streptomyces*
HGVAMTPGSRTYLCQLDALSGTGALNPTNPACRDALSQSGANALYNW
FAVLDSNAGGRGAGYVPDGSLCSAGDRSPYDFSAYNAARADWPRTHL
TSGATLKVQYSNWAAHPGDFRVYLTKPGWAPTSELAWDDLQLVQTVS
NPPQQGGAGTNGGHYYWDLALPSGRSGDALMFIQWVRSDSQENFFSC
SDI
WP_048404375.1 *Chromobacterium violaceum*
HGTMEIPINRTYSCFLEGPEAPKTPACQAAKQAGGTQAMYDWNGINQN
PSGDNHQAVVPDGTLCSGGKAEFKGFNLAREDWRTTNIVPDANGNYE
FIYKATAPHAFKFYVTKNGWNPTQALKWSDLELFGTVGNPQQDAGKR
YHMTLKLPQGKTGKHIIFNVWKRSDSEEAFYSCSDV
WP_076226074.1 *Chromobacterium pseudoviolaceum*
HGTMEIPINRTYSCFLEGPEAPKTPACQAAKQAGGTQAMYDWNGINQN
PSGDNHQAVVPDGTLCGGGKAEFKGFNLAREDWRTTNIVPDANGNYE
FIYKATAPHAFKFYVTKNGWNPTQALKWSDLELFGTVGNPQQDAGKR
YHMTLKLPQGKTGKHIIFNVWKRSDSEEAFYSCSDV
WP_071113852.1 *Chromobacterium sphagni*
HGTMEVPINRTYSCYKEGAESPKSAACQAAKQAGGVQAMYDWNGIN
QNPPGDNHQAVVPDGTLCAAGQSKFKGFNLARTDWPATNIVPNASGN
FEFIYKAPAPHAFRFYVTKNGWNPSQPLKWSDLELFGTYGNPPLDANQ
RYHMTMKLPTGKTGRHIIYNVWKRSDSEEAFYSCSDV
WP_047244066.1 *Chromobacterium subtsugae*
HGTMEVPINRTYSCFLEGPEAPKTAACQAARQVGGTQAMYDWNGINQ
NPQGDNHQAAVPDGTLCGGGKAEFKGFNLARRDWRATSIVPDANGNY
EFIYKATAPHAFKFYITKNGWNPDLPLKWSDLELFGTAGNPQQDAAKR
YHMTLKLPQGKTGQHIIFNVWKRSDSEEAFYSCSDV

WP_107801020.1 *Chromobacterium panama*
HGTMEVPINRTYNCFKEGAESPKSAACQEARRVGGTQAMYDWNGINQ
NPQGDNHAAVVPDGKLCAGGQDKFKGFNLGRTDWPKTKLVPDANGN
YEFVYHATAPHSFKFYVTKNGWDQTKPLKWSDLDPFHTVGNPPLDAN
KRYHMVAKLPAGKSGPHIIFNVWKRADSEEAFYSCSDV

WP_081689029.1 *Inquilinus limosus*
MPISRVYACYQEGPETPKSEACKAAKAVGGAQAFYDWNGVRQGDAG
GNHQAVVPDGQLCSGGGAEFRGLDLARPDWTKSSIAPTAAGDYTFVW
KATAPHAFQYYITKNGWSPNAPLKWSDLEAFATVGSAAQKVGDQYR
MTVKLPQGRKGDHVIYGIWQRADSREAFYACSDV

WP_057949360.1 *Lysobacter enzymogenes*
HGSMVHPKSRIYACKQGDAENPSDPACRAAWQVSGSAMFYDWMSINR
ADADGRHRAVVPDGQLCSGGNPTFAGLDLARADWQAQPIAADANGRF
AFLFKGTAPHAWTFYITRDGWNPERPLHWSDLEPFCTLGNVPLGSDGN
YRLDCPLPK-RSGRHVIYNTWQRSDSTEAFYTCMDV

WP_010482246.1 *Pseudomonas geniculata*
HGTMTTPISRVYACFQGNPENPTNPACAAAKAVGGSQAFYDWNGINQ
ANANGNHQAVVPDGKLCSGNNPTFRGLDVNRSDWQTTPIHADANGKF
TFVFKATAPHAWRFYVTREGWQPGSPLGWADLQEFCTLGNTPLSADG
TYKLQCTLPQ-RTGQHVIYNTWQRSDSTEAFYTCMDV

WP_049412354.1 *Stenotrophomonas maltophilia*
HGTMTTPISRVYACFQGNPENPTNPACAAAKAIGGSQAFYDWNGINQA
NANGNHQAVVPDGKLCSGNNPTFRGLDVNRSDWQTTPIHADANGKFT
FVFKATAPHAWRFYVTREGWQPGSPLGWADLQEFCTLGNTPLSADGT
YKLQCTLPQRTGQHVIYNTWQRSDSTEAFYTCMDV

WP_053443696.1 *Stenotrophomonas maltophilia*
HGTMTTPISRVYACFQGNPENPTNPACAAAKAIGGSQAFYDWNGINQA
NANGNHQAVVPDGKLCSGNNPTFRGLDVNRSDWQTTPIHADANGKFT
FVFKATAPHAWRFYVTREGWQPGSPLGWADLQEFCTLGNTPLSADGT
YKLQCTLPQ-RTGQHVIYNTWQRSDSTEAFYTCMDV

WP_074725696.1 *Stenotrophomonas pavanii*
HGTMTTPVSRVYACFQGNPENPTNPACAAAKAVGGSQAFYDWNGINQ
ANANGNHQAVVPDGKLCSGNNPTFRGLDVNRSDWPTTPIQPDANGKF
TFVFKATAPHAWRFFVTREGWQPGSPLRWADLQEFCTLGNTPLSADGT
YKLQCTLPQ-RSGQHVIYNTWQRSDSTEAFYTCMDV

WP_100553964.1 *Stenotrophomonas lactitubi*
HGTMSKPSSRVYACYQGNPENPTNPACAAAKAIGGAQPFYDWSGINQ
ANANGNHQAVVPDGQLCSGGNSKYRGLDLDRSDWSTTPIRADGNGRY
TFEFLAPAPHAWKFYVTREGWQPGSPVGWSDLQEFCTLGNVPLSEGGV
YKLDCPLPQ-RTGQHLIYNTWQRSDSAEAFYTCMDV

WP_061513394.1 *Legionella pneumophila*
HGSMEVPISRTYQCYKEGPESPKSAACKAAVQIGGTQPLYNWHEVNQA
AANDRHRELIADGQLCAGGRDFFKGFNLARNDWTTTVVHPDSNGRFQ
FVYVATAPHKFKFYVTKDGYDFNTPLKWSDLEPFCTITSVALANGRYQ
MDCPLPANKTGKRIFYVIWQREDSPEAFYSCSDV

WP_095984428.1 *Cystobacter fuscus*
HGSIEIPISRVYNCYKEGPETPQSAACKAAIAYGGTQAFYDWNGVRQG
NANGQHRALIPDGKLCSAANESHKGLDLARSDWPAKRIAPNAQGRFDF
VYHATAPHAFQFFVTRQGYNPTQPLKWSDLEPFCTVGTTPLQNNRYTL
NCPFPTGRTGRHVIYNIWQRSDSPEAFYACVDV

WP_071901846.1 *Cystobacter ferrugineus*
HGSIEIPISRVYNCYKEGPETPQSAACKAAIAYGGTQAFYDWNGVRQG
NANGQHRALIPDGKLCSAANESHKGLDLARSDWPAKRIAPNAQGRFDF
VYHATAPHAFQFFITRQGYNPSQPLKWSDLEPFCTVGSTPLQNNRYTLN
CPFPTGKTGRHVIYNIWQRSDSPEAFYACVDV

WP_046713188.1 *Myxococcus fulvus*
HGSMEVPLSRVYSCFKEGPETPKSAACKAAIQSGGTQALYDWNGVRQ
GNANGRHRELIADGKLCSAANESHRGLDLARTDWPATLISPDSSNRFDF
VFHATAVHAFHLYVTKEGYNPALPLKWSDLEPFCTLTAPAPQNNRYRF
TCPFPQGRTGQHVIYAIWQRSDSPEAFYACTDV

WP_074725695.1 *Stenotrophomonas pavanii*
HGTLSKPFSRVYACYQGNPENPSHPACAAAKAIGGAQPFYDWAGINQA
EADGNHQAVVPDGELCSGGNSKYRGLDLNRSDWQTSPIRADARGRYT
FEFKAPAPHAWRFYVTREGWQPGSPLRWADLEAFCTLGNVPLSGDVY
KLDCPLPK-RSGQHVIYNTWQRSDSKEAFYTCADV

WP_088497014.1 *Stenotrophomonas maltophilia*
HGTMTTPVSRVYACFQGNPENPTNPACAAAKAIGGSQAFYDWNGINQ
ANANGNHQAVVPDGKLCSGNNPTFRGLDVNRSDWQTTPIQPDANGRF
TFVFKATAPHAWRFFVTREGWQPGGPLRWADLQEFCTLGNTPLSADG
TYKLQCTLPQ-RSGQHVIYNTWQRSDSTEAFYTCMDV

Ch-LPMO

Monooxygenases proteins from the species used in the construction of the phylogenetic tree

WP_098855344.1 *Bacillus thuringiensis (Query)*
MKKNSLQKMKKVILSGGVLLTGLLTFGFSEKASAHGYVESPASRSYLC
KQGVNVNCGPIQYEPQSVEGIGGFPQLGPSDGQIAGAGHFPALDVQTV
DRWKKVTLNGGTNTFKWKLTAPHSTKEWKYYITKKGWNPNKPLTRS
DLDLVPFYVKNDGGARPGTTVTHEANVPTDRSGYHLILAVWEIADTGN
AFYQVIDVNLLNNGLVSNFAFNNVVQTPTLF

WP_000742280.1 *Bacillus cereus*
MKKNSLQKMKKVILSGGVLLTGLLTFGFSEKASAHGYVESPASRSYLC
KQGVNVNCGPIQYEPQSVEGIGGFPQLGPSDGQIAGAGHFPALDVQTV
DRWKKVTLNGGTNTFKWKLTAPHSTKEWKYYITKKGWNPNKPLTRS
DLDLVPFYVKNDGGARPGTTVTHEANVPTDRSGYHLILAVWEIADTGN
AFYQVIDVNLLNNGLSSNFAFNNVVQTPTLF

WP_097855297.1 *Bacillus toyonensis*
MKKTSLQKMKKVMLSGGILLTGLLTFGFSEKASAHGYVESPASRSYLC
KQGVNVNCGPVQYEPQSVEGIGGFPQLGPSDGQIAGAGHFPALDAQSV
DRWKKITLNGGTNTFKWKLTAPHSTKEWKYYITKKDWNPNMPLTRSS
LDIVPFYVKNDGGVRPGTTVTHEANVPTDRSGYHLILAVWEIADTGNA
FYQVIDVNLVNNGLNSNFAFNNAVQVPTLF

WP_017149289.1 *Bacillus bingmayongensis*
RKSKHFQGFQKAMLSCGVLFTGVLTFGFTEKASAHGYVESPASRSYLC
KQGVNTNCGPIQYEPQSVEGPGSFPQLGPSDGQIAGAGHFPDLDVQTVD
RWKKVPLSGGKNTFQWKLTAPHSTKEWKYYITKKDWDPNKPLSRSDL
DLVPFYVKNDGGIRPGTTVIHEANVPTDRNGYHLILAVWEIADTGNAF
YQVIDVNLVNNGVTDSAHQSSYLPSSHLF

WP_033676157.1 *Bacillus gaemokensis*
RSRRKLKNSIRYSLLCGMCMVSLVSFGFAQKASAHGYVESPASRSYLC
KQGVNLNCGPIQYEPQSVEGPGLFPQLGPSDGKIAGAGHFPDLDVQTV
DRWKKVTLSGGKNKFQWKLTAPHSTKEWKYYITKKDWNPNKPLTRF
DLDLVPFYVKNDGGVRPGTTVIHEANVPTDRSGYHLILAVWEIADTGN
AFYQVIDVNLVNNGLEGKMEFNSNIHFPLRF

WP_084157953.1 *Bacillus manliponensis*
RKGNRLKDSIRISLLCGACMIGLSTFSFAGKASAHGYVESPASRAYLCK
QGVNVNCGLVQYEPQSIEARGSFPQLGPSDGQIAGGGIFPELDAQSIDR
WEKVTLNGGKNTFQWKLTAPHSTREWKYYITKKDWDPNKPLTRSQLE
TIPFYQKNDGGVRPGNTVVHEANVPTDRSGYHLILAVWEIADTPNAFY
QVIDVNLVNNGLPNVISSLHTQPSVFPY

WP_089967265.1 *Lihuaxuella thermophila*
MAQRSF--
WCKWMVSCGMMLLVGATLVFAESASAHGYIESPASRAYLCKQGLNK
DCGQIRWEPQSVEGPGSFPQSGPPDGQLAGTSRFPELNVQTADRWHKV
NIRGGNNTFQWYLTAPHATKEWKYYITRKGWNPDKPLARADLELFCH
YYDGGKRPPNTVTHTCNVPTDRSGYHVILGVWEIADTGNAFYQAMDV
NLVNGVSNVQVPSAHHKSSIGNY

WP_006676405.1 *Paenibacillus dendritiformis*
IGSTLLSKVSPLFTAFGLIVLGLASAIFADSASAHGYIESPASRAYQCKLG
MNTNCGQVQYEPQSVEAKGNFPEGGPADGHIAGGGIFAPLDEQSADR
WNKVKMQGGTNTFQWHLTAPHATSEWKYYITKKDWDPNKPLTRADL
DPVPFCTIQDGGKKPPATVTHECSVPTDRSGYHLILGVWEIADTGNAFY
QVIDVDLVNDGSGIELPSAPAQGDKVKH

WP_087435284.1 *Paenibacillus apiarius*
IWNHLVSKVSPLFAAFGLTALGLATLIFADSASAHGYIESPTSRAVLCKQ
GLNKDCGQVQYEPQSVEGKGSFPQGGPADGQIAGGGVFAELNEQTAE
RWHKVTLQGGKNTFKWYLTAPHATSEWKYYITKKDWDPNKPLTRAD
LDPVPFCSIDDGGKRPPNNVTHECTVPTDRSGYHLILGVWEIADTGNAF
YQVIDVNLMNGGTGEQAPAAPGDGDQVKH

WP_087445137.1 *Paenibacillus thiaminolyticus*
IGSTLLSKVSPLFTAFGLIVLGLASAIFADSASAHGYIESPASRAYQCKLG
MNTNCGQVQYEPQSVEAKGNFPVSGPADGHIAGGGIFAPLDEQSADR
WNKVKMQGGTNTFQWHLTAPHATSEWKYYITKKDWDPNKPLTRADL
DPVPFCTIQDGGKKPPATVTHECSVPTDRSGYHLILGVWEIADTGNAFY
QVIDVDLVNDGSEIELPAAPGQGDKVKH
WP_036222403.1 *Lysinibacillus sphaericus*
MNTKFTKLSKKGLSAALLAAGIIGISTAPNAYAHGFVEKPASRAALCTQ
ALNLMCGSIMYEPQSLEAPKGFPQSGPMDGKIASAGFGGILDQQTTDR
WFKNTITGGPNTFTWKYTAPHLTSKWHYYITKRNWNPNKPLTRADLEP
IGTVAHNGSAASTNLTHTINVPTDRSGYHVILAVWDVADTSNAFYNVI
DVNLVNNDGNVDTEAPTQPGDKVEY
WP_082332692.1 *Lysinibacillus contaminans*
MKLHIFNSKSLSVIFGALLMLVLTFTFSSTASAHGFVSKPESRGLLCKTG
ANINCGGVQYEPQSLEAPGNFPIGGPADGQIAGANTFPKLDEQSKYRWT
KVPMKSGANTFEWTLTAAHATAKWDYYLTKQDWNPDQPLKRADLEL
FCSINDGGKKPPFNVKHNCEVPERSGYQVILAVWSIADTPNAFYNVIDA
NFDGTTIPPIDPGPIDPGDKVKY
WP_036094696.1 *Listeria newyorkensis*
MKKFLGKSVLFVVAFALLFATFAPGASAHGYISKPESRAYLAKLGINQN
AGAIQWEPQSVEATGNFPLGGPSDGTIAGGGKFPEMDAQTPDRWHHID
LLGGQNNFTWTLTALHRTKEYKYYITKKGWNPNSLLTRGDLELLTTID
AHNEVPSTTVNHQVPLPTDRNGYYVVLGVWEIADTGNAFYQAVDVNL
INNGTMTLK
WP_070239340.1 *Listeria monocytogenes*
MKKITKIGMFFAVFMLAVVLFETTASAHGYISKPASRVYLANKGINVG
VGSAQYEPQSVEAPKGFPASGPADGSIAGGGKYSLLDAQTANRWAKV
DIESGPLTVEWTLTAPHRTSSWQYFITKKGWDPNKPLTRASLEPLTTIEA
DGSVPNALTKQEINIPDDRSGYYVILGVWNIADTGNAFYQVIDANIINSS
VAPAVDNEAPTGDKVIY

WP_010772090.1 *Enterococcus caccae*
MKKKLLGLALMSTIILGGGLLSAEEASAHGYVEKPAARGYQGSLDKNT
KYGMVITNPQSLEFDKGFPQAGPADGQIASGGQITDMDSTGLNRWTKQ
DINTGVNTFTWHYTAPHSTTKWHYYMTKVGWNPDKPLSRADFDFLGE
VKHDGSAASNNKSHQITVPENRSGYHVILAVWDVADTTNAFYNVIDV
NVKGGGEITPPVEETPAGDTVTF
WP_069664391.1 *Enterococcus termitis*
MMKKLVGLTMMATICLAGAGVLFSDEANAHGYVESPVSRGYQGALD
KNTKYGNVITNPQSLEAPKGFPEKGPADGRIASGGQIGDLDSTGVDRW
TKQEAKTGVNTFTWHYTAPHATSKWHYYITKNGWNPEKPLARENMEL
IGEIKHDGSTASNNKSHQVTIPSDHTGYHVVLAVWDVADTSNAFYNVI
DLNLKSDGGTTPPVEETPAGEKVVY
WP_034688872.1 *Enterococcus mundtii*

MKKSTLIGLGFILAGIGSTLTSVEVAAHGYVSEPASRGYQGSLDKNTKY
GAVINEPQSLEAPKGFPAAGPEDGQIASGADFLLDQQTASLWTKQQLT
AGANDFTWTFTANHSTTKWHYYMTKAGWDQNDVLTRDDLEFIGEVG
NNGQLASTNPTHSINIPNDRLGYHVILAVWDVADTKNGFYNVIDVDVK
GETVLADLQK

WP_076633771.1 *Lactobacillus plantarum*
VKTKKNNVFRYLLVIIGMMVALLGFEAQASAHGFVTNPGGRAYLGST
GLNTNIGSVMYEPQSIEAPKNTIDGKIASAGKFSQLDEQTANRWYKTPV
KAGNLDVTWRLTARHKTSTWDYYITKPGWNPNAPLKFSDFKKIASYN
DNGAMPSEFMTHQVNIPSNEKGYQVLLSVWNIADTGNAFYQVSDINVQ

WP_021731441.1 *Lactobacillus*
LSTKKNRILRYLFVIVGMMIAFLGSVVQASAHGFVTNPGGRAYLGSTGL
NTNIGSVMYEPQSIEAPKNTFIDGKIASAGNFAQLDEQNANRWYKTPVR
AGKLSVTWQLTARHKTSTWDYYLTKPGWNPNAPLKFSDFKKIASYND
NGAMPNEFVTHQVNIPADEKGYQILLSVWNIADTGNAFYQVSDLNVQ

WP_055538846.1 *Streptomyces neyagawaensis*
PRTFTRKRAKWYAAGVGLATMGALVFSSGASGHGYTDLPVSRQKLCQ
NGTVTNCGDIQWEPQSVEGPKGFPGSGPADGQICSAGRFGQLNSPTGG
AWPTTRVTGGQNTFRWQFTAMHATTDFKYYVTKPGWNQNHALARSD
LNLTPFFTVPYNGQRPPSTLSHTGRLPSGLSGHHVILAVWTIADTGNAF
YACSDVTF

WP_055712380.1 *Streptomyces torulosus*
PRTFTRKRAKWYAAGVGLATMGALVLSSGASGHGYTDLPVSRQKLCQ
NGTVTNCGDIQWEPQSVEGPKGFPGSGPADGQICSAGRFGQLNSPTGG
AWPTTRVTGGQNTFRWQFTAMHATTAFTYYVTKPGWNQNHALARSD
LNLTPFFTVPYNGQRPPSTLSHTGRLPSGLSGHHVILAVWTIADTGNAF
YACSDVTF

WP_005486591.1 *Streptomyces bottropensis*
MRHHARNRTKWYAAAVGLATTGAFVLSSGATGHGYTDLPVSRQKLC
QNGSVTNCGDIQWEPQSVEGPKGFPGSGPADGRLCSGGRFNQLDASTG
GAWPTTRVTGGQNTFRWQFTAMHATSDFKYYVTRAGWNQNHALTRS
DLNLTPFFTVPYNGQRPPSTLSHTGRLPSGLSGHHVILAVWTIADTANA
FYACSDVTF

WP_033531196.1 *Streptomyces galbus*
MRKHTRSRTKWYAAGVGLATTGALVLSSGATGHGYTDLPVSRQKLCQ
NGTVTNCGDIQWEPQSVEGPKGFPGSGPADGQICSGGRFNQLNASTGG
AWPTTRVTGGQNTFRWQFTAMHATTDFKYYVTKAGWNQNHALARSD
LNLTPFFTVPYNGQRPPATLSHTGRLPSGLSGHHVIVAVWTIADTTNAF
YACSDVTF--

WP_046731057.1 *Streptomyces humi*
MRKRTKVSAVVLGVATAGAFALSTGASGHGYTDLPISRQKLCANGTV
TNCGDIQYEPQSVEGPKGFPASGPADGQLCNGGRFAQLSSPTGGAWPT
TRVTGGQTTFRWQFTAMHATTDFKYYVTKQGWNQNHALARSDLNTT
PFLTVPYNGQRPPATLSHSGTLPSGLSGHHVILAVWTIADTGNAFYACS
DVTF

WP_093654234.1 *Streptomyces wuyuanensis*
MRKRIGAAMVGLAVAGTTLLATGSASSHGYTDSPISRQKLCANGTVGN
CGEIQWEPQSVEGPKGFPGAGPADGKLCAGGRFAQLDDPRGGAWPTT
RVSSGQNTFRWQFTAMHATTDFKYYITRNGWNPSQPLTRAALDPQPFL
TIPYNGQRPPGTLSHSGTLP-SKSGRHLILAVWTIADTPNAFYACSDVQF

WP_026416933.1 *Actinomadura oligospora*
PGTRTFTPRKTTLLLAAVAGLGLPLLSAAPALAHGYTTSPPSRALLCKQ
GVVRNCGPIQWEPQSTEGPKGFPSAGPADGTLCAAGRWAPLDDPRGG
AWPARSLTSGASSFTWTLTAAHATTSFRYFVTRDGWNPAQPLTRAQLD
PTPFLQVDYGGRIPPYSVTHTGTLPV-
KHGRHMILAVWTIADTANAFYQCSDVDFG

WP_017615731.1 *Nocardiopsis salina*
MQNRVLRSSAAVAGSTALIFALMPAGTASAHGYVSDPPSRQAQCAAG
VVECGPIQWEPQSVEGPSGLMSCSGGAFSELDDDSYGWDVTPVGTTSL
EWTITAAHSTAEWEYFIDGDLVDSFDEGGAQPPSNFTHDVDLSGYSGE
QTILARWNVADTDNSFYACIDVNGAL

WP_039991854.1 *Serratia odorifera*
MNATSRTLLSVGLLGAAMFGASHQASAHGYVESPASRSYQCKLQLNT
QCGSVQYEPQSVEGLKGFPQSGPADGHIASADTFFELDQQTPTRWNKL
NLHSGPNSFTWNLTARHSTTSWRYFITRQGWDASQPLTRAAFDLTPFC
QFNDGGAIPAAKVTHQCNIPADRTGSHVILAVWDVADTGNAFYQAIDV
NLSK

WP_044552811.1 *Serratia liquefaciens*
MNNTSRTLMSLGLLSAALFGVSQQAAAHGYVETPASRAYQCKLQLNT
QCGSVQYEPQSVEGLKGFPQAGPADGHIASADTFFELDQQTPTRWNKI
NLHTGANSFTWRLTARHSTTSWRYFITKPNWDASQPLTRASFDLTPFCQ
QNDGGAIPAAQVTHQCNIPADRSGSHVILAVWDVADTTNAFYQAIDVN
LTK

WP_054307289.1 *Serratia rubidaea*
MNTMTRTLMTLGLLSAATLGVSHTASAHGYVESPASRAYQCKQQLNT
QCGSVQYEPQSVEGLKGFPQAGPPDGHIASADTFFELDQQTPTRWNKL
NLHTGENTFTWNLTARHSTTSWRYFITKPGWDASQPLSRASFDLTPFCQ
YDDGGAIPAAKVSHKCTIPADRSGSHVILAVWDVADTGNAFYQAIDVN
LSN

WP_085118064.1 *Serratia proteamaculans*
MNKTSRTLMSLGLLSAALFGVSQQAAAHGYVETPASRAYQCKLQLNT
QCGSVQYEPQSVEGPKDFPQAGPADGHIASADTFFELDQQTPTRWNKI
NLHTGPNSFTWTLTARHSTTSWRYFITRPNWDASQPLTRASFDLTPFCQ
QNDGGATPAAQVTHQCNIPADRSGSHVILAVWDVADTTNAFYQAIDV
NLSK

WP_050539174.1 *Yersinia mollaretii*
MKYIRANILILTGTLTLSSYFSSLAHAHGYIQSPPSRAYLCYQGINHYCG
AVQYEPQSVEGYKNFPYSGPEDGRIASGGPFSALDAQTPTRWQKTNVH

SGTMNLQWQLTATHSTTHWRYYLTKQGWNSSQPLTRASFDLVPFCQY
NDNGRVPNTRVVHQCSLPLDRSGYHVILAAWDIADTNHAFYQVVDISF
PH

WP_051714045.1 *Shewanella*
VPKFAIPKLTQLSLVALALTAGSSLLSQTASAHGYVISPESRSYACKTGS
NVNCGAIQWEPQSVEGPSGFPESGPADGKIASAAAFSPLDEQSPSRWSK
RDIKAGWNDFSWQFTANHVTRNWRYYLTRQGWDQNQALSRASFDLA
PFCVIDGGMVQPPKLVTHNCYVPEDRTGYHVILAVWEVGDTTNSFYNA
IDVNLSSGAVVPGEWTDIGCKPWPY

WP_099457584.1 *Shewanella xiamenensis*
VPKFAIPKLTQLSLVALALTAGSSLLSQTASAHGYVISPESRSYACKTGS
NVNCDAIQWEPQSVEGPSGFPESGPADGKIASAAAFSPLDEQSPSRWSK
RDIKAGWNDFSWQFTANHVTRNWRYYLTRQGWDQNQALSRASFDLA
PFCVIDGGMVQPPKLVTHNCYVPEDRTGYHVILAVWEVGDTTNSFYNA
IDVNLSSGA-VVPGEWTDIGCKPWPY

WP_009837614.1 *Pseudoalteromonas tunicata*
MKKLNKLFLAMAPLATLVSLPFAQTAVAHGYISKPESRGYLCRLTQNT
NCGNVVYEPQSLEGPDRFPESGPADGHIASAGAFSQLNAQTISRWTKRP
IKAGPNEFTWTFTANHSTRDWRYFITKTTWDPNSPLTRDQFEAVPFCEY
SGHYKQPPRQVTHLCNVPADRNGYHIVLGVWDVGDTGMSFYNVVDL
MIDNGDTSNVYWQDVGCKPWPY

WP_082916329.1 *Pseudoalteromonas neustonica*
MNILHKTLLAMSPLALAMLLPVSNTASAHGYISKPESRGYLCRLRENA
NCGNVVYEPQSLEGLDRFPESGPVDGHIASAGGFGQLDSQSINRWTKR
HIKAGPNEFSWTLTAAHSSRDWRYFITKKSWNPNSPLTRDQFEAVPFCE
YQSYNKQPPRQLTHLCNVPSDRSGYHIVLGVWDVADTPMSFYNVVDV
MIDNGDTSNVSWQDVGCKPWPY

WP_072034004.1 *Rahnella aquatilis*
MKKYLMTPVLSLIFSGSLMLGQEASAHGYVSSPASRAYECNLQNNKNC
GPVQYEPQSIEGPGNFPIGGPPDNRIPSGGQYHELDAQTGERWVKNSIKS
GPITFTWTLTAQHSTKSWKYYMTKPGWDINKPLTRDSFNLTPFCEFYD
DGAVPPPVVNHSCNFPSGNIGYNVILAVWNIQDTSHAFYQAIDVDITK

WP_042289422.1 *Citrobacter*
MKLSKIALAVATLTVASSALAHGYIESPASRAYMCKLGQNVDCGSVQY
EPQSVEKTSGFPTGAPPDGQLASAGQYSQLDRQSLNAWTKSPMTAGPH
QFVWHHTAPHKTTNWRYYITKQNWDPNKPLTRDQFDLTPFCTINGNG
QAPAVTQSMNCNVP-ERTGYQVIYGVWEIADTTNSFYQAIDVDFGNSG-
NVTPDETPGVCKPFPY

WP_077227974.1 *Leclercia adecarboxylata*
MKLSKIALALATLTVASSALAHGYVESPASRAYMCKLGKNIDCGTVQY
EPQSVERTSGFPTGAPPDGQLASAGQYSQLDRQSLNAWTKTPITAGKNT
FTWYHTAPHKTVNWRWYITKQDWNPNKPLTRDQFESTPFCTVNGNGQ

APSPRQEMSCNVP-QRTGYQVIYAVWEIADTTNSFYQAIDVDFGNGG-
NVTPDETPAVCKKAPY
WP_045353104.1 *Enterobacter cloacae complex*
MKLSKIALTVATLTVASSALAHGYIESPASRAYMCKLGQNIDCGSVQY
EPQSVEKTSGFPTGAPPDGQLASAGNYSQLDKQSLNTWTKNPMTAGPH
DFVWHHTAPHKTTNWRYYITKQNWDPNKPLTRDQFELTPFCTINGNG
QAPAMTKSMTCNVP-ERTGYQVIYGVWEIADTSNSFYQAIDVDFGNGG-
NVTPDETPAICKPFPY

WP_059288204.1 *Enterobacter hormaechei*
MKLSKIALAVTTLTVASSALAHGYIESPASRAYMCKLGQNIDCGSVQY
EPQSVEKTSGFPTGAPPDGQLASAGSYSQLDKQSLNAWTKNPMTAGPR
EFVWHHTAPHKTTNWRYYITKQNWDPNKPLTRDQFELTPFCTINGNGQ
APAMTKSMTCNVP-ERTGYQVIYGVWEIADTANSFYQAIDVDFGNGG-
NVTPDDTPAVCRPVPF
WP_053064552.1 *Klebsiella oxytoca*
MQLKKLSIYTAALLFTSSALAHGYVAFPPSRAYQCNTGKNSDCGSVQW
EPQSVEQASGFPEGAPPDGQLASAGNFSQLDSQSPTRWAKSAIKSGENN
FIWHHSAPHKTTNWRYYITKQNWDQNKPLTRSDFESKPFCQIDGNGMT
PAIEVTHSCNVP-
ERTGYQVIYAVWEIADTANSFYQAIDVDFGGTGDDAENGSLWTTCRPF
PY
WP_064352001.1 *Klebsiella*
MQLKKLSIYTAALLFTSSALAHGYVAFPPSRAYQCNTGKNSDCGSVQW
EPQSVEQTSGFPEGAPPDGQLASAGNFSQLDSQSPTRWAKSAIKSGENN
FIWHHSAPHRTTNWRYYITKQNWDQNKPLTRSDFESKPFCQIDGNGMT
PAIEVTHSCNVP-
ERTGYQVIYAVWEIADTANSFYQAIDVDFGGTGDDAENGSLWTTCRPF
PY
WP_084204703.1 *Aeromonas popoffii*
MAAKIQLNHIATALALLASGGVLAHGYISQPESRNYLCKTGGNSQCGA
VQWEPQSVEGPSGFPQSGPQDGQIASAGNWSELNAQTSDRWARREVQ
AGPFAISWTFTANHVTRNWRYYLTKQDWNPNQSLTRAAFDLTPFCVV
DGNMVQPPKQVTHRCTLPERTGYQLILGVWEVGDTSNSFYNLIDANFK
GGTQPPLSWSQGGCRPFPS
WP_079516857.1 *Kosakonia oryzae*
MELLMKLSKVFLAVTTLCMAGGALAHGYVTEPASRAALCTANQNKD
CGAPQYEPQSVEGPDGFPAAGPADGHLASAGSMVNLDQQSASRWTKH
PMKAGKNSFTWKFTAAHKTGSWKYYITKANWNPNQPLTRNSFEAQPF
CEVTGNGAIANTNPKHECDVP-
ERDGYQVIMAAWDVSDTGATFYNVIDVDFGGNNPAPAPGDEEENCQK
VSY

WP_062702299.1 *Chryseobacterium indologenes*

MINRKFFFPVLLIMALLVPSFNLSAHGYVVSPASRGYQGSLDKATLYGT
VINEPGSLEALKGFPALGPADGKIASGSGNTLLDIQTADRWKKTNITAG
VNTFIWKYLAYHATAKWHYYMTKPGWNPNKPLTRQDLELIGEVTHNG
TPPQDNVPHHITVPANRTGYHVILAVWDVADTGNAFYNVIDVNVTSGT
GVSTPPAVPTGGTRVFY

WP_084694528.1 *Chryseobacterium vrystaatense*

MIKLRILFSAVLVLCTLIFSSHLSAHGYVMSPASRGYQGSLDKTSLYGSV
INEPGSLEAKKGFPALGPADGMIASGSGNTVLDLQTADRWKKTNITTG
VNTFIWKYLAYHATAKWHYYMTKPGWNPNQPLSRQSLELIGEVIHNG
TPPQDNTPHHITIPANRQGYHIILAVWDVADTANAFYNVIDVNVQSTVT
PTTPAVPTGLGTKVFY

WP_090024402.1 *Chryseobacterium oleae*

MIKLRIFFSALLLCILTFSSKLSAHGYVMSPASRGYQGSLDKATIYGSVIN
EPGSLEAKKGFPAFGPADGMIASGSGNTILDIQTADRWKKTNVTTGVN
TFIWKYLAYHATAKWHYYMTKPGWNPNQPLSRQNLELIGEVIHNGTPP
QDNTPHHITIPANRQGYHVILAVWDVADTANAFYNVIDVNVQSTVTPT
QPATPTGLGTKVFH

WP_082798703.1 *Chryseobacterium cucumeris*

MKTSKFFFALLLMLAMLVPSLHLSAHGYVLSPASRGYQGSLDKAAIYG
SVINEPGSLEAPKGFPAAGPADGKIASGSGDTTLDIQTADRWKKTNISTG
VNAFIWKYSAYHATAKWHYYMTKQGWNPNQPLSRQDLELI--
GTVMHNGTPPQDNVSHQITVPANRTGYHIILAVWDVADTTNAFYNVID
VNVTSGTGVSAPATPTGLGTKVFH

WP_089695525.1 *Chryseobacterium culicis*

MMTRKFFFQVFLMMSILISLIHLSAHGYVVSPASRGYQGSLDKASIYGS
VINEPGSLEAPKGFPASGPADGKIASGSGDTTLDIQTADRWKKTNITTGI
NSFIWKYLAYHATAKWHYYMTKQGWDPNKPITRQDLELGTIAHNGTP
PQDNVSHQITVPANRTGYHIILAVWDVADTTNAFYNVIDVNVTSGTGV
AAPAIPTGLGTKVFH

WP_083997416.1 *Chryseobacterium angstadtii*

SINRSYKKHLKIMIKCRILFSLFLMLLFSSHLFAHGYVLSPASRGYQGSL
DKATLYGSVINEPGSLEAKKGFPAFGPADGRIASGSGNTVLDLQTADR
WKKTNITAGVNTFIWKYSAYHATAKWHYYMTKPGWDPNQPLKRQDL
ELIGEIVHNGTPPQDNVPHQITVPNNRSGYHIILAVWDIADTVNAFYNVI
DVNVQTGGPITAPATPTGLGTKVFH

APPENDIX II

ENDO 1,4 BETA Xln A

<u>P18429</u> *Bacillus subtillis (Query)*
KNFLVGLSAALMSISLFSATASAASTDYWQNWTDGGGIVNAVNGSGG
NYSVNWSNTGNFVVGKGWTTGSPFRTINYNAGVWAPNGNGYLTLYG
WTRSPLIEYYVVDSWGTYRPTGTYKGTVKSDGGTYDIYTTTRYNAPSI
DGDRTTFTQYWSVRQSKRPTGSNATITFSNHVNAWKSHGMNLGSNWA
YQVMATEGYQSSGSSNVTVW

<u>XYL NODE 45</u> *ancestral enzyme*

KKLLAVLLAAAMSLSMFAVTASAAGTDYWQNWTDGGGTVNAVNGS
GGNYSVNWSNTGNFVVGKGWTTGSPNRVINYNAGVFAPSGNGYLTLY
GWTRNPLIEYYVVDSWGTYRPTGTYKGTVNSDGGTYDIYTTTRYNAPS
IDGDTQTFTQYWSVRQSKRPTGSNATITFSNHVNAWASHGMNLGSNW
SYQVLATEGYQSSGSSNVTVW

C-LPMO

<u>AAD27623.1</u> Streptomyces viridosporus (query)
HGVAMVPGSRTYLCQLDAITGTGALNPTNPACRDALNKSGSSALYNW
FAVLDSRAAGRGPGYVPDGTLCSAGDRSPYDFSAYNAARADWPRTHL
TSGATVKVQYSNWAAHPGDFRVYLTKPGWSPTSPLGWNDLELIQTVT
NPPQQGSPGTNGGHYYWDLKLPSGRSGDRLIFMQWVRSDSQENFFSCS
DI
<u>LPMO NODE 50</u> ancestral *enzyme*
HGAMVFPGSRTYLCYQDGVGGGGALQPTNPACAAAVAQGGTQPLYD
WFGVLRSNAAGRHREIIPDGQLCSGGTRGPYDFSAYNLARTDWPTTQL
TAGATIQFRYNNWAPHPGTWYLYVTKDGWDPTRPLKWSDLEPFDTVT
NPPLSNGPGTDDGHYYWDATLPNGKSGRHIIYSIWQRSDSPEAFYSCSD
V

Ch-LPMO

<u>WP_098855344.1</u> *Bacillus thuringiensis (Query)*
MKKNSLQKMKKVILSGGVLLTGLLTFGFSEKASAHGYVESPASRSYLC
KQGVNVNCGPIQYEPQSVEGIGGFPQLGPSDGQIAGAGHFPALDVQTV
DRWKKVTLNGGTNTFKWKLTAPHSTKEWKYYITKKGWNPNKPLTRS
DLDLVPFYVKNDGGARPGTTVTHEANVPTDRSGYHLILAVWEIADTGN
AFYQVIDVNLLNNGLVSNFAFNNVVQTPTLF

<u>LPMO NODE 53</u> *ancestral enzyme*
VMKKMLMKSKKISIAVALLMAGLLTLLFASSASAHGYVSSPASRAYLC
SLGKNTNCGSVQYEPQSVEAPKGFPAAGPADGQIASGGKFSQLDEQTA
NRWTKTPIKAGANTFTWHFTAPHATTKWHYYITKQGWNPNKPLTRAD
LELIPFCEIDDNGATPSTNVTHQVNPSDRSGYHVILAVWDIADTGNAF
YQVIDVNVGGGGATTAPANPTATGDKVTY

List of Abbreviations

AcCNC: Nanocrystals prepared by sulfuric acid treatment

AFM: Atomic force microscopy

Anc C-LPMO: Ancestral cellulose lytic polysaccharide monooxygenase

Anc Ch-LPMO: Ancestral chitin lytic polysaccharide monooxygenase

Anc EG: Ancestral endoglucanase

Anc EG+CBM: Ancestral endoglucanase with carbohydrate binding module attached

Anc LPMO: Ancestral lytic polysaccharide monooxygenase

ANC XLN: Ancestral xylanase

B. subtilis: Bacillus subtilis

BC: Bacterial cellulose

BKP: bleached Kraft pulp

Bt: *Bacillus thuriegiensis*

CBH: Exoglucanase

CBM: Carbohydrate-Binding Module

CFCh: Chitin nanofibers

CI%: Crystallinity index

CNC: Cellulose nanocrystals

CNCh: Chitin nanocrystals

CNF: Cellulose nanofibers

CP/MAS ^{13}C NMR: Cross-polarization magic angle spinning ^{13}C Nuclear Magnetic Resonance

DNA: Deoxyribonucleic acid

DNS: 3,5-dinitrosalicylic acid

DTG: Derivative of thermogravimetric curves

E. coli: Escherichia coli

EG: Endoglucanase

En CNCh: Enzymatic produces nanochitin

EnCNC: Enzymatic produced nanocrystals

FTIR: Fourier-transform infrared spectroscopy

IPTG: Isopropyl β-D-1-thiogalactopyranoside

IR: infrared

LPMO: Lytic polysaccharides monooxygenases

MW: molecular weight

OD: Optical density

Q: Query

S. viridosporus: Streptomyces viridosporus

TEA: Triethylamine

TEMPO: 2,2,6,6-tetramethylpiperidine-1-oxyl

TGA: Thermogravimetric analysis

UBKP: Unbleached Kraft pulp

XLN: xylanase

XRD: X-ray diffraction.

Acknowledgements

I would like to finish this thesis mentioning all the people who have made it possible.

First of all, thank you Prof. Txema Pitarke for giving me the opportunity to accomplish my master and PhD thesis in nanoGUNE, developing my PhD in such a great place would not have been possible without your support.

I am especially grateful to my supervisor Raúl Pérez Jiménez for giving me the chance to join nanobiomechanics group and learn cutting edge science. Thank you for your support, confidence, advices and your sense of humor. Thanks also to my co-director David de Sancho, without your collaboration I would not have finished my thesis. Thank you.

My thanks to all the members (former and current) of nanobiomechanics group. Thank you Aitor for showing me ASR and Marie Fertin for perfoming my first experiments with me. I would like to specially thank Alvaro and Jorg for their unconditional support from the bigining, nunca olvidaré aquellas tardes que terminaban con una caña en mano. My most special thanks to Leire, eskerrikasko emandako guztiagatik. But, my hurt is with Borja and Antonio; Borja, contigo he aprendido lo más bonito de la ciencia, no ponerle limites a la imaginación y el trabajo en equipo, sabes que has sido mi mayor pilar, y puedo decir con orgullo que no solo en el trabajo, si no en la vida también. Gracias a ti Antonio, tu luz me ha iluminado en muchas ocasiones. Gracias a los dos, por tanto.

No quiero olvidarme de las maravillosas personas que nanogune me ha dado. Nestor y Javi, por los que empezamos y terminaremos juntos! Nos queda mucho por recorrer. Milesker Andoni, beti prest, laguntzeko edo parrandarako. Teresa, siempre con una sonrisa y buenos consejos. Pero especialmente gracias a ti Sarai; Milesker txiki, nere aizpa bezela izateagatik, zure adiskidetasuna ezinbestekoa izan da neretzako.

Un especial agradecimiento a mis compañeros de Madrid: Berni, Xabi y Patri.vosotros sois el claro ejemplo de como se puede ser competitivo y buena persona al mismo tiempo.

Nanogunetik kanpo dauden pertsona gertukoenak ere oso presente ditut. Maria, zure eskua eman zenidan eta zientziarekiko grina erakutsi zenidalako, ez iezadazu inoiz askatu. Milesker Janire, Helena, Martin eta Mikel izan ginelako, garelako eta izango garelako. Eskerrikasko nere lagunei, zuen terapia orduak ez dute prezriorik! Milesker Eneko, egunik zatarrena alaitzeko dezun abilezeiagatik.

Azkenik, nere familia eskertzea gustatuko litzaidake. Aita eta Ama, eman didazuen guztiagatik, zuek gabe hau ez zen posible izango. Eskerrikasko Amaia, euki daitekeen aizpa hoberena zeralako. Milesker Marta, Luis eta Maier, neri hegoak emateagatik. Baita zuei ere Belen eta Lulu, milesker beti-beti hor zaudetelako. Eta milesker zuri Itzal, tesia iteko zein bizitzeko euki daitekeen bidelagunik onena zea, milesker nere bizitza maitasunez betetzeagatik.

Amaia, Ana, Itzal eta Iñigori. Izateagatik..