PhD thesis:

Study of *Thermotoga maritima* β-galactosidase: immobilization, engineering and phylogenetic analysis

by David Talens-Perales

Supervised by:

Julio Polaina Molina

Julia Marín Navarro

Valencia, 2016

Los Doctores Julio Polaina Molina y Julia Marín Navarro pertenecientes al Instituto de Agroquímica y Tecnología de Alimentos del Consejo Superior de Investigaciones Cientificas, hacen constar que: La Tesis Doctoral titulada "Study of *Thermotoga maritima* β-galactosidase: immobilization, engineering and phylogenetic analysis", presentada por Don David Talens Perales para optar al grado de Doctor en Biotecnología por la Universidad de Valencia, ha sido realizada en el Instituto de Agroquímica y Tecnología de Alimentos (IATA-CSIC) bajo su dirección, y que reúne los requisitos legales establecidos para ser defendida por su autor. Y para que así conste a los efectos oportunos, firman el presente documento en Paterna, a 22 de Julio de 2016.

Julia Marín Navarro

Julio Polaina Molina

This work was developed at the Department of Food Biotechnology in the Agrochemistry and Food Technology Institute (CSIC), Valencia, Spain. This project was carried out within a JAEpredoc program and FPU program sponsored by CSIC and the Ministerio de Educación, Cultura y Deporte respectively. This project was also supported by grants BIO2010-20508-C04-02, BIO2013-48779-C4-3-R from Spain's Secretaría de Estado de Investigación, Desarrollo e Innovación and EU H2020-634486-INMARE from EU Horizon 2020 Program.

Dedicado a

mi familia y a ti,

gracias por no dejarme vencer

Agradecimientos

Cuando uno decide embarcarse en un doctorado no sabe muy bien qué es lo que está haciendo. Simplemente se deja llevar y muchas veces, una vez ya estás subido en ese tren, no te atreves a bajar o buscas las fuerzas necesarias para no caerte por el camino. Tu vida pasa directamente a girar entorno a un librito que contiene una pequeña parte del esfuerzo, en este caso de cinco años, pero que tan solo es la punta del iceberg de lo que has hecho para llegar a tenerlo. Todo, absolutamente todo, gira entorno a la tesis y en ocasiones obligas un poco a los demás a seguir tu ritmo, nuchas veces estresante. Es por ello que las primeras líneas de los agradecimientos quiero dedicarlas a pedir disculpas a todos aquellos a los que haya podido agraviar con mi mal humor, mis horarios estabulados y un largo etc.

Es algo más que establecido que en los agradecimientos se incluya a los directores de tesis, tanto al Dr. Julio Polaina como a la Dra. Julia Marín-Navarro. A Julia agradecerle la paciencia que ha tenido conmigo, sobre todo a la hora de las correciones y también su apoyo, experimental y moral. Siempre dispuesta a echarme un cable y ponerme los pies en el suelo ante diseños experimentales incompletos, artículos ilegibles y en la corrección de la presente tesis. A Julio por supuesto, por su dirección, por sus consejos y correcciones,

tanto en el laboratorio, como también en la vida diaria. Son muchas horas las que uno pasa en el trabajo y durante 5 años he tenido dos padres, uno biológico y uno científico, con el que he compartido muchos momentos, tanto científicos como de mi vida personal. Por supuesto agradecer a toda la gente que ha pasado por el laboratorio durante estos años, Santi, Álvaro, Nicole, Roberto, Mary, Beltrán, Menandro...seguro que me dejo a gente, porque como sabréis muchos de los que estáis oteando esta tesis, un laboratorio es un sitio muy dinámico en cuanto a personal. Por supuesto no me olvido de Lola, que junto con Julio y Julia ha estado acompañándome día tras día durante el doctorado. Agradecer también a toda la gente del IATA que ha hecho posible que hoy esta tesis se defienda: personal de limpieza, administración, informática...todo debe estar perfectamente sincronizado para que las cosas salgan bien. También debo agradecer a Daniel Hudson y a su grupo por acogerme durante los meses de estancia en la Universidad de Tübingen.

No puedo dejar de lado a mis padres y a mi hermana. A mis padres por ayudarme a ser quien soy, para lo bueno y para lo malo. Por haber confiado en mí desde que decidí estudiar Biología, apoyándome en todas mis decisiones, fuesen más o menos acertadas y poniéndome las cosas fáciles aunque eso significase complicar su día a día. También quiero mencionar a mis abuelos, los que están y los que no, porque todos ellos han dado un golpecito a la hora de forjar quien soy. Debo agradecer a mi pareja todo el soporte día tras día, así como a su familia que han estado ayudando en todo cuanto he necesitado. Mucha gente me dejo por el camino, sobre todo amigos, sin embargo los que me conocen saben que en mi cabeza siempre habrá un rinconcito para ellos capaz de despertar una pequeña sonrisa. No tiene sentido alargarse más con los agradecimientos no vaya a ser que sea más larga que la propia tesis.

Como decían mis padres: al final todo llega. Y este, es el final de una etapa.

Summary

β-Galactosidases are biotechnologically relevant enzymes with an important role in the food industry. Their main application is the manufacture of lactose-free dairy products. They are also used to prevent the formation of crystals in refrigerated dairy products, to accelerate cheese ripening and as a tool for the treatment of whey generated in cheese industries (Adam et al., 2004; Husain, 2010). The retaining catalytic mechanism of these enzymes yields a covalent intermediate between the enzyme and the galactosyl group of the substrate. During hydrolysis this galactosyl group is transferred to a water molecule, but lactose can also act as acceptor (transglycosylation), yielding a galactooligosaccharide (GOS). When lactose acts as acceptor of the galactosyl group the enzyme catalyzes the synthesis of galactooligosaccharides (GOS) (Davies et al., 1997). This process is named transglycosylation. Therefore, these enzymes can be used for the synthesis of GOS, with known prebiotic properties. Some of the benefits of GOS ingesta are improving intestinal microbiota (Vulevic et al., 2008), anti-adhesive effect against pathogenic microorganisms in the intestinal tract (Shoaf et al., 2006; Quintero et al., 2011), as well as increased calcium absorption (van den Heuvel et al., 2000; Whisner et al., 2013). Enzymes used in the industry for GOS synthesis are obtained from microorganisms such as *Kluyveromyces lactis* and *Aspergillus oryzae*, producing β-(1,6) GOS type,

whereas those from *Bifidobacterium bifidum* and *Bacillus circulans* are used for β-(1,3) and β-(1,4) GOS production respectively (Rodriguez-Colinas et al., 2011). Some of the β-galactosidases employed in industry (from *Kluyveromyces lactis, Bacillus circulans* and *Bifidobacterium bifidum*) belong to the GH2 family, according to CAZy classification based on sequence similarity (Lombard et al., 2014).

Lactose is poorly soluble in water at room temperature (Gänzle, 2012) but its solubility increases with temperature. Therefore there is interest in characterizing thermostable β-galactosidases. The use of high temperatures in industrial processes, allowed by these type of enzymes, also reduces the risks of contamination by mesophilic organisms.

Immobilization of enzymes provides certain advantages, facilitating their re-utilization and yielding a product free of enzyme (without additive). Moreover, in some cases, immobilization stabilizes enzymes against external agents such as temperature, pH or the presence of proteases, and can even improve their kinetic properties (Torres-Salas et al., 2011).

In the present study we used the β-galactosidase of hyperthermophilic bacteria *Thermotoga maritima* (TmLac) classified in the GH2 family, as study material, the results obtained for each of the specific objectives raised are summarized below:

Immobilization of TmLac to facilitate its use in industrial processes

In this study we have used two different immobilization methodologies. On one hand, covalent immobilization was performed on polyvinyl alcohol magnetic beads activated with epoxy groups. On the other hand non-covalent

immobilization based on tagging with carbohydrate binding domains (CBMs) was tested. CBMs facilitate the substrate recognition (usually polysaccharides) by glycosyl hydrolases (Boraston et al., 2004). Immobilization of TmLac using CBMs from four different families (CBM2, CBM5, CBM9 and CBM19) also aims to characterize the binding specificity of these domains to two of the most abundant polysaccharides in nature, cellulose and chitin. CBM5 and CBM2 were obtained from *Pyrococcus furiosus* chitinase A and B, respectively. CBM9 was obtained from *Thermotoga maritima* xylanase A and CBM19 from *Saccharomyces cerevisiae* chitinase Cts1. Parameters such as support binding feasibility, enzyme activity of the bound enzyme and recyclability were analyzed for both types of immobilization. For CBM mediated immobilization the reversibility and the specificity of interaction has also been studied.

Covalent immobilization efficiency was 20 mg per g of support. Activity assays showed an increase of the total activity of the bound enzyme compared with the free version, using both the chromogenic *p*-nitrophenyl-β-D-galacto pyranoside (*p*NP-β-Gal) or lactose at different concentrations as substrate. Furthermore immobilization increased enzyme stability at low temperatures. This could be explained by an increased rigidity of the enzyme by multipoint covalent attachment to the support, favoring a more stable conformation (Mateo et al., 2000; Cowan and Fernandez-Lafuente, 2011).

During analysis of TmLac hybrids with different CBMs, TmLac_CBM5 and Tmlac_CBM19 were discarded, since the former displayed only unspecific binding whereas the latter lost thermostability. Hybrids with CBM9 and CBM2 were bound specifically to cellulose at both pH 6.5 and pH 8.5 and, with higher efficiency (about two fold) at the more acidic pH (8 mg enzyme per g of support). CBM2 hybrid also bound efficiently chitin at pH 8.5 (8 mg of enzyme per g of support). However TmLac_CBM9 binding was unstable in

the presence of glucose and galactose, both hydrolysis products of lactose, and therefore the hybrid TmLac_CBM2 was finally selected.

Hydrolysis of 5 % lactose, concentration equivalent that found in milk, was analyzed. The covalently immobilized enzyme could be reused 4 times, achieving a complete hydrolysis after each incubation period (3 hours) at 75°C. In the case of the immobilized enzyme tagged with CBM2, the activity decreased up to 30 % after the first cycle of incubation (3.5 hours) at 75°C. This drop in activity may be explained by the release of the hybrid from the support or the progressive loss of enzyme activity.

Analysis of the structural determinants of transglycosylation

In collaboration with Dr. Julia Sanz Aparicio (IFQR-CSIC, Madrid) we are trying to solve the three-dimensional structure of the protein by X-ray crystallography. Results obtained so far at low resolution (3 Å) have revealed that the quaternary structure of TmLac is an octamer.

Transglycosylation experiments with lactose at a high concentration (25 %) determined that TmLac synthesizes β-3'-galactosyl-lactose and β-6'-galactosyl-lactose in a 3:1 molar ratio. Based on these results putative key residues for transglycosylation were analyzed. For this, a structural model of the enzyme was developed using as template the solved structure of β-galactosidase (LacZ) from *E. coli* (Juers et al., 2012). Using Autodock4 a molecular docking of the major synthesized GOS was carried out, assuming that the structure of this complex would be similar to the reaction intermediate in the transglycosylation reaction. Based on this study 12 mutations were designed at 6 different positions (V93, V94, W959, D568, F571 and N574) to determine the role of these residues

in the transglycosylation process.

Mutant analysis suggested that W959 is critical for the synthesis of β-3'-galactosyl-lactose. The W959A and W959C mutants showed a reduction of 80 % in the synthesis of the trisaccharide. The results suggest that this residue may be a binding platform of the lactose acting as acceptor of the galactosyl group, through π-C-H interactions. On the other hand, N574S, N574A substitutions increased the synthesis of β-3'-galactosyl-lactose between 30 and 40 % while β-6'-galactosyl-lactose synthesis was not affected. This effect could be attributed to the disruption of a putative hydrogen bond between the residue N574 and the galactosyl moiety to be transferred, giving a more flexible conformation that would facilitate the transglycosylation process . The F571L mutation specifically increased β-3'-galactosyl-lactose synthesis up to 40 %. This mutation could favour the rotation of the W959 residue to a more favorable position for the transglycosylation reaction, by the release of steric impediments. The combinations of mutations N574S/F571L and N574A/F571L increased two-fold the synthesis of β-3'-galactosyl-lactose specifically.

Phylogenetic analysis of GH2 β-glycosidases

Most of β-galactosidases used in industry belong to the family GH2. All of them have in common the catalytic domain GH2C, but show a great variability in the presence and arrangement of other non-catalytic domains. To understand the different domain architectures (DAs) present in the family 2, their phylogenetic relationships and possible connections between structure and function, different bioinformatic methodologies have been applied.

From the proteins classified in family GH2 in CAZy database, only those that contained the canonical catalytic domain (GH2C) defined in Pfam, were

selected. All these sequences contain two β-sandwich domains at the N-terminal end identified as GH2N and GH2. This composition was described by Juers et al., (1999) as a characteristic feature of the family GH2. Analyzing the variable C-terminal end of these proteins, 5 DA types were distinguished. One of the most represented corresponds to the GH2N+GH2+GH2C composition (Type 1), without any additional domain, which includes proteins classified as glucuronidases. A variant of this DA (Type 2) consists of a large subunit (LacL), homologous to DA type 1, and a small subunit (LacM) labeled as Bgal_Small_N in Pfam. The active enzyme is the result of the interaction of LacL/LacM subunits. The most common DA is the type 3, wherein the domain GH2C is joined to Bgal_Small_N through a β-sandwich intermediate domain. C-terminal ends of type 4 enzymes are not structurally identified, while the C-terminal region of type 5 has a large variability of domains identified (BIG1, F5/F8, BIG4, etc.) connected to GH2C through a common β-sandwich domain. The β-galactosidases most widely studied and used in industry belong to types 2, 3 and 5.

Phylogenetic analysis with GH2C sequences shows a close relationship between the evolution of the catalytic domain and the presence of certain C-terminal domains (specifically, Bgal_Small_N and BIG1). On the other hand, the interpretation of the phylogenetic tree allowed us to propose an evolutionary model for the enzymes of the GH2 family. One of the most significant results obtained from this analysis is that the type 2 β-galactosidases (composed of two types of subunits) would be the result of a gene disruption of type 3 after which the beta β-sandwich, acting as a hinge between GH2C and Bgal_Small_N would be lost. This study also shows that although some type 4 proteins, with uncharacterized C-terminal ends, appear close to type 1, 3 or 5, others have GH2C sequences unrelated to any of the enzymes characterized so far. The

structural and functional characterization of this group could provide novel useful enzymes for industry.

Resumen

Las β-galactosidasas son enzimas biotecnológicamente relevantes por su papel en la industria de alimentos. Su principal aplicación es la fabricación de productos lácteos sin lactosa. Se utilizan además para prevenir la formación de cristales en productos refrigerados derivados de la leche, acelerar la maduración del queso y como herramienta para tratar el suero generado en las industrias queseras (Adam et al., 2004; Husain, 2010). En el mecanismo catalítico de retención de estas enzimas se forma un intermediario covalente entre la enzima y un grupo galactosilo proviniente del sustrato. En la hidrólisis este grupo galactosilo es transferido a una molécula de agua, pero la lactosa puede actuar también como aceptor (transglicosilación), sintetizándose un galactooligosacárido (GOS) (Davies et al., 1997). Estas enzimas pueden ser utilizadas para la síntesis de GOS, cuyas propiedades prebióticas son conocidas. Algunos de los efectos beneficiosos de los GOS son la mejora de la microbiota intestinal (Vulevic et al., 2008), efectos antiadhesivos frente a microorganismos patógenos en el tracto intestinal (Shoaf et al., 2006; Quintero et al., 2011), así como el incremento en la absorción de calcio (van den Heuvel et al., 2000; Whisner et al., 2013). Las enzimas utilizadas en la síntesis de GOS a nivel industrial se obtienen de microorganismos como *Kluyveromyces lactis* y *Aspergillus oryzae* que producen GOS del tipo β-(1,6) mientras que las

de *Bifidobacterium bifidum* y *Bacillus circulans* producen GOS β-(1,3) y β-(1,4) respectivamente (Rodriguez-Colinas et al., 2011). Algunas de las β-galactosidasas empleadas en la industria (de *Kluyveromyces lactis, Bacillus circulans* y *Bifidobacterium bifidum*) pertenecen a la familia GH2, según la clasificación del CAZy (Lombard et al., 2014) basada en similitud de secuencia.

La lactosa es poco soluble en agua a temperatura ambiente (Gänzle, 2012) pero su solubilidad aumenta con la temperatura. Es por ello que hay interés en caracterizar β-galactosidasas termoestables. El uso en la industria de procesos a alta temperatura, permitidos por este tipo de enzimas, reduce además los riesgos de contaminación por mesófilos.

La inmovilización de enzimas aporta ciertas ventajas, por ejemplo facilitando su reutilización y dando lugar a un producto libre de la misma (sin aditivo). Además, en ocasiones la inmovilización estabiliza la enzima frente a agentes externos como la temperatura, el pH o la presencia de proteasas, e incluso puede mejorar sus propiedades cinéticas (Torres-Salas et al., 2011).

En el presente estudio se ha utilizado la β-galactosidasa de la bacteria hipertermófila *Thermotoga maritima* (TmLac) clasificada en la familia GH2 como material de estudio. Los resultados obtenidos para cada uno de los objetivos planteados se muestran a continuación:

Inmovilización de la TmLac para facilitar su uso en procesos industriales

En este estudio hemos empleado dos metodologías de inmovilización distintas. Por un lado, inmovilización covalente sobre esferas magnéticas de polyvinil-alcohol activadas con grupos epoxy, y por otro inmovilización no covalente basada en el etiquetado con dominios de unión a carbohidratos (CBMs). Los

CBMs facilitan el reconocimiento del sustrato (normalmente polisacáridos) por parte de las glicósido hidrolasas (Boraston et al., 2004). La inmovilización de TmLac usando CBMs de cuatro familias distintas (CBM2, CBM5, CBM9 y CBM19) tiene además el objetivo de caracterizar la especificidad de unión de estos dominios a dos de los polisacáridos más abundantes en la naturaleza, la celulosa y la quitina. CBM5 y CBM2 se obtuvieron de las quitinasas A y B de *Pyrococcus furiosus*, respectivamente. CBM9 proviene de la xilanasa A de *Thermotoga maritima* y CBM19 de la quitinasa Cts1 de *Saccharomyces cerevisiae*. Para los dos tipos de inmovilización estudiados se han analizado distintos parámetros como son la capacidad de unión al soporte, la actividad de la enzima unida y su reciclabilidad. En el caso de la inmovilización por CBMs se ha estudiado también la reversibilidad y la especificidad de la interacción.

La eficiencia de la inmovilización covalente fue de 20 mg por g de soporte. Los análisis de actividad mostraron un incremento de la actividad total de la enzima unida respecto a la libre, usando tanto el sustrato cromogénico *p*-nitrofenil-β-D-galacto-piranósido (*p*NP-β-Gal) como lactosa a distintas concentraciones. Además la inmovilización incrementó la estabilidad a bajas temperaturas. Esto podría explicarse por una mayor rigidez de la enzima al establecer enlaces covalentes con el soporte en distintos puntos de la estructura favoreciendo una conformación más estable (Mateo et al., 2000; Cowan and Fernandez-Lafuente, 2011).

Durante el análisis de los híbridos de TmLac con los distintos CBMs, TmLac_CBM5 y Tmlac_CBM19 se descartaron dado que el primero mostró solo unión inespecífica mientras que el segundo perdió la termoestabilidad. Los híbridos con CBM9 y CBM2 se unieron de forma específica a celulosa tanto a pH 6.5 como a pH 8.5, aunque con mayor eficiencia (alrededor del doble) al pH más ácido (8 mg de enzima por gramo de soporte). En el híbrido con CBM2

también se unió de forma eficiente a quitina a pH 8.5 (8 mg de enzima por g de soporte). No obstante la unión de TmLac_CBM9 se mostró inestable en presencia de glucosa y galactosa, productos de hidrólisis de la lactosa, por lo que finalmente se seleccionó el híbrido TmLac_CBM2.

Se analizó la hidrólisis de lactosa al 5 %, concentración equivalente a la de la leche. La enzima inmovilizada de forma covalente se pudo reutilizar 4 veces, consiguiéndose una hidrólisis completa tras cada periodo de incubación (3 horas) a 75ºC. En el caso de la enzima inmovilizada a través de CBM2, la actividad disminuyó un 30 % tras el primer ciclo de incubación (3.5 horas) a 75ºC . Esta caída de actividad podría deberese a la liberación del soporte o la pérdida progresiva de actividad del híbrido.

La inmovilización covalente es irreversible pero requiere la purificación previa de la enzima, mientras que la inmovilización a través de CBMs aporta la ventaja de la especificidad de la interacción, que puede permitir acoplar purificación e inmovilización en un solo paso. Además el soporte en este último caso es más económico que una resina activada.

Análisis de los determinantes estructurales de la transglicosilación

En colaboración con la Dra. Julia Sanz Aparicio (IFQR-CSIC, Madrid) se está intentando resolver la estructura tridimensional de la proteína mediante cristalografía de difracción de rayos X. Los resultados a baja resolución (3 Å) obtenidos hasta el momento han desvelado que la estructura cuaternaria de la TmLac es octamérica.

Los experimentos de transglicosilación con lactosa a elevada concentración (25 %) determinaron que la TmLac sintetiza β-3'-galactosil-lactosa y β-6'-

galactosyl lactosa con una relación molar 3:1. En base a estos resultados se analizó qué residuos podrían ser claves para el proceso de transglicosilación. Para ello, se construyó un modelo estructural de la enzima tomando como molde la estructura resuelta de la β-galactosidasa (LacZ) de *E. coli* (Juers et al., 2012). Mediante Autodock4 se realizó un acoplamiento molecular (*docking*) del GOS sintetizado de forma mayoritaria, asumiendo que la disposición de dicha molécula en el centro activo sería similar al intermediario de reacción en el proceso de transglicosilación. En base a este estudio se diseñaron 12 mutaciones en 6 posiciones distintas (V93, V94, W959, D568, F571 y N574), para averiguar el papel de estos residuos en el proceso de transglicosilación.

El análisis de los mutantes sugirió que el residuo W959 es crítico tanto para la hidrólisis como para la síntesis de β-3'galactosil-lactosa. Los mutantes W959A y W959C mostraron una reducción del 80 % en la síntesis del trisacárido. Los resultados sugieren la participación de este residuo como plataforma de unión a la lactosa aceptora del grupo galactosilo mediante interacciones de tipo π-C-H. Por otro lado, las mutaciones N574S, N574A incrementaron la síntesis del β-3'-galactosil-lactosa entre un 30 y un 40 %, mientras que la síntesis de β-6'-galactosil-lactosa no quedó afectada. Este efecto podría atribuirse a la ruptura de un potencial puente de hidrógeno entre el residuo N574 y el grupo galactosilo que ha de transferirse, dando una conformación más flexible que facilitaría el proceso de transglicosilación. La mutación F571L incrementó específicamente la síntesis de β-3'-galactosil-lactosa en un 40 %. Esta mutación podría favorecer la rotación del W959 a una posición más favorable para la transglicosilación, por la supresión de impedimentos estéricos. La combinación de mutaciones N574S/F571L y N574A/F571L duplicó específicamente la producción de β-3'-galactosil-lactosa.

Análisis filogenético de las β-glicosidasas de la familia GH2

La mayoría de las β-galactosidasas utilizadas en la industria pertenecen a la familia GH2. Todas ellas tienen en común el dominio catalítico GH2C, pero muestran una gran variabilidad en cuanto a la presencia y disposición de otros dominios no catalíticos. Para entender las diferentes arquitecturas de dominios (DAs) presentes en la familia 2, así como sus relaciones filogenéticas y posibles conexiones estructura-función, se han aplicado distintas metodologías bioinformáticas.

De las proteínas clasificadas en la familia GH2 en la base de datos CAZy se seleccionaron aquellas que contienen el dominio catalítico canónico (GH2C) definido en la base de datos Pfam. Todas las secuencias seleccionadas contienen dos dominios de tipo β-sandwich en posición N-terminal identificados como GH2N y GH2, composición descrita por Juers et al., (1999) como característica de la familia GH2. Analizando la parte variable C-terminal de estas proteínas distinguimos 5 tipos de arquitecturas de dominios. Uno de los más representados se corresponde con la composición GH2N+GH2+GH2C (Tipo 1), sin ningún otro dominio adicional, que incluye proteínas catalogadas como glucuronidasas. Una variante de esta DA (Tipo 2) está formada por una subunidad grande (LacL), homóloga a la DA de tipo 1, y una subunidad pequeña (LacM) etiquetada como Bgal_Small_N en Pfam. La enzima activa es el resultado de la interacción de las subunidades LacL/LacM. La DA más frecuente es la de tipo 3, en la que el dominio GH2C está unido al dominio Bgal_Small_N a través de un dominio intermedio β-sandwich. Las enzimas de tipo 4 tienen extremos C-terminal que no están identificados estructuralmente, mientras que la región C-terminal de las de tipo 5 presenta

una gran variabilidad de dominios identificados (BIG1, F5/F8, BIG4, etc.) conectados al GH2C a través de un dominio β-sandwich común. Las β-galactosidasas más ampliamente estudiadas y utilizadas a nivel industrial pertenecen a los tipos 2, 3 y 5.

El análisis filogenético realizado con las secuencias GH2C muestra una estrecha relación entre la evolución del dominio catalítico y la presencia de determinados dominios en C-terminal (en concreto, Bgal_Small_N y BIG1). Por otro lado, la interpretación del árbol filogenético nos ha permitido proponer un modelo evolutivo para las enzimas de la familia GH2. Uno de los resultados más significativos obtenidos a partir de este análisis es que las β-galactosidasas de tipo 2 (compuestas por dos tipos de subunidades), serían el resultado de una disrupción génica de las de tipo 3 tras la cual se habría perdido el dominio β-sandwich que actúa como bisagra entre GH2C y Bgal_Small_N. Este estudio también muestra que aunque algunas proteínas de tipo 4, con extremos C-terminal no caracterizados, aparecen próximas a las de tipo 1, 3 o 5, otras presentan secuencias GH2C no relacionadas con ninguna de las caracterizadas hasta la fecha. La caracterización estructural y funcional de este grupo podría aportar nuevas enzimas útiles para la industria.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Die Lactase, ein neues Enzym.

Von

M. W. Beyerinck

in

Delft.

Mit 2 Figuren.

Figure 1.1: Title of the first scientific article where a β-galactosidase was mentioned (Beijerinck, 1889).

1.1 Role of β-galactosidases in food industry

The enzyme β-galactosidase (lactase) was reported firstly by Beijerinck at the Department of Microbiology and Enzymology at Delft University in 1889 (Figure 1.1). The enzyme, secreted by yeasts, was able to hydrolyze lactose into glucose and galactose. β-Galactosidases are widely spread in microorganisms (bacteria, archea, fungi and yeasts), plants and animals. Besides their predominant hydrolytic activity, β-galactosidases also catalyze transglycosylation to synthesize galactooligosaccharides and other derivatives (Park and Oh, 2010). Currently, microbial β-galactosidases are essential enzymes for different food applications related to lactose processing. Two main applications involving these enzymes are the development of lactose-free milk products for consumers with intolerance (Adam et al., 2004) and the synthesis of prebiotic galactooligosaccharides (Gänzle, 2012). Other uses are related to the improvement of food properties. Lactose hydrolysis increases the sweetness

of products without extra sugar adition, reducing calories. Moreover lactose hydrolysis reduce precipitation problems in products such as condensed milk or ice-cream (Adam et al., 2004; Panesar et al., 2006). Finally, β-galactosidases can be used for the transformation of lactose from whey generated in cheese industry. High lactose content is associated with an increased biochemical and chemical oxygen demand in environment. The hydrolyzed waste may be used as substrate for other industrial applications such as bioethanol production (Guimarães et al., 2010)

1.2 Lactose intolerance

Lactose is the most abundant sugar in milk (5% w/v) (Rosado 1997; Kunz et al., 2000), and represents the major carbon source for humans in the earliest stages of life (Adam et al., 2004). Lactose can not be absorbed directly in the gut and needs to be hydrolyzed previously by a β-galactosidase. In mammals this enzyme is the lactase-phlorizin hydrolase (LPH). The enzyme is a major glycoprotein of the microvillus membrane in the human intestinal mucosa. It is synthesized as a single chain precursor with a subsequent intracellular proteolytic processing (Sterchi et al., 1990). During the first stages of development lactase activity is maximum. After a period between 2 and 12 years two phenotypes are differentiated: lactase non-persistant (LNP), which can not tolerate lactose ingestion and lactase persistent (LP), which can consume lactose without problems (Mattar et al., 2012). These phenotypes are directly related to the expression levels of the LPH gene which is dramatically reduced in LNP adults (Wang et al., 1998). Lactase activity reduction is the main cause of lactose maldigestion, and is well distinguished from the congenital lactose intolerance which is a rare disease caused by an autosomic recesive

mutation that affects infants since their birth (Mattar et al., 2012).

The mechanism of the variation of lactase expression in adults is not understood. However, there are some polymorphisms in enhancer regions upstream of the gene that encodes LPH which are more common among LP subjects (Mattar et al., 2012). In a lactose tolerant person lactose is hydrolyzed by the lactase located in the brushes of the gut enterocytes (Figure 1.2 A), and glucose and galactose are absorbed by cells through a sodium cotransporter. This cation transport creates an osmotic gradient that promotes the absorption of water molecules, reducing the luminal water. In a LNP phenotype lactose is not digested and it is accumulated in the colon producing adverse physiologic effects (Figure 1.2 B). High amounts of lactose in the colon set an osmotic gradient which yields an influx of water into the lumen, producing diarrhoea. Moreover, lactose is hydrolyzed by β-galactosidases from the intestinal microbiota, yielding gases that cause blowing, intestinal distension, abdominal pain, etc (Lomer et al., 2008; Ingram et al., 2009). The LP phenotype is very common in north-western Europe and is associated to herding practices (Figure 1.3). However LNP is the major phenotype in the world, representing ca. 65% of the global population (Ingram et al., 2009).

1.3 Oligosaccharides and prebiotic properties

Oligosaccharides are low molecular weight carbohydrates with a polymerization degree between 2 and 10 connected through linear or ramified glycoside linkages (Weijers et al., 2008). These molecules show a vast structural diversity and have different biological roles in organisms being crucial for development, growth and cell function (Raman et al., 2005). In food technology oligosaccharides have been commercialized since the 80s as low calorie bulking agents, but

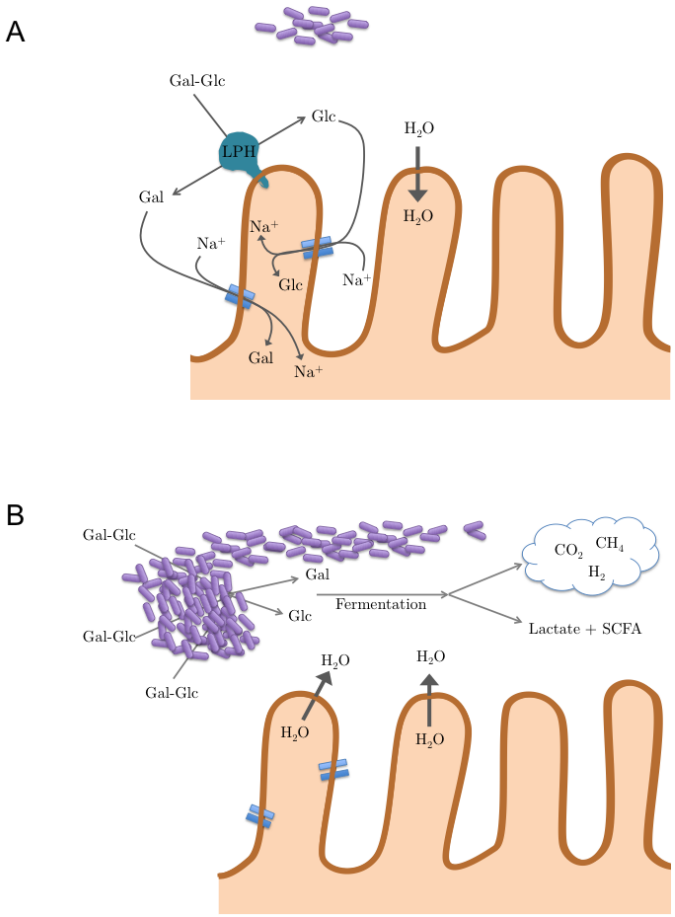


Figure 1.2: Lactose tolerance and lactose intolerance mechanism. (**A**) In a LP phenotype, lactase-phlorizin hydrolase (LPH) in the brush border hydrolyzes lactose (Gal-Glc) into galactose (Gal) and glucose (Glc), which are rapidly absorbed into the bloodstream taking luminal water with them. (**B**) In a LNP phenotype, there are several mechanisms for the symptoms. Unabsorbed lactose passing through the colon generates a high osmotic load producing the increase of water into the lumen. Moreover, lactose is fermented by intestinal microbiota producing gases such as CO_2, CH_4 or H_2. SCFA: short chain fatty acids. Adapted from: (Lomer et al., 2008).

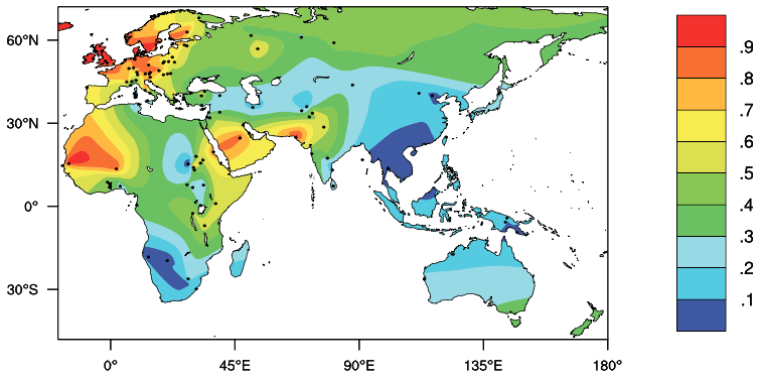


Figure 1.3: Map of Old World lactose persistence (LP) phenotype frequencies. Dots represent collection locations. Colours and colour key show the frequencies of the LP phenotype estimated by surface interpolation. Figure from Itan et al., 2010.

in the last years they have received much interest in food and pharmaceutical industries because of their prebiotic properties and the increased consumer interest in healthy products (Patel and Goyal, 2011). They are also used as jelling agents, antioxidants, humectants and in drug delivery. The concept of prebiotic was proposed by Gibson and Roberfroid in 1995 as "food ingredients that are not hydrolyzed by gut enzymes and have a healthy effect stimulating the growth and/or activity of one or a limited number of bacteria in the colon" (Gibson and Roberfroid, 1995). In 2007 the concept was redefined by Robertfroid as "a selectively fermented ingredient that allows specific changes, both in the composition and/or activity in the gastrointestinal microflora that confers benefits upon host well-being and health" (Roberfroid, 2007).

To consider a food ingredient as prebiotic the following criteria must be accomplished:

1. Resistance to gastric acidity, hydrolysis by mammalian enzymes, and gastrointestinal absortion

2. Fermentation by intestinal microbiota

3. Selective stimulation of growth or activity of microbiota that contribute to health and well-being

Oligosaccharides with prebiotic properties are listed in Table 1.1. Galactooligosaccharides, considered in this study, fulfill the requirements of Roberfroid's definition of a prebiotic. Different health benefits are associated to the ingest of such compounds as will be detailed below. Products with prebiotic effects, such are galactooligosaccharides (GOS) or fructooligosaccharides (FOS), are used as ingredients of infant formulae for preventing infections and to stimulate the immune system. Currently the Directive 2006/141/EC on infant formulae allows the adition of GOS-FOS in ratio 9:1 in a quantity of 0.8 g/100 ml of prepared product (Roberfroid et al., 2010).

1.3.1 Gut microbiota improvement and host health benefits

Gut microbiota has been identified as a key aspect for health and well-being. The microbiota of the human gastro-intestinal tract constitutes a complex ecosystem influenced by different factors. Non-digestible carbohydrates with prebiotic effects stimulate the growth of bacterial genera/species characterized exclusively by saccharolytic fermentation which are likely to be more beneficial to host than those including proteolytic/putrefactive fermentation. The products of carbohydrate fermentation include short chain fatty acids (SCFA), mainly acetate, propionate and butyrate, and other metabolites that act as electron acceptors and gases like H_2, CO_2, CH_4 and H_2S.

Table 1.1: Chemical description of different oligosaccharides, applications and properties. 1: Crittenden and Playne (1996); 2: Hernot et al. (2009); 3: Macfarlane et al. (2008); 4: Moreno et al. (2014); 5: Whisner et al. (2013); 6: Cardelle-Cobas et al. (2011); 7: Hernandez-Hernandez et al. (2012); 8: Prasad et al. (2007); 9: Schuster-Wolff-Bühring et al. (2010); 10: Mu et al. (2013); 11: Closa-Monasterolo et al. (2013); 12: Monsan and Ouarné (2009); 13: Sabater-Molina et al. (2009); 14: Goffin et al. (2011); 15: Bode (2009).

Substrate	Oligosaccharide type	Chemical description	Applications/Properties	References
Lactose	Galactooligosaccharides (GOS)	(D-Galactose)$_n$ monomers connected by β-(1,3), β-(1,6) or β-(1,4) linkages and bond in most cases to a terminal glucose moiety at the reducing end	Functional food (immune stimulation, prebiotic, anti adhesive effect against pathogens and calcium absorption improvement)	1, 2, 3, 4, 5
	Lactulose and lactulose derived GOS (Lu-GOS)	Lactulose: 4-O-β-D-galactopyranosyl-D-fructose Lu-GOS: (D-galactose)$_n$-lactulose	Pharmaceutical (anti constipation agent, treatment for hepatic encephalopathy) Functional food (prebiotic effect)	1, 6, 7, 8, 9
	Lactosucrose	O-β-D-galactopyranosyl-(1-4)-O-α-D-glucopyranosyl-(1-2)-β-D-fructofuranosidase (lactosyl fructoside)	Functional food (prebiotic effect and increased mineral absorption)	1, 10
Inuline/sucrose	Palationose/isomaltulose	[6-O-α-D-Glucopyranosyl-D-fructose]$_n$	Functional food (prebiotic effect) and food additive (substitute sweetener)	1
	Fructooligosaccharides (FOS)	(D-fructose)$_n$ linked,in most cases to a terminal sucrose at the reducing end ^{1}F-FOS: D-fructose units interconnected through β-(2-1) linkages (inuline type) ^{6}F-FOS: D-fructose units interconnected through β-(2-6) linkages (levan type) ^{6}G-FOS: D-fructose units connected to the C6 terminal glucose (neoseries)	Functional food (prebiotic effect, relieve constipation) Industrial uses (processed food adjuvant, cosmetics)	1, 11, 12, 13
Starch	Isomaltooligosaccharides (IMOS)	(D-glucose)n connected by α-(1,6), α-(1,3), α-(1,2) linkages	Functional food (prebiotic and anti-cariogenic effects)	14
Other	Human Milk Oligosaccharides (HMOs)	Oligosaccharides formed by D-glucose, D-galactose, N-acetylglucosamine, L-fucose and sialic acid as basic building blocks	Prebiotic effect, food additive, anti-adhesive effect	15

SCFA accumulation causes a drop in pH from ileum to the caecum which changes microbiota composition, and prevents the overgrowth of pathogenic, acid-sensitive bacteria, like *Enterobacteriaceae* and *Clostridiaceae* genera (Duncan et al., 2009; Cherrington et al., 1991). Moreover, SCFA play a role in the host metabolism of lipids, glucose and cholesterol in various tissues. SCFA reduce cholesterol plasma concentrations in rodents and humans (den Besten et al., 2013). Another health-beneficial effect associated to prebiotics is related to the gut associated lymphoid system which provides an initial defence against intestinal pathogens. The influence of microbiota on the host immune system may proceed through SCFA interaction with the immune cells and entherocytes, modifying their activity (Roberfroid et al., 2010) or by direct interaction with intestinal epithelial cells and dendritic cells (Lebeer et al., 2010). Indeed ingestion of 5.5 g/day of GOS mixture for 10 weeks increased phagocytosis, natural killer cell activity and production of anti-inflammatory cytokines, while the production of pro-inflammatories was reduced (Vulevic et al., 2008).

1.3.2 Anti-adhesive effects

Intestinal tract diseases begin with the recognition by the pathogen, or its toxins, of a cell surface receptor from gut epithelial cells. The anti-adhesive strategy relies on the structural similarity between the carbohydrates of the epithelial cell surface receptors and the prebiotic oligosaccharides (Sinclair et al., 2009). The use of a GOS mixture reduced significantly the adherence to epithelial cells of an *E.coli* entheropathogenic strain (Shoaf et al., 2006), *Cronobacter sakazakii* (an oportunistic pathogen implied in meningitis, necrotizing enterocolitis, and sepsis in neonates) (Quintero et al., 2011) and Ctx toxin from *Vibrio cholerae* (Sinclair et al., 2009).

1.3.3 Other benefits

GOS increase Ca^{2+} absorption in animal models, post-menopausal women and young girls. A collateral effect of the increment of fermentation is the hypertrophy of the intestinal mucosa that increases the surface area for greater mineral diffusion (van den Heuvel et al., 2000; Whishner et al., 2013). Another benefit associated to GOS ingesta is the protective effect against the development of colorectal tumours, preventing the growth of microorganisms that yield aromatic or heterocyclic amines with carcinogenic effects (Bruno-Barcena and Azcarate-Peril, 2015). Finally, the use of GOS has been reported to reduce an inflammatory inter-leukin, suggesting their potential application to treat inflammatory processes and allergies associated with this interleukin (Verheijden et al., 2015).

1.4 Oligosaccharide synthesis

Oligosaccharides can be obtained by two major methods, depolymerization of polysaccharides and synthesis from precursors of lower degree of polymerization. Both processes can be carried out either by chemical or enzymatic technologies (Courtois, 2009). However, chemical synthesis is expensive and reports low yields and therefore, for large scale production, enzymatic methods are the most common. Both glycosyl hydrolases (GHs) and glycosyltransferases (GTs) are used. These enzymes are classified in different families on CAZy database based on sequence similarity (Lombard et al., 2014).

Glycosyl hydrolases (EC 3.2.1) are enzymes capable to hydrolyze glycosidic bonds. They are classified in retaining or inverting depending on whether the reaction mechanism preserves or not the conformation (α or β) of the anomeric

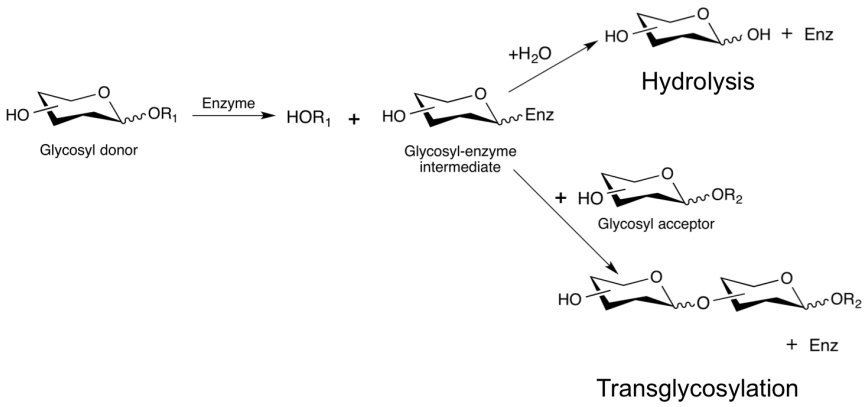


Figure 1.4: Reaction scheme of retaining glycosyl hydrolases. Adapted from Davies et al., 1997.

carbon atom. The inverting mechanism occurs in a single step involving two critical residues: one acts as a general base that deprotonates a water molecule, which carries out the nucleophile attack on the anomeric carbon whereas the other one acts as general acid and protonates the leaving sugar. On the other hand, retaining glycosidases proceed through a double substitution mechanism. In the first step a nucleophilic residue attacks the anomeric carbon of the substrate and the leaving group is protonated by an acid/base catalyst. The result is a covalent intermediate between the enzyme and the glycosyl donor, which is subsequently released by the nucleophilic attack of a water molecule activated by the acid/base catalyst (hydrolysis). Alternatively, a transglycosylation reaction may occur when the covalent intermediate is released by a second sugar molecule acting as glycosyl acceptor (Figure 1.4) (Davies et al., 1997).

Other enzymes able to catalyze transglycosylation reaction require activated donors. This is the case of *Leloir* glycosyltransferases and phosphorylases (EC

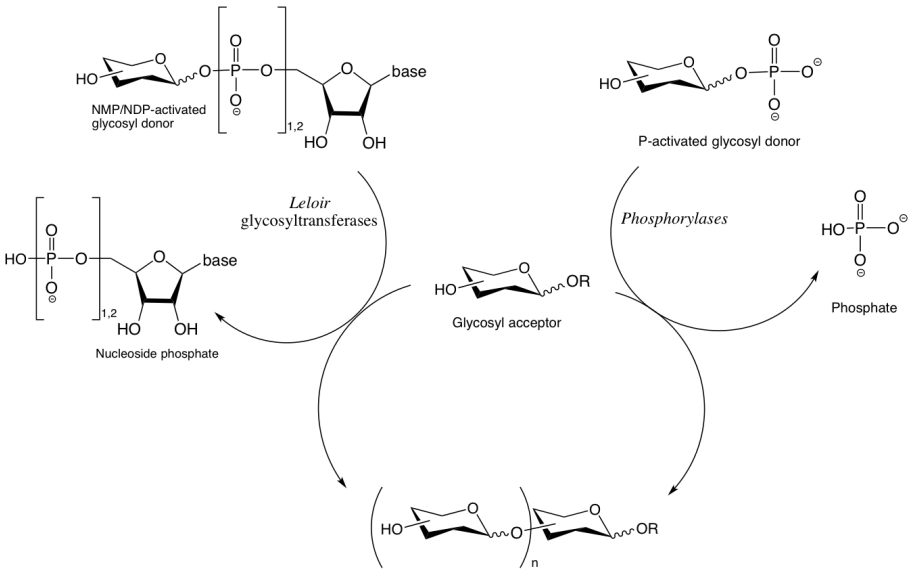


Figure 1.5: Synthesis of oligosaccharides with phosphorylases (right side) and *Leloir* glycosyltransferases (left side). Adapted from Weijers et al., 2008.

2.4) (Figure 1.5). The acceptor substrate may be a monosaccharide but also an heteropolysaccharide, oligosaccharide, nucleic acid or lipid (Lairson et al., 2008; Weijers et al., 2008). Like glycosidases the mechanism can be inverting or retaining. Phosphorylases in their reverse reaction catalyze the transfer from phosphorylated monosaccharides to a carbohydrate acceptor, releasing the phosphate molecule. GTs classified as *Leloir enzymes* use glycosyl donors activated with nucleotide mono- or di- phosphate (CMP, UDP, GDP or GTP), releasing the nucleoside phosphate (Weijers et al., 2008). Despite the diversity of products synthesized by this type of glycosyltransferases the requirement of costly activated substrates, make them less attractive than the enzymes using nonactivated sugars for industrial production and limit their use to high-added-value oligosaccharides.

1.5 Biochemical and structural properties of β-galactosidases

β-Galactosidases are classified by the Enzyme Commission as EC 3.2.1.23/108 and hydrolyze D-galactosyl residues connected through β-glycosidic linkages to polymers, oligosaccharides or secondary metabolites (Adam et al., 2004; Husain, 2010). They are distributed in four glycosyl hydrolase families, namely GH1, GH2, GH35 and GH42 (Lombard et al., 2014). These families use a retaining mechanism for hydrolysis and belong to the clan GH-A characterized by a TIM-barrel domain. Only certain enzymes belonging to families GH1 and GH2 have lactose as the natural substrate, whereas the rest act on different galactose-containing glycosides (Husain, 2010).

Microbial β-galactosidases are preferred for industrial applications due to high efficiency of enzyme production per biomass unit compared to other sources. One of the galactosidases most widely used in the industry for lactose hydrolysis belongs to the GH2 family and is obtained from *Kluyveromyces lactis*. Fungal enzymes are useful to treat substrates such as whey since they have generally an acidic pH optimum range (between 2.5 - 5.4). In contrast, yeast galactosidases have optimum pHs close to neutral (between 6.0 - 7.0). Bacterial β-galactosidases are also widely used for lactose hydrolysis because of their good stability being the acid lactic bacteria one of the major groups of interest (Husain, 2010).

Due to their retaining mechanism, β-galactosidases yield GOS when lactose acts as both galactosyl donor and acceptor. The transglycosylation/hydrolysis ratio depends on the structural conformation of the enzyme active site, but it is also influenced by other factors such as lactose concentration (Bruins et

al., 2003; Ji et al., 2005), water activity and reaction temperature (Millqvist-Fureby et al., 1998). Enzyme specificity determines the linkage type (β-(1,2), β-(1,3), β-(1,4) or β-(1,6)) and the polymerization degree of the synthesized GOS. The most common products are the trisaccharides β-6'-galactosyl-lactose, β-3'-galactosyl-lactose and β-4'-galactosyl-lactose. GOS structural differences have been related to their stability and prebiotic potential (Torres et al., 2010; Hernandez-Hernandez et al., 2012).

Commercial enzymes for GOS synthesis belong to the GH2 family and are extracted from *Kluyveromyces lactis* and *Aspergillus oryzae*, to obtain β-(1,6) GOS; *Bifidobacterium bifidum*, for β-(1,3) GOS and *Bacillus circulans*, for β-(1,4) GOS (Rodriguez-Colinas et al., 2011). Since high lactose concentration increases GOS yield (Gänzle, 2012) the use of high temperatures is desirable for GOS production to increase the lactose solubility. Moreover working at high temperature reduces the risk of microbial contamination by mesophiles. Therefore, a considerable effort has been taken in order to identify thermostable β-galactosidases. Some examples of thermostable enzymes with tranglyscosylating capacity are GH1 β-galactosidases from *Sulfolobus solfactaricus* and *Pyrococcus furiosus*, that yield β-(1,3), β-(1,6) and β-(1,4)-galactosyl-lactose; GH2 β-galactosidase from *Streptococcus thermophilus* that yields β-(1,6)-galactosyl-lactose and the GH42 β-galactosidase from *Thermus* sp., that yields β-(1,3)-galactosyl-lactose (Greenberg and Mahoney, 1983; Petzelbauer et al., 2000; Akiyama et al., 2001; Ji et al., 2005). We have selected the thermostable β-galactosidase from *Thermotoga maritima* to characterize the GOS chemical profile of the wild-type enzyme and to carry out a study of the structural determinants of transglycosylating activity within the active site, which has not been previously addressed with GH2 enzymes (Section 4.6).

The structure of GH2 β-galactosidases includes a TIM-barrel domain that

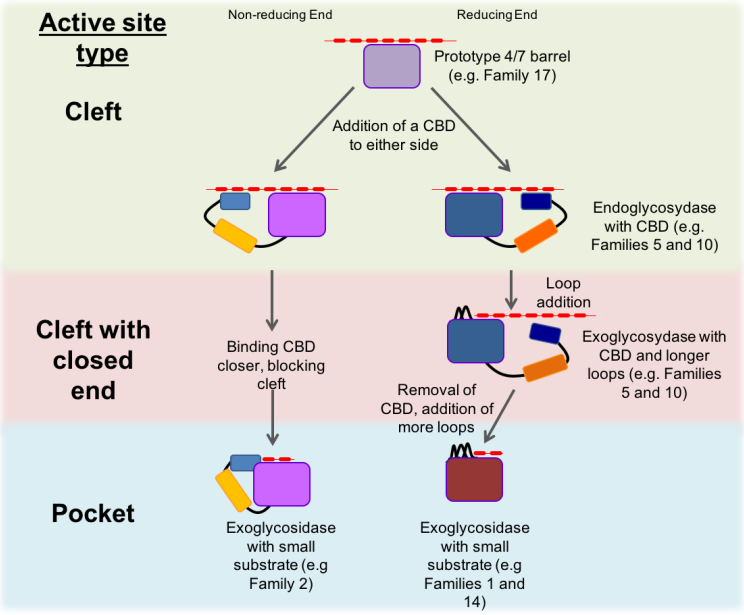


Figure 1.6: Model for the evolution of the 4/7 superfamily of enzymes. A presumed prototypical α/β barrel enzyme with a long, groove-like active site cleft and an extended polysaccharide substrate is shown at the top. The addition of a cellulose binding domain (CBD) to either terminus of the α/β barrel could give rise to both endoglycosidases and exoglycosidases (e.g., families 5 and 10). The generation of an enzyme that hydrolyzes small substrates might have occurred by either of the routes shown utilizing loops and/or a CBD to turn the active site from a cleft into a pocket. Adapted from: Juers et al., 1999.

contains the catalytic residues and two β-sandwich domains at the N-terminal side. Juers et al., 1999 studied the evolutionary relationship between β-galactosidases and other glycohydrolases. Their work suggested the existence of a primitive TIM barrel catalytic domain able to hydrolyze polysaccharides, that after the incorporation of two additional domains in N-terminal position, restricted hydrolysis capacity to small substrates such disaccharides. The primitive cleft conformation of the catalytic site changed to a pocket-shape, more adapted to small substrates (Figure 1.6). Despite this common cassette,

β-galactosidase architecture shows some divergence. Most enzymes, named hereafter canonical β-galactosidases, contain two relatively well conserved β-sandwich domains at the C-terminal end of the catalytic module. In some cases functional β-galactosidases are bicistronic enzymes composed by two subunit types (LacL and LacM). LacL is homologous to the N-terminal domains and the TIM-barrel, whereas LacM is homologous to the C-terminal β-sandwich domain of canonical galactosidases. Interestingly, some β-galactosidases with high tranglycosylating efficiency such as those from *Bacillus circulans* and *Bifidobacterium bifidum*, show a completely divergent modular arrangement at the C-terminal end. In this work, we have carried out a bioinformatic analysis to gain insights into the different architectures of β-galactosidases and the phylogenetic relationships among their catalytic domains (Section 4.7).

1.6 Enzyme immobilization

The use of enzymes is a common and old practice in industry for many different processes (e.g biomass convertion, textile industry, production of biodiesel, antibiotic synthesis, etc) (Polaina and MacCabe, 2007). A more recent application of this methodology is enzyme immobilization, which consists in fixing the enzyme to a solid surface, or to itself creating nets, by different techniques (Husain, 2010; Homaei et al., 2013; Sheldon and van Pelt, 2013). Immobilized enzymes provide different advantages over the corresponding free versions. Immobilization facilitates the separation of the enzyme from the product mixture, allowing its reutilization and the obtention of enzyme-free products. It also allows the implementation of enzymes in biosensors (Ball et al., 2003; Lukacheva et al., 2007; Sezginturk and Dinckaya, 2008; Ammam and Fransaer, 2010). Moreover, in most cases this strategy stabilizes

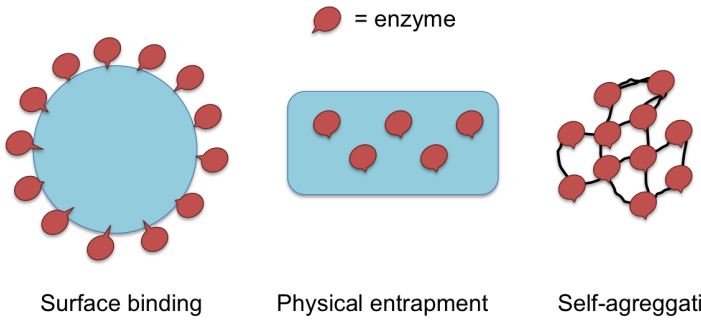


Figure 1.7: Different methods for immobilizing enzymes. Modified from: Sheldon and van Pelt (2013).

the enzyme against external agents such as temperature, pH or proteases, increasing its life-time. Immobilization may also modify the kinetic properties of the enzyme either in a positive or negative way (Torres-Salas et al., 2011). These changes are explained by structural alterations induced by the interaction with the support or by changes in the diffusional properties of the enzyme or the substrates (Homaei et al., 2013).

A universal immobilization protocol for enzymes is not available. Immobilization efficiency depends on the enzyme structure at different levels (from primary to quaternary). *In silico* studies and structure modelling may be used to select an immobilization method as the most adequate for a specific enzyme, but in most cases the best immobilization method for a particular enzyme is determined empirically (Torres-Salas et al., 2011). Enzyme immobilization methods are classified in three main groups: covalent or non-covalent surface binding, physical entrapment and self-aggregation by cross-linking (Figure 1.7) (Mateo et al., 2007b; Torres-Salas et al., 2011; Sheldon and van Pelt, 2013).

- **Surface binding.** Immobilization to the surface of solid supports of different nature can be performed covalently or not covalently. The

fixation can be done using organic polymers, either synthetic (e.g. polyvinyl, polystyrene), or natural (e.g. starch, chitosan and cellulose) or inorganic compounds (e.g. zeolite and silica). Some of these supports are used to cover the surface of a particle or nano-particle with a magnetic core (e.g. magnetite). Proteins immobilized to these magnetic particles can be easily separated by the application of an external magnetic field (Torres-Salas et al., 2011).

Covalent immobilization has as major advantage the high stability of the linkage (virtualy irreversible) between the enzyme and the carrier, preventing the decay of the activity by enzyme releasing from the support. Moreover, covalent binding may occur simultaneously at different points of the same molecule (multipoint attachment) yielding a more stable structure. Organic and inorganic materials can be functionalized by different groups (e.g. amino, carboxyl, epoxide, thiol). Amino and carboxy activated resins require further chemical activation (e.g. with glutaraldehyde or carboxidiimide) to react with either amino or carboxylic groups from proteins. Thiol groups react with cysteine residues but binding is not stable under reducing conditions and in many instances cysteines are not accessible at the protein surface. Epoxy resins react with nucleophilic groups (thiol, amino or phenol). Epoxide activated supports show certain advantages compared to other agents because they do not require a previous chemical activation and the target residues, mainly lysines, are abundant and exposed at the surface of proteins. Chemical reaction between epoxide and amino group needs a previous adsorption of the protein to the support, usually by hydrophobic interactions promoted by the addition of ammonium sulphate at high concentrations. In second generation epoxy-activated resins, this initial adsorption step is promoted

by a secondary derivatitation of the support that promotes electrostatic, complexing or disulfide bond interactions (Mateo et al., 2007a). Reactive lysine residues must be uncharged and therefore highly alkaline conditions are preferred for binding. However this type of covalent immobilization is not useful for highly glycosylated proteins because the reactive groups are not accessible (Torres-Salas et al., 2011). For glycosylated proteins other immobilization methods have been developed, based on the covalent bond formation between polyurethane foams and the hydroxyl groups of carbohydrates (Bakker et al., 2000).

Non-covalent immobilization may be based on specific or unspecific hydrophobic or electrostatic (ionic or non-ionic) interactions. Supports such as membranes or microparticles of nitrocellulose or polylysine or, most recently, carbon nano-tubes have been used for unspecific protein adhesion (Feng and Ji, 2011). The major advantage of this method is that the protein is not chemically modified and that binding conditions are less aggressive than those used for covalent immobilization. However, this kind of immobilization is less stable. Alternatively, non-covalent binding may involve bioaffinity interactions. In this case the binding support may be modified by a ligand (e.g. protein A/G, streptavidine or antibodies) with specific affinity for the protein of interest or a tag fused to it. The main advantage of this approach is that the target protein may be purified and immobilized in one single step (Homaei et al., 2013).

- **Physical entrapment.** It consists in the encapsulation of the enzyme during polymerization of support. The polymeric matrix can be made of organic or inorganic polymer matrices, such as polyacrilamide and silica sol-gel, respectively (Sheldon and van Pelt, 2013). This type of immobilization is the best way to avoid the negative influence of external

agents, such as proteases, which cannot diffuse into the polymeric matrix (Homaei et al., 2013).

- **Self-aggregation.** The method consists in the cross-linking of the enzyme using a bifunctional agent such as glutaraldehyde that yields carrier-less macroparticles. The absence of carrier reduces a large portion of non-catalytic elements in the reaction mixture, allowing better space-time yields and productivity (Sheldon and van Pelt, 2013). Protein matrices can be synthetized as cross linked enzyme crystals (CLECs) or cross-linked enzymes aggregates (CLEAs). Both CLECs and CLEAs can be encapsulated in order to protect them from external agents, resulting in a combined immobilization procedure (Torres-Salas et al., 2011; Sheldon and van Pelt, 2013).

In this work two different types of surface immobilization have been carried out: covalent immobilization using epoxy activated magnetite supports and a non-covalent method based in specific bioaffinity interactions (Sections 4.2 and 4.3). In the latter case the enzyme was linked to carbohydrate binding modules (CBMs), with affinity for polysaccharides, as will be explained in the following section.

1.7 Generation of hybrid enzymes containing Carbohydrate Binding Modules (CBMs)

CBMs act in nature as auxiliary domains of enzymes that hydrolyze polysaccharides, such as cellulose, chitin, starch, glycogen, inulin, pullullan or xylan (Guillén et al., 2010). These domains bind polysaccharides and have three general effects: a proximity effect bringing the catalytic machinery onto

Table 1.2: CBMs classification based on fold (Oliveira et al., 2015).

Fold family	Fold	CBM families (CAZy classification)
1	β-sandwich	2, 3, 4, 6, 9, 11, 15, 16, 17, 20, 21, 22, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 40, 41, 42, 44, 47, 48, 51, 70
2	β-trefoil	13, 42
3	cystein knot	1
4	unique	5, 12
5	OB fold	10
6	hevein fold	18
7	unique (hevein-like fold)	14

the corresponding substrate, a targeting function to specific regions of the polysaccharides, and a disruptive effect loosening the structure of tightly packed polysaccharides (Boraston et al., 2004). In few cases, CBMs are involved in the adhesion of the enzyme to the cell-wall rather than exerting a specific substrate binding role (Montanier et al., 2009). The most common structural element of these domains is a beta-sheet. CBMs are classified in 74 families according to sequence similarity (CAZy database), which are grouped in seven fold families (Table 1.2 and Figure 1.8).

Within CBMs ligand binding sites, aromatic residues play a substantial role interacting with the sugar rings of the polymer by π-C-H stacking contacts. Based on the orientation of these residues CBMs are classified in three main groups. CBMs belonging to type A show two or three aromatic residues exposed, and linearly arranged, at the binding site. They are named surface-binding CBMs and usually interact with crystalline polysaccharides thanks to

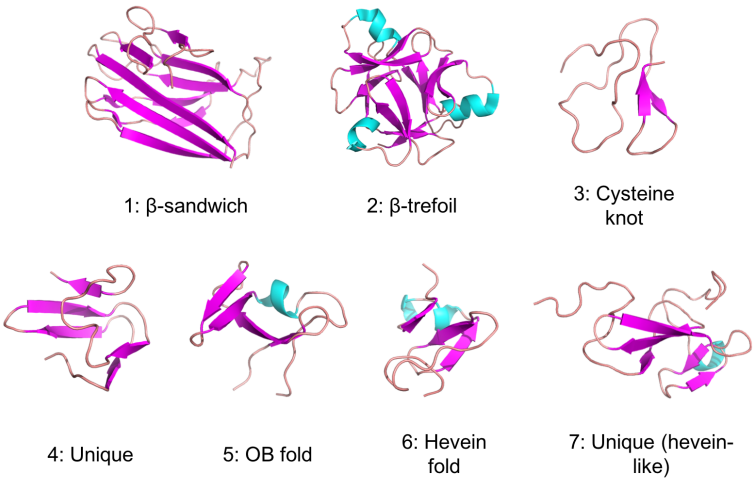


Figure 1.8: CBM structures from the different CBMs folding types. 1: β-sandwich (PDB: 1GU3); 2: β-trefoil (PDB: 3KMV); 3: Cystein knot (PDB: 4BMF); 4: Unique (PDB: 1ED7); 5: OB fold (PDB: 1E8R); 6: Hevein fold (PDB: 1HEV); 7: Unique (hevein-like fold) (PDB: 1DQC).

the planar architecture of the binding site (Figure 1.9 A). CBMs type B bind amorphous polysaccharides through a cleft in the module that restricts binding to free single polysaccharide chains (Figure 1.9 B). Type C or lectin-like CBMs show a pocket-shaped biding site which imposes a restriction to interact only with oligosaccharides or polysaccharides ends (mono-, di- or trisaccharides) (Figure 1.9 C) (Guillén et al., 2010). Despite these structural features are suggestive of certain binding properties of the CBMs, modules within the same type show different substrate specificities.

CBMs have different biotechnological applications by means of the construction of bi-functional proteins with various purposes such as: functionalization of biomaterials for cell immobilization (Wierzba et al., 1995; Andrade et al., 2010), bioremediation tools for the elimination of organo-phosphorous molecules or heavy metals from the environment (Richins et al., 2000; Xu et al., 2002),

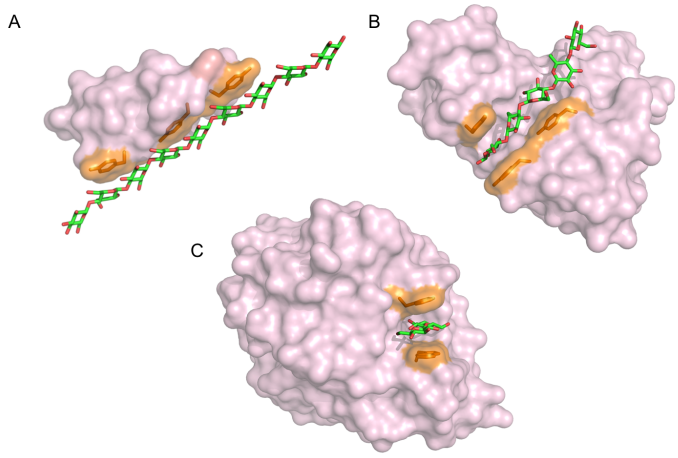


Figure 1.9: Classification of CBMs based on the conformation of the ligand binding site. (**A**) Type A CBM from *Trichoderma reesei* cellobiohydrolase I (PDB 1CBH) (Cellulose was manually docked). (**B**) Type B CBM from *Cellulomonas fimi* endo-1,4-glucanase C (PDB 1GU3) and (**C**) Type C CBM from *Thermotoga maritima* xylanase 10A (PDB 1I82). Carbohydrates are green coloured. The aromatic residues in the binding surfaces are coloured in orange. Adapted from Guillén et al., 2010.

molecular probes to construct a cohesin-dockerin microarray for the study of cellulosomes (Haimovitz et al., 2008), recombinant protein purification, enzyme engineering, etc. (Oliveira et al., 2015). One of the uses of CBMs is the attachment of these domains to enzymes for protein immobilization (Myung et al., 2011). For the current study four CBMs from different families were selected. CBM2 from *Pyrococcus furiosus* chitinase A (ChiA); CBM5 from *P. furiosus* chitinase B (ChiB), CBM9 from *Thermotoga maritima* Xylanase A and CBM19 from *Saccharomyces cerevisiae* Cts1 chitinase. CBM2 and CBM5 show a type A surface interaction with slight differences in the spatial arrangement of the aromatic residues (Nakamura et al., 2008) whereas CBM9 is classified as type C (Notenboom et al., 2001). So far, none of the CBM19 modules has been characterized structurally or biochemically. The purpose of this study was to characterize domains with affinity for cellulose or chitin, the first and the second most abundant polymers on Earth, which may serve as inexpensive immobilization supports for proteins tagged with such modules.

Chapter 2

Objectives

The general objective of this work is the study of a thermostable β-galactosidase from *Thermotoga maritima* (TmLac), belonging to glycosyl hydrolase family 2 (GH2). The specific objectives of the work are:

1. **Immobilization of TmLac to facilitate its use in industrial processes.**

2. **Analysis of the structural determinants of the transglycosylating activity to increase galactooligosaccharide production.**

3. **Phylogenetic analysis of GH2 β-glycosidases.**

Chapter 3

Materials and methods

3.1 Microbial strains and culture media

Escherichia coli strains used in this work are listed in Table 3.1. Cultures were incubated for 16 hours in Luria Bertani broth (0.5% yeast extract, 1% peptone and 0.5% sodium chloride) at 37°C. For overproduction of heterologous proteins, *E. coli* cultures were grown in 2XTY media (1% yeast extract, 1.6% tryptone and 0.5% sodium chloride) at 37°C before induction with IPTG. *Saccharomyces cerevisiae* strain BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) was grown in YPD media (1% yeast extract, 1% peptone and 2% glucose) at 30°C during 16 hours. Liquid cultures were grown with orbital shaking at 200 rpm. Ampiciline and chloramphenicol were added to final concentration of 100 mg/l and 68 mg/l respectively when required. For solid medium 2% agar was added.

3.2 DNA manipulation and analysis

3.2.1 Enzymatic treatment of DNA

Treatment with restriction endonucleases, phosphorylation of the 5'-ends with polynucleotide kinase, and ligation with T4 ligase were performed following the recommendations described in Green et al. (2012).

3.2.2 DNA gel electrophoresis

Gels for DNA analysis were prepared in TBE buffer (89 mM Tris, 89 mM boric acid and 2 mM EDTA) with agarose at 0.7-1% (w/v) of concentration. Preparative electrophoresis in order to purify DNA fragments was carried out

Table 3.1: *E.coli* strains used in the present work. Origin of the microbial strains are defined as: (1) commercial, (2) laboratory collection, (3) this work.

Strain	Genotype /compositions	Origin
XL1-BLUE (Stratagene)	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1Δlac [F proAB lacIqZΔM15Tn10 (Tetr)]	1
ROSETTA2 (Novagen)	F' $^-$ ompT hsdSB (rB$^-$ mB$^-$) gal dcm pRARE2 (Cam^R)	1
XRA	XL1Blue transformed with pRARE2 (CamR) isolated from ROSETTA2	2
TmLac-X	XL1-BLUE transformed with pTmWT	3
TmLac-ROS	ROSETTA2 transformed with pTmWT	3
TmLac-XRA	XRA transformed with pTmWT	3
TmLac_N574A-XRA	XRA transformed with pTm_N574A	3
TmLac_N574S-XRA	XRA transformed with pTm_N574S	3
TmLac_W959F-XRA	XRA transformed with pTm_W959F	3
TmLac_W959C-XRA	XRA transformed with pTm_W959C	3
TmLac_W959A-XRA	XRA transformed with pTm_W959A	3
TmLac_F574L-XRA	XRA transformed with pTm_F571L	3
TmLac_V94Q-XRA	XRA transformed with pTm_V94Q	3
TmLac_V94T-XRA	XRA transformed with pTm_V94T	3
TmLac_V93T-XRA	XRA transformed with pTm_V93T	3
TmLac_F571L/N547S-XRA	XRA transformed with pTm_F571L/N547S	3
TmLac_F571L/N547A-XRA	XRA transformed with pTm_F571L/N547S	3
TmLac_E242A/E243A/E244A-ROS	ROSETTA2 transformed with pTm_E242A/E243A/E244A	3
TmLac_E697A/K698A/K699A-ROS	ROSETTA2 transformed with pTm_E697A/K698A/K699A	3
TmLac_K1029A/E1030A/K1031A-ROS	ROSETTA2 transformed with pTm_K1029A/E1030A/K1031A	3
TmLac_CBM9-ROS	ROSETTA2 transformed with pTm_CBM9	3
TmLac_CBM2-ROS	ROSETTA2 transformed with pTm_CBM2	3
TmLac_CBM5-ROS	ROSETTA2 transformed with pTm_CBM5	3
TmLac_CBM19-ROS	ROSETTA2 transformed with pTm_CBM19	3

in TAE buffer (40 mM Tris, 20 mM acetic acid and EDTA 1 mM pH 8) with low melting temperature agarose (0.7-1% (w/v)). DNA fragments were purified using "QIAquick Gel extraction kit" from Qiagen.

3.2.3 DNA isolation

Plasmidic DNA was purified from 3 ml *E. coli* cultures. DNA purification was carried out by the alkaline lysis method (Bimboim and Doly, 1979), for analytical purposes, or with a commercial kit, for preparative purpose (GeneJet Plasmid Extraction Kit (Thermo Scientific)).

Yeast genomic DNA was obtained from 4 ml yeast culture using a fast DNA isolating method (Polaina and Adam, 1991) with modifications introduced in Latorre-Garcia et al. (2007).

3.2.4 DNA cloning vectors

The vector used for gene cloning and expression was pQE-80L (Qiagen). This vector contains an ampiciline resistance gene as positive selection factor. Gene expression is triggered by a hybrid promoter inducible by galactose or IPTG (isopropil-β-D-galactopiranoside). A histidine tag next to the multiple cloning site (MCS) is fused in frame to the 5' side of the inserted sequence. The presence of the *lacI* repressor ensures tight control at the transcriptional level of the cloned gene (Figure 3.1).

3.2.5 DNA amplification by PCR

DNA amplification was carried out with Phusion ^{TM} polymerase II Hot Start High-Fidelity DNA Polymerase (Thermoscientific) (3-5 U/μl) or Roche

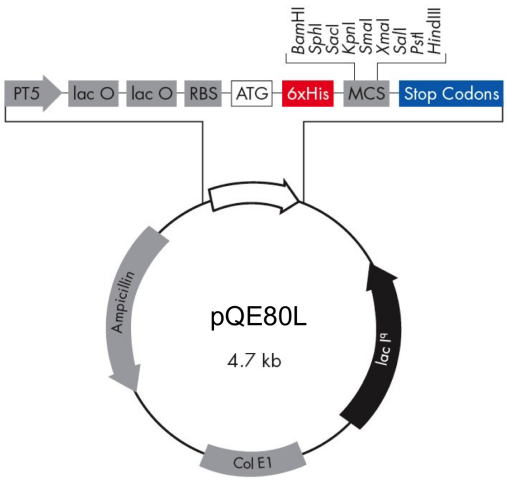


Figure 3.1: PQE-80L vector used for the overexpression of *T. maritima* β-galactosidase gene.

Expand High Fidelity (3 U/μl) according to instructions of the manufacturer. PCR products were purified with a commercial kit (QIAEX II Gel Extraction Kit (Qiagen)).

3.2.6 Generation of WT TmLac construct

The β-galactosidase encoding gene from commercial *T. maritima* was amplified from genomic DNA (DSM3109) (1 ng/μl) using primers JM771 and JM772, which contained restriction sites for *Sac*I and *Pst*I to facilitate cloning into pQE-80L vector (Table 3.2). The primers were designed based on the TmLac gene sequence published by Moore et al. (1994). The purified PCR product and vector pQE-80L were digested with restriction enzymes *Sac*I and *Pst*I at 37°C for 1.5 hours, and purified after electrophoresis. Subsequently, the vector and the gene were ligated by incubation with T4 DNA ligase at 16°C overnight. The ligation mixture was used to transform XL1-BLUE strain and the resulting plasmid was named pTmWT.

Table 3.2: Primers used in the cloning of β-galactosidasa gene from *T. maritima*. Restriction sites are indicated in italics.

Primer	Sequence	Characteristics
JM771	CAC*GAGCTC*AAGAATATGCCCTACGAATGGG	Forward with *Sac*I restriction site
JM772	GTG*CTGCAG*TCACCTCACGTAGATAGTTTTTCTC	Reverse with *Pst*I restriction site
DT889	AGCC*GTCGA*CCCTCACGTAGATAGTTTTTCTCGTG	Reverse without stop codon and *Sal*I restriction site

Figure 3.2: Constitution of CBM-containing TmLac pQE_80L plasmid.

3.2.7 Construction of TmLac_CBM hybrids

To generate TmLac_CBM hybrids, the TmLac encoding sequence, excluding the stop codon, was previously cloned within the *Sac*I/*Sal*I sites of pQE80L, using primers JM771/DT889 (Table 3.2). Subsequently, DNA sequences encoding different CBMs were amplified from genomic DNA of the corresponding organism and fused to the C-terminus of the β-galactosidase gene of *T. maritima* as shown in Figure 3.2. CBM2 and CBM5 from *P. furiosus*; CBM19 from *S. cerevisiae* and CBM9 from *T. maritima* were selected. Genomic DNA from *T. maritima* (DSM310) and *Pyrococcus furiosus* (DSM3638) were pucharsed from DSMZ collection, and DNA from *S. cerevisiae* BY4741 was isolated as previously described (Section 3.2.3). Primers used to amplify the different CBMs

are listed in Table 3.3. In some cases, a linker of two flexible residues was inserted between the TmLac and the CBM region. Resulting plasmids were named pTm_CBM9, pTm_CBM2, pTm_CBM5 and pTm_CBM19, accordingly to the CBM module used in each case.

3.2.8 Site-directed mutagenesis

Site-directed mutagenesis was performed using a modification of the protocol described by Hemsley et al. (1989). The method consists in the amplification of the plasmid carrying the target gene using diverging oligonucleotides with the desired mutation. Subsequent phosphorylation of 5' ends by polynucleotide kinase (Thermo Scientific) and final religation with T4 DNA ligase (Thermo Scientific) were carried out (Figure 3.3). Amplification was performed using a DNA polymerase with proof reading ativity that produces blunt ends (Phusion High-Fidelity DNA Polymerase, Thermo Scientific) following manufacturer instructions with the primers shown in Table 3.4. *E. coli* XRA strain was used as the final recipient strain for protein overproduction.

3.2.9 DNA sequencing

The accuracy of gene amplification of TmLac gene, CBMs and all generated mutants was tested by DNA sequencing carried out by the Central Service for Experimental Research from Universitat de València, using a 3730 DNA analyzer (Applied Biosystems).

Table 3.3: PCR primers used to generate TmLac_CBM constructs. Restriction sites are indicated in italics.

Primer	**Sequence**	**Comments**
DT878	CTATCC*GTCGA*CACGCTTTCCCCAACAACG	Forward primer for CBM19 amplification introducing *Sal*I restriction site.
DT879	GGACTA*AAGCTT*TTAAAAGTAATTGCTTTCCAAATAAGAG	Reverse complementary primer for CBM19 amplification introducing *Hind*III restriction site.
DT890	ATGA*GTCGAC*GGCGCAGGAATAATGGTAGCGACAGC	Forward primer for CBM9 amplification introducing *Sal*I restriction site and Gly + Ala between TmLac and CBM as linker (underlined).
DT891	GGAGAC*AAGCTT*TCACTTGATGAGCCTGAGGTTAC	Reverse complementary primer for CBM9 amplification introducing *Hind*III restriction site.
2DT21	ATGA*GTCGAC*GGCGCAAGCACTTCAATGGAAAACCTAGTC	Forward primer for CBM5 amplification introducing *Sal*I restriction site and Gly + Ala between TmLac and CBM as linker (underlined).
2DT22	GGAGAC*AAGCTT*AGTTGGCTCTGCCTCACCC	Reverse complementary primer for CBM5 amplification introducing *Hind*III restriction. site.
2DT23	ATGA*GTCGAC*GGCGCAACAACTACCCCTGTCCCAGTCTC	Forward primer for CBM2 amplification introducing *Sal*I restriction site and Gly + Ala between TmLac and CBM as linker (underlined).
2DT24	GGAGAC*AAGCTT*AATTACTTGTCCGTTTATTTCTAGGGTTATTTC	Reverse complementary primer for CBM2 amplification introducing *Hind*III restriction site.

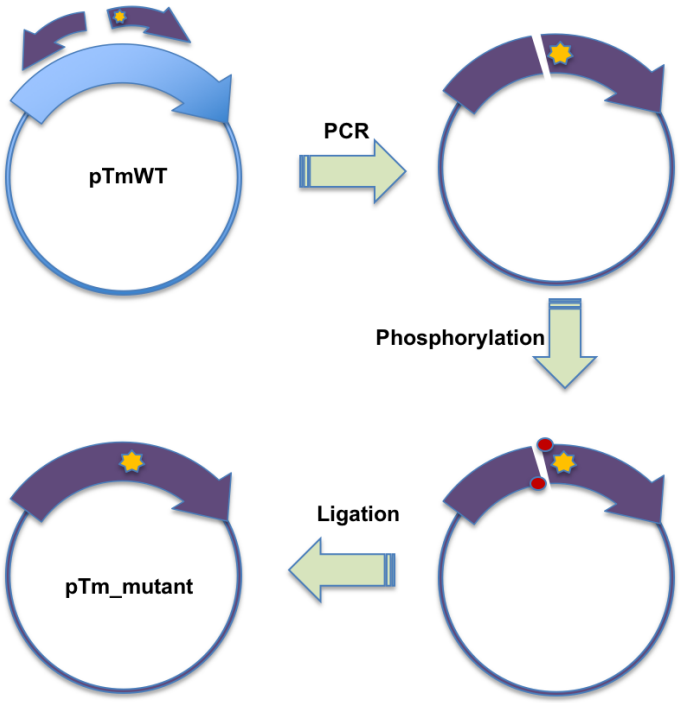


Figure 3.3: Scheme of site directed mutagenesis protocol. Light blue arrow corresponds to the β-galactosidase gene. The plasmid to be mutated (pTmWT) was amplified using two primers which anneal adjacent in opposite directions (purple arrows) one of which was designed including the mutation of interest (star). The mutant plasmid was recovered after phosphorylation (red dots) and religation of the amplified fragment (Hemsley et al., 1989).

Table 3.4: Primers used to generate mutant versions of TmLac. Plasmid pTmWT was used as template. Underlined sequences correspond to introduced mutations.

Primer	**Sequence**	**Plasmid**
DT931	GTTTATCCATTTGAACCGAACCC	pTm_V93T
DT930	CGTGTTCGTGTAGATGGGCTTTCC	
DT932	ACTTATCCATTTGAACCGAACCCTCC	pTm_V94T
DT932	CACGTTCGTGTAGATGGGCTTTC	
DT933	CAGTATCCATTTGAACCGAACCCTCC	pTm_V94Q
DT934	CACGTTCGTGTAGATGGGCTTTC	
DT927	AGCGGAAACTTCTGTATAAACGGTGTG	pTm_D568S
DT929	ATTCGGTGTGTCACCGAAGTCAC	
DT928	GCCGGAAACTTCTGTATAAACGGTGTG	pTm_D568A
DT929	ATTCGGTGTGTCACCGAAGTCAC	
DT829	TATAAACGGTGTGGTACTGCCCG	pTm_F571L
DT935	CATAAGTTTCCGTCATTCGGTGTGTC	
DT830	TATAAGCGGTGTGGTACTGCCCGATAG	pTm_N574S
DT831	CAGAAGTTTCCGTCATTCGGTG	
DT925	TATAGCCGGTGTGGTACTGCCCGATAG	pTm_N574A
DT831	CAGAAGTTTCCGTCATTCGGTG	
DT942	TATAGACGGTGTGGTACTGCCCGATAG	pTm_N574D
DT831	CAGAAGTTTCCGTCATTCGGTG	
DT859	GCGGGTGCGATGCCTCATCTG	pTm_W959A
DT834	GCTGTCGTCTCCTCCAAGGC	
DT859	GCGGTGCGATGCCTCATCTGG	pTm_W959C
DT834	GCTGTCGTCTCCTCCAAGGC	
DT833	TTGGTGCGATGCCTCATCTGG	pTm_W959F
DT834	AGCTGTCGTCTCCTCCAAGG	
DT830	CATAAGTTTCCGTCATTCGGTGTGTC	pTm_F571L/N574S
DT935	TATAAGCGGTGTGGTACTGCCCGATAG	
DT925	CATAAGTTTCCGTCATTCGGTGTGTC	pTm_F571L/N574A
DT935	TATAGCCGGTGTGGTACTGCCCGATAG	
2DT8	CAGCCAAAGACCTTGAAGTAACACTCATCACC	pTm_E242A/E243A/E244A
2DT9	CCGCCTCACCGAGATTTCTCATCTCTAC	
2DT10	GCCGCATCCATTTCAGATGGAGTGTCACTC	pTm_E697A/K698A/K699A
2DT11	CGCAAATGCCGGTGCTTTCAGAAG	
2DT12	GCCGCACAGGTGGTTCTCTTTGTTGATG	pTm_K1029A/E1030A/K1031A
2DT13	CGCGCTCAGTGGAGTGTCGTTCAG	

3.3 Gene expression, protein extraction and purification

Two host strains were used for heterologous gene expression: *E. coli* ROSSETA2 and XRA. In preliminar assays ROSSETA2 strain showed better yields for TmLac production, and this was the choice to overproduce and purify proteins for crystallization assays. However enzymatic assays of the different TmLac mutants were performed directly from total protein extracts. Therefore, in order to avoid a putative interference of *E. coli* endogenous β-galactosidase, the strain XRA (Δlac) was selected in these cases.

Transformant bacteria were cultured at 37°C in 2XTY medium with ampicillin (100 mg/ml) and chloramphenicol (68 mg/ml) until an OD of 0.6 at 600 nm was reached. Gene induction was carried out with IPTG (1 mM final) for 3 to 5 hours at 37°C or for 16 hours at 16°C, depending on the optimum conditions for each protein. The cells were collected by centrifugation at 4000 *g* for 20 minutes at 4°C and the pellet was concentrated 20-100 fold in buffer A (20 mM phosphate buffer, pH 7.4, 10 mM imidazole, 500 mM NaCl) supplemented with protease inhibitor cocktail (Complete EDTA-free (Roche)). Highly concentrated cell suspensions used for protein purification were incubated 30 minutes with lysozyme (1 mg/ml). In all cases, cells were disrupted by sonication in a B. Braun-BIOTECH LABSONIC with 5 pulses of 30 seconds at 185 W and a 0.5 frequency, alternating with 30 seconds of incubation on ice.

Protein extracts were subjected to a heating treatment of 15 minutes at 85°C in order to denature mesophilic proteins. Afterwards, samples were centrifuged at 15000 *g* for 20 minutes at 4°C and the supernatant was further

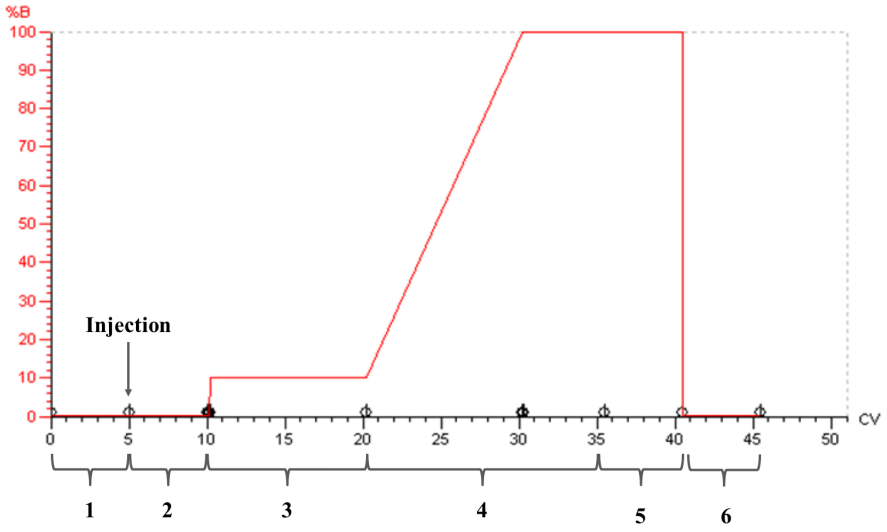


Figure 3.4: Ni-affinity cromatography steps used for purification of β-galactosidase from *Thermotoga maritima*. (1): equilibration, (2) washing unbound proteins, (3) washing of loosely bound proteins, (4) elution of tightly bound proteins,(5) final column wash, (6) re-equilibration. X-axis represents column volumes (CV) and Y-axis % of elution buffer (%B).

purified using a nickel affinity column (Nickel Affinity 5 ml Cartridge Company ABT). Chromatography conditions are detailed in Figure 3.4. The column was equilibrated with buffer A (10 mM imidazole, 500 mM NaCl and 20 mM phosphate buffer, pH 7.4) and elution was carried out with increasing concentrations of buffer B (500 mM imidazole, 500 mM NaCl and 20 mM phosphate buffer pH 7.4). The equipment used for chromatography was an ÅKTA-Purifier (GE Healthcare). Tmlac_CBMs hybrids were purified using the same conditions described previously, but reducing the column volume to 1 ml (HisTrap FF crude Healthcare) and without the previous heat treatment step. Fractions with more than 95% purity (as determined by SDS-PAGE), were dialyzed in 20 mM Tris-HCl buffer pH 7.5 with 50 mM NaCl and concentrated when required by ultra-filtration with a 20 kDa pore size filter (Pierce).

Crude protein quantification was carried out by Bradford assay (Bradford, 1976) and purified protein quantification was performed spectrophotometrically using theoretical extinction coefficients based on sequence composition.

3.4 Protein electrophoresis and western blot

Samples were mixed in a 1:1 proportion with 2X protein loading buffer (125 mM Tris-HCl, 4% SDS, 20% glycerol, 0.05% bromophenol blue and β-mercaptoethanol) and heated at 95°C for 10 minutes. Electrophoresis on polyacrylamide gel 10% (100 V 1 hour) was performed as described in Sambrook et al. 1989, using the OmniPage Mini device (Cleaver Scientific). Gels were stained with Coomassie blue solution (Coomassie brilliant blue R250 0.25% (w/v), methanol 50% (v/v)) and 10% acetic acid (v/v). Alternatively, for western blot analysis, the proteins were transferred to nitrocellulose membranes (1h, 100V) with a commercial transfer system Mini Trans-BlotR (BioRad) in TB buffer (25 mM Tris, 150 mM glycine and 20% methanol, pH 8.3). After a blocking treatment with 3% BSA in TBS (10 mM Tris, 150 mM NaCl pH 7.5) for 30 minutes at room temperature, membranes were incubated overnight at 4°C with rabbit anti-His antibody (His-probe H-15, Santa Cruz Biotech) diluted 1:200 in TBS and 3% BSA. After several washes of 10 min with TBS-T (0.1% Tween 20 in TBS) the membrane was incubated with the secondary antibody, conjugated with peroxidase (anti-Rabbit IgG HRP, Santa Cruz Biotech) diluted 1:5000 in TBS with 2.5% milk powder, for 2 hours at room temperature. After washing with TBS-T the membrane was analysed with a chemiluminescent substrate (ECL western blotting substrate Pierce, Thermo Scientific) and the signal was detected with a CCD (Fujifilm LAS-1000 plus) camera. Image analysis for band quantification was performed using

Quantity One software (Bio-Rad).

3.5 Protein immobilization

3.5.1 Covalent immobilization of TmLac to magnetic beads

For covalent enzyme immobilization, magnetic beads with epoxy activated groups (M-PVA Chemagen E02) were used. The beads, preserved in acetone, were washed with binding buffer (50 mM sodium borate buffer, 50 mM KCl pH 10) and equilibrated twice for 10 minutes with the same buffer. After every wash, the spheres were separated from the liquid using a magnet. The enzyme (0.25 mg/ml) was incubated with the spheres (1.5 mg/ml) in binding buffer with 2 M $(NH_4)_2SO_4$ for 4 hours at 37°C with shaking at 180 rpm. Unreacted epoxy groups were blocked by incubation with 250 mM Tris pH 8 buffer for 2 hours with shaking at room temperature. Finally the enzyme loaded beads were re-suspended in 20 mM Tris, 50 mM NaCl pH 7.5 buffer to be preserved at 4°C until use.

3.5.2 CBM-mediated TmLac immobilization

Non-covalent immobilization was carried out using either chitin or cellulose as support. These experiments were performed using either a crude cell extract or purified protein.

Protein crude extract (1 μg/μl) or purified protein (1 μM) was incubated with 10 mg/ml chitin (Sigma C7170) or acid washed-cellulose (Fluka 22184-F) in buffer 20 mM phosphate pH 6.4 or 20 mM Tris pH 8.5 at 37°C with

shaking (200 rpm) during 4 hours. Control samples without polysaccharide were incubated in parallel under the same conditions.

Binding assays with extracts were used to analyze binding specificity of the TmLac_CBM hybrids compared to the rest of *E. coli* proteins. In this case, after incubation with the polysaccharide, samples were centrifuged (5 min. 15000 *g*) and pellets were subjected to sequential washing steps to reduce the unspecific binding (Figure 3.5). Finally pellets with bound enzyme were re-suspended in 250 μl of assay buffer (50 mM phosphate, 10 mM NaCl, 1 mM $MgCl_2$, pH 6.5) for an SDS-PAGE analysis. When purified protein was used, samples containing chitin or cellulose were centrifuged after binding and the β-galactosidase activity of the supernatant (unbound enzyme) and that of the control samples (total enzyme) was assayed with *p*-nitrophenyl β-D-Galactopyranoside (*p*NP-β-Gal). These data were used to calculate the percentage of bound enzyme to each support (Figure 3.6). In order to study binding stability, pellets were washed with 800 μl of the corresponding binding buffer (pH 6.5 or 8.5) to remove any excess of unbound protein and were resuspended and incubated with assay buffer (50 mM phosphate, 1 mM $MgCl_2$, pH 6.5) containing or not a mix 1:1 of glucose and galactose (150 mM each) during 30 min. at 75°C. After centrifugation the supernatant (including released enzyme) was collected. The pellet (bound enzyme) was washed with 800 μl of assay buffer and finally resuspended in 250 μl of the same buffer. Both released and bound samples were analyzed by SDS-PAGE (Figure 3.6).

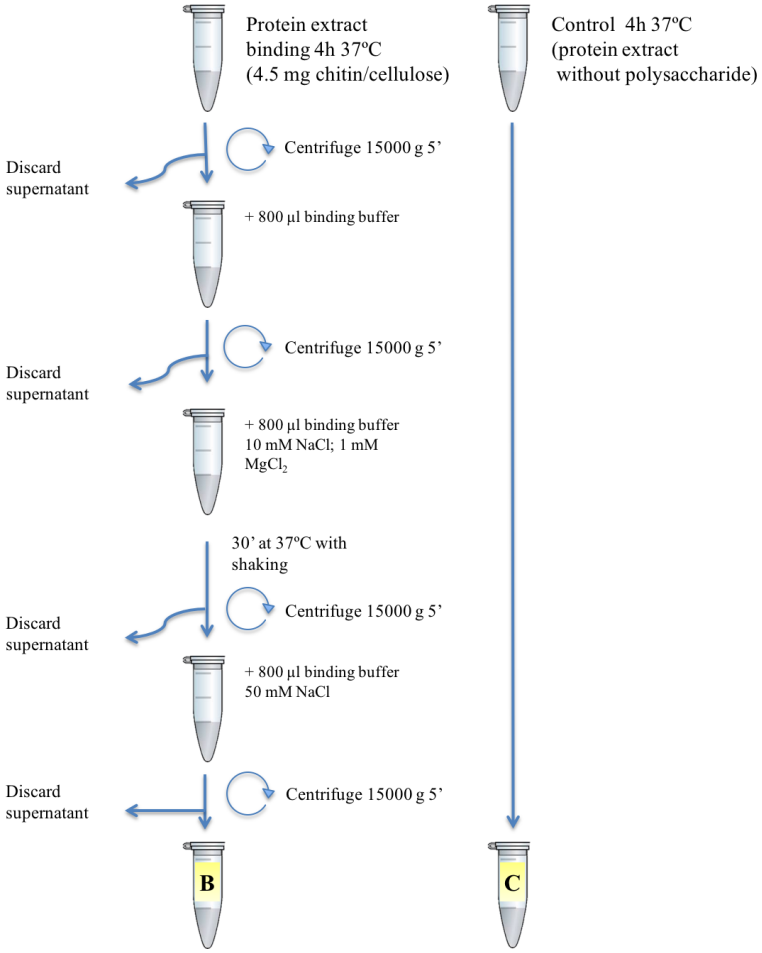
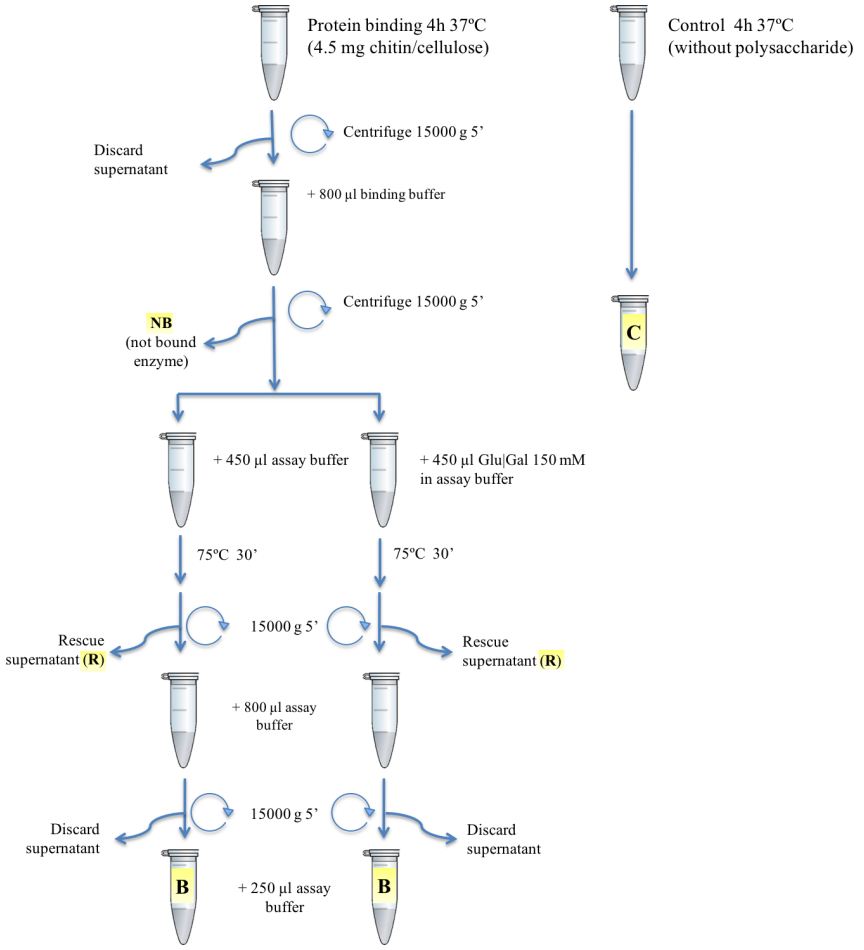


Figure 3.5: Analysis of binding affinity to cellulose or chitin with protein extracts containing WT TmLac or TmLac_CBM hybrids (B = bound enzyme; C = control).



$$\% \, binding = \left[1 - \frac{A_c - A_{NB}}{A_c}\right] \cdot 100$$

Figure 3.6: Analysis of binding affinity and binding stability with purified WT TmLac and TmLac_CBM hybrids. B and R correspond to the bound and released fractions after treatment of immobilized enzyme at 75°C in different media. The percentage of binding was calculated as indicated where A_C and A_{NB} are the β-galactosidase activity of the control (C) and not bound (NB) samples, respectively. B and R correspond to the bound and released fractions after treatment of immobilized enzyme at 75°C in different media.

3.6 Enzyme assays

3.6.1 β-galactosidase assays with *p*NP-β-Gal

Enzymes were incubated with 5 mM *p*NP-β-Gal in assay buffer (50 mM phosphate, 10 mM NaCl, 1 mM $MgCl_2$, pH 6.5) at 75°C. The reactions were stopped by adding 1 ml of 1 M Na_2CO_3. The activity was spectrophotometrically monitored by measuring the optical density at 400 nm which corresponds to the maximum absorbance of *p*-nitrophenyl at basic pH.

3.6.2 β-galactosidase assays with lactose

Enzyme preparations were incubated with lactose at different concentrations in assay buffer (50 mM phosphate, 10 mM NaCl, 1 mM $MgCl_2$, pH 6.5) or milk at 75°C. The reaction was stopped by heating at 95°C for 10 minutes. The amount of glucose released was determined using a commercial glucose assay kit (Glucose (GO) Assay Kit, Sigma). One unit of activity was defined as the amount of enzyme releasing one μmol of glucose per minute using lactose at 25% (w/v) as substrate.

3.6.3 Measurement of enzyme GOS production

Enzyme preparations (0.7 - 0.8 U/ml) were incubated with lactose at 25% (w/v) in assay buffer at 75°C under shaking (200 rpm). The reaction was stopped by heating at 95°C for 10 minutes. The product profile was analyzed using anion exchange chromatography equipped with a CarboPac PA-100 column and an amperometric detector (Dionex, Thermo Scientific). The program used for each chromatographic separation is shown in Table

Table 3.5: Program used for the analysis of synthesized GOS by anion exchange cromatography in a DIONEX equipment. The solutions used in the mobile phase were: (A) water; (B) 1 M NaOH; (C) 1 M NaOAc and (D) water. Time 0 represents the time of sample injection and negative times correspond to washing and equilibration steps. Concentration changes were performed with linear gradients.

Time (min)	%A	%B	%C	%D
-35.0	25	30	20	25
-30.1	10	30	50	10
-30.0	45	10	0	45
0.0	45	10	0	45
2.0	45	10	0	45
5.0	35	30	0	35
12.5	27,5	30	15	27,5
12.6	45	10	0	45

3.5. Sugar standards (β-6'-galactosyl-lactose, β-3'-galactosyl-lactose, glucose and galactose) were analyzed in the same conditions to identify and quantify reaction products. All samples were clarified before injection through a filter of 0.45 μm pore size.

3.7 Bioinformatic tools

3.7.1 Analysis of DNA and protein sequences

DNA sequence analysis was carried out with DNAMAN 4.03 (Lynnon Biosof) and Serial Cloner 2.6.1 (http://serialbasics.free.fr/Softwares.html).

Protein and DNA sequence alignments were made with Bioedit (Hall, 1999) using Clustal W algorithm (Larkin et al., 2007). Protein molecular weight and theoretical extinction coefficient were calculated using ProtParam server (Walker et al., 2005).

3.7.2 Protein structure modeling and docking

Structure modeling of TmLac was carried using the I-TASSER server (Zhang 2008; Roy et al., 2010 and Zhang et al., 2015). The structure obtained were analyzed with Pymol Software (Schrödinger, LLC, 2010). Docking of β-3'-galactosyl-lactose and β-4'-galactosyl-lactose in TmLac and *Bacillus circulans* (PDB: 4YPJ) β-galactosidase was performed using Autodock sofware (Morris et al., 2009). β-3'-galactosyl-lactose and β-4'-galactosyl-lactose coordinates were obtained by Glycam Web Tool (http://glycam.org (Accesed in 2016)) (Kirschner et al., 2008). Using AutodockTools, hydrogens for ligand and receptor were added and charges were assigned with the t method. All the hydroxyl bonds and glycosidic bonds were made rotatable, but not the sugar ring bonds. Autodock 4.0 was executed 100 times using Lamarckian algorithm (LGA), a population of 150, mutation rate of 0.02 and a rate crossover of 0.8. Simulations were performed using a maximum of 2500000 energy evaluations and a maximum of 27000 generations. Results of the docking were grouped using a cut value of 2.0 Å.

3.7.3 Analysis of domain architectures

GenBank accesion numbers of protein sequences classified in the GH2 family were retrieved from the CAZy database (Lombard et al., 2014) and mapped in the Uniprot database (Consortium, 2015) using the tool "Retrieve/IDmap-

ping" (http://www.uniprot.org/uploadlists/(Accessed November 2015)). This operation yielded the UniProt code, GI number and Pfam domains of the different sequences, including the start and end coordinates of the domain limits (envelope) (Finn et al., 2014). A binary vector of 69 digits was assigned to each sequence using a PYTHON language script (*Domain_study.py*, detailed in Appendix A). This vector indicates the presence (1) or absence (0) of a specific Pfam domain and its relative position in order from N-terminus to C-terminus, including the existence of tandem repeats. The amino acid sequences of the different proteins, in FASTA format, were retrieved from the NCBI database, using the Batch Entrez Tool (http://www.ncbi.nlm.nih.gov/sites/batchentrez (Accessed in November 2015)). Finally, a matrix was generated that included the GI number, the binary vector, and the associated domain architecture (DA). We define DA as the linear composition of domains of a given sequence in N-terminal to C-terminal order. Only sequences that contained a single catalytic GH2C domain, with coverage higher than 70%, were selected for further analysis. The set of DAs was refined to take into account C-terminal sequences corresponding to domains not identified by Pfam. We took into consideration the presence or absence of an extra C-terminal sequence, its length and existing homology to stablish a classification of the selected proteins. For this analysis, the protein sequences were also analyzed using the Interpro platform (Mitchell et al., 2015) in order to get information complementary to that provided by Pfam. The pipeline for DA analysis is sumarized in the Figure 3.7.

3.7.4 Phylogenetic analysis

The sequence of the catalytic domain (GH2C) was extracted for a selection of proteins representative of the different domain architectures (DAs).

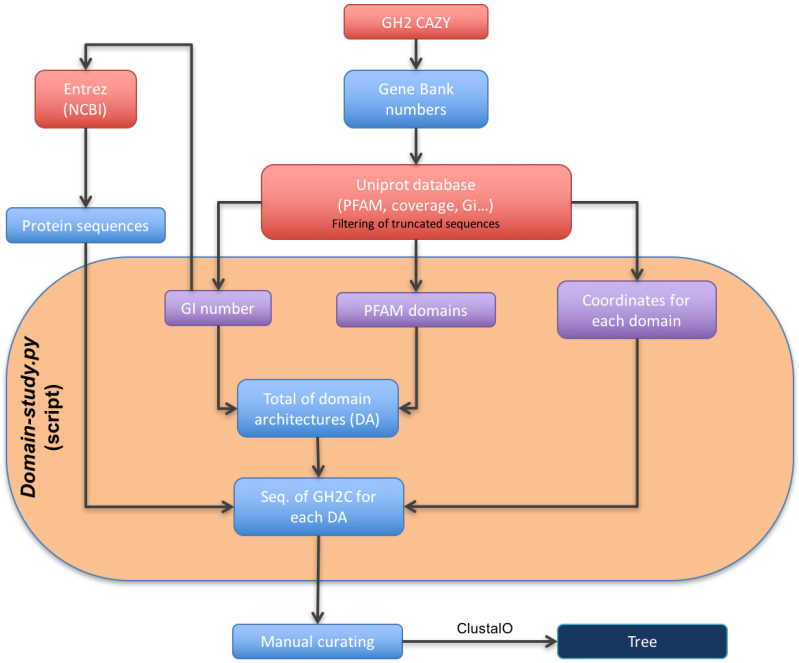


Figure 3.7: Pipeline followed for the obtention of the GH2 domain composition and tree construction.

Alignment of the GH2C sequences was performed using Clustal O MSA (Sievers et al., 2011) and CLCsequence viewer (QIAGEN). Trees were built using Maximum Likehood Parsimony algorithm with JTT matrix (Jones et al., 1992) and a boostrap of 100 replicates using MEGA6 software (Tamura et al., 2013). Results were visualized on Dendroscope Software (Huson and Scornavacca, 2012) and represented using FigTree software (http://tree.bio.ed.ac.uk/software/figtree/). For the classification of proteins with unidentified C-terminal domains, these sequences stretches were extracted and aligned as described above. Subclusters were obtained from the resulting phylogenetic trees (not shown).

Chapter 4

Results

4.1 Production of *Thermotoga maritima* β-galactosidase

PQE-80L was chosen as expression vector for the production of TmLac carrying a His-tag on its N-terminal end. This vector has a promoter-operator region containing two sequences of the *lac* operator that increase repressor affinity in absence of an inducer, reducing the possibilities of undesired basal expression. Since gene expression may be decreased by the presence of rare codons, the TmLac sequence was analyzed with "Rare Codon Calculator" software from NIH MBI Laboratory for *Structural Genomics and Proteomics Server* (http://nihserver.mbi.ucla.edu/RACC/). Up to 59 of the 65 arginine residues existing in the protein are encoded by rare codons in *E.coli*, some of them in tandem repeats, which represent 9% of total codons. Therefore, we compared TmLac production using two host strains: XL1-Blue, a regular strain for heterologous gene expression, and ROSSETA2, which contains the plasmid pRARE2, that encodes for 12 tRNAs corresponding to rare codons in *E. coli*. Expression was much more efficient in the latter strain (Figure 4.1).

Recombinant, histidine-tagged TmLac was purified from induced cultures of *E. coli* ROSETTA2 transformants by heating treatment, followed by a clarification step by centrifugation and affinity chromatography through a nickel column. Samples at different steps of the purification protocol were analyzed by SDS-PAGE (Figure 4.2). TmLac migrated as a prominent band in the crude extract corresponding to the expected mass of 129 kDa and was recognized by an anti-histidine antibody in a immunoblot assay. After heat treatment (85°C) most of the endogenous *E. coli* proteins precipitated whereas, a significant fraction of TmLac was recovered in the soluble fraction

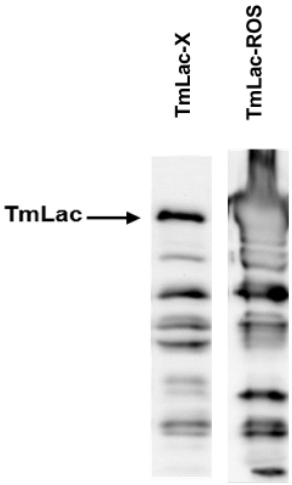


Figure 4.1: Differences in TmLac gene expression between strain XL1-Blue (TmLac-X) and ROSETTA2 (TmLac-ROS). TmLac from a cellular extract of IPTG induced cultures was analyzed by SDS-PAGE and immunolabelled with anti-HIS antibodies. The band corresponding to the expected size of TmLac is shown with an arrow.

in agreement with the thermophile nature of the enzyme. This single step treatment rendered a TmLac sample with 95% of purity. Some of the minor bands of lower size were recognized by the anti-His antibody suggesting that they were possibly caused by proteolytic cleavage of the protein at exposed loops that did not compromise the assembly. Further purification by nickel-affinity column resulted in a preparation with a higher ratio of full length to fragmented enzyme.

To determine the molecular mass of native TmLac, the purified protein was analyzed by size exclusion cromatography (Figure 4.3). The protein eluted at a volume corresponding to ca. 800 kDa (±200 kDa). This result indicates that the quaternary structure of the enzyme is composed of 5-9 subunits. The enzyme showed tendency to aggregate at this concentration since two peaks

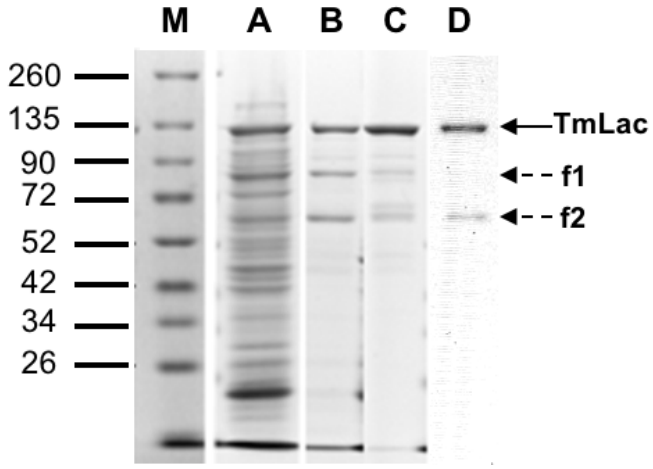


Figure 4.2: SDS-PAGE analysis of protein samples obtained during the purification process. *Lane A* soluble extract. *Lane B* soluble protein after heat treatment (85°C, 15 min.). *Lanes C and D* protein recovered after nickel-affinity cromatography. *Lane M* molecular mass markers. Protein bands were deleted by Coomasie blue R250 staining (*lanes A-C*) or by immunoblotting with an anti-His antibody (Santa Cruz Biotech, sc-803) (*lane D*). Bands corresponding to TmLac and proteolytic fragments (f1 and f2) are indicated by arrows.

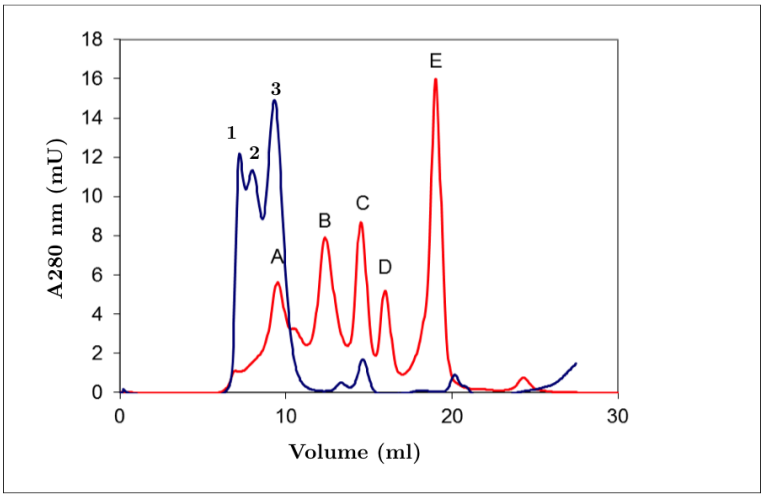


Figure 4.3: Size exclusion cromatography of TmLac. A TmLac sample (2.5 mg/ml) was injected in a Superdex 200 column 10/300 (GE-Healthcare) previously equilibrated with buffer (50 mM phosphate buffer pH 7.0, 150 mM NaCl) and eluted with 2 column volumes at 0.5 ml/min. Peaks corresponding to native TmLac *(3)* or to aggregates equivalent to 12-mer *(2)* or 20-mer *(1)* are indicated. Red curve corresponds to the molecular mass standard (Bio-Rad, cat #151-1901). A = Thyroglobulin (bovine) (670 kDa); B = γ-globulin (bovine) (158 kDa); C = Ovalbumin (chicken) (44 kDa); D = Myoglobin (horse) (17 kDa); Vitamin B12 (1.35 kDa).

of higher molecular mass, corresponding to a mass equivalent to ∼12 and ∼20 monomeric units, were also observed in the chromatogram.

4.2 TmLac covalent immobilization

TmLac covalent immobilization was assayed using an epoxy-activated resin as support. Purified enzyme was incubated with epoxy-coated magnetic beads at high pH and in the presence of ammonium sulphate. A control sample of enzyme without beads was incubated in the same conditions. After incubation,

the beads were captured using a magnet to separate the unbound fraction, and then treated with Tris buffer to wash out loosely bound enzyme and block the remaining free epoxy groups. Upon incubation, the control sample showed β-galactosidase activity whereas none was detected in the not-bound fraction nor in the washing solution, indicating that virtually 100% of the enzyme was linked to the epoxy beads (20 mg of enzyme/g of beads). Equivalent amounts of free and bound enzyme were incubated under denaturing conditions that would disassemble the oligomeric form of the enzyme and consequently released the non-covalently linked monomers from the beads. The amount of TmLac was 60% indicating that 40% of the monomers were bound covalently (Figure 4.4).

Immobilized enzyme showed higher activity than the free form using either *p*NP-β-Gal or lactose at different concentrations as substrates (Figure 4.5). Activity of both free and bound enzyme, using semi-skimmed milk as the substrate, was similar to that obtained with a solution of 5% lactose (ca. the concentration of lactose in milk) (Figure 4.5 B). This result indicates the absence of a matrix effect in the milk that compromises the activity of TmLac.

Thermal stability of the enzyme after low temperature incubation (4^{o}C) was analyzed. β-Galactosidase activity of concentrated enzyme ($>$ 0.3 mg/ml), either in free or bound form, was stable for several weeks. In contrast, diluted free enzyme (6 μg/ml) showed significant destabilization, loosing activity progressively down to 50% after 6 hours of cold incubation. However, the activity of the bound enzyme, was stable at the same concentration for the same time (Figure 4.6).

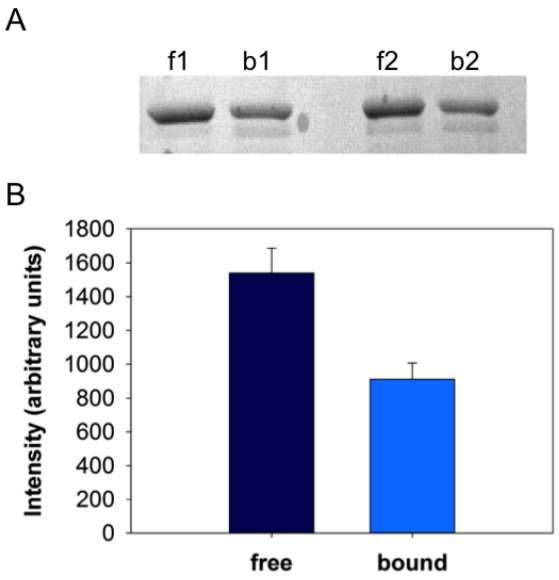


Figure 4.4: TmLac released from the support after denaturing conditions. **A** A duplicate sample of bound enzyme (b1, b2) (0.2 mg/ml) was incubated with 1% of SDS at 95°C during 10 minutes to induce holoenzyme disassembly and the release of non-covalently linked subunits from the beads. An equivalent amount of free enzyme (f1, f2) was treated in the same conditions. After centrifugation of the sample (15.000 g x 10 min), the supernatant was analyzed by SDS-PAGE and Coomassie staining. **B** Band intensity was quantified with software Quantity-One (Bio-Rad). *Error bars* indicate standard deviation of duplicates.

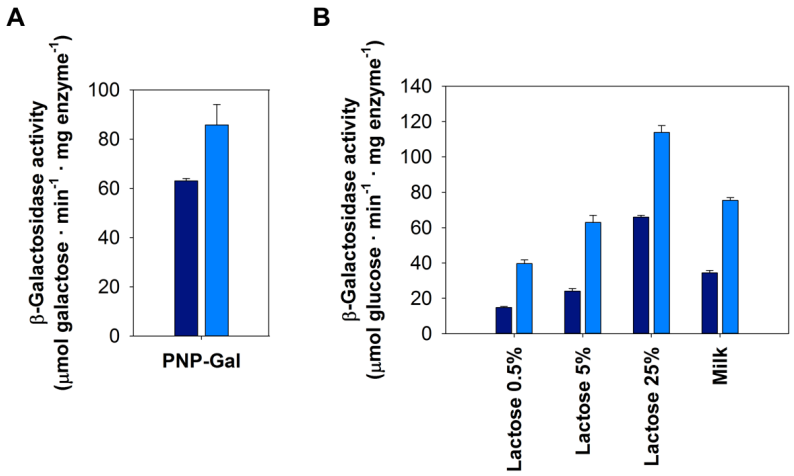


Figure 4.5: β-Galactosidase activity of free (dark blue) and bound (light blue) TmLac was assayed with different substrates. **A** *p*NP-β-Gal (5 mM); **B** lactose from milk or buffered solutions, at different concentrations. *Error bars* indicate standard deviation of triplicates.

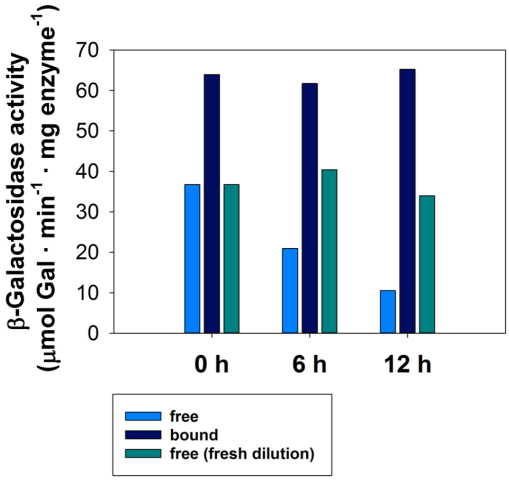


Figure 4.6: Activity of free and bound enzyme after incubation at 4°C. β-Galactosidase activity was measured with *p*NP-β-Gal using free or bound enzyme diluted at 6 μg/ml and kept at 4°C for the indicated times. A freshly diluted sample of free enzymes was analyzed at each time as a control.

4.3 Production and analysis of hybrid TmLac_CBM enzymes

We have assayed non-covalent attachment to polysaccharides with hybrid versions of TmLac containing carbohydrate binding modules. This study also sought to obtain basic information about the binding affinity to chitin or cellulose of previously uncharacterized CBMs.

4.3.1 Production of hybrid *Thermotoga maritima* β-galactosidase fused to different CBMs

Four CBMs from different families were selected: CBM2 from *Pyrococcus furiosus* chitinase A (ChiA), CBM5 from *P. furiosus* chitinase B (ChiB), CBM9 from *Thermotoga maritima* Xylanase A and CBM19 from *Saccharomyces cerevisiae* Cts1 chitinase. These domains were fused to the C-terminal end of TmLac and the hybrid enzymes were characterized. Recombinant TmLac_CBMs were produced in *E. coli*. Soluble extracts before and after a heat treatment were analyzed by SDS-PAGE (Figure 4.7). The electrophoretical mobilities of recombinant proteins were in agreement with the molecular masses of TmLac_CBM5 (136 kDa), Tmlac_CBM2 (148 kDa), Tmlac_CBM2 (139 kDa) and TmLac_CBM19 (138 kDa) predicted by ProtParam (Walker et al., 2005). The hybrids with CBMs from thermophilic sources (CBM9, CBM5 and CBM2) showed a resistance to heat treatment similar to the wild-type enzyme. However TmLac_CBM19 was not recovered in the soluble fraction after heating, probably because the denaturation of the CBM19 module triggered the aggregation of the enzyme. Therefore, this construct was discarded from further experiments.

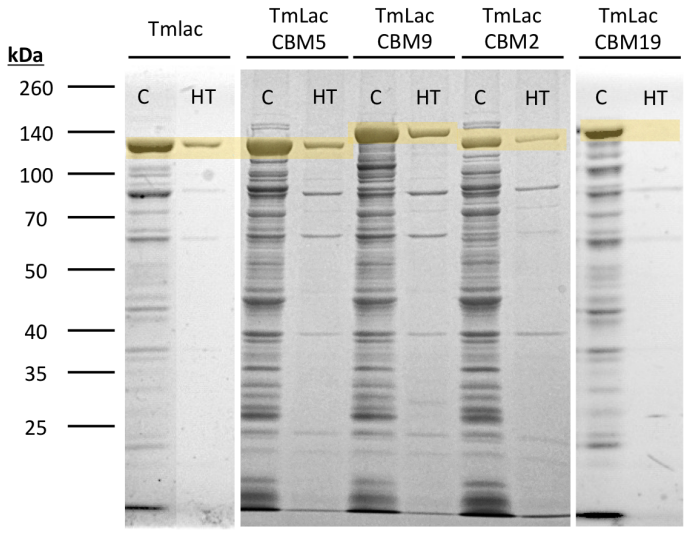


Figure 4.7: Soluble protein of cellular extract from *E. coli* transformants expressing wild-type or hybrid enzymes. SDS-PAGE and Coomassie blue staining were carried out with clarified samples before (C) and after heating 85°C during 5 minutes (HT). TmLac-ROS, TmLac_CBM9-ROS and TmLac_CBM19-ROS were induced at 37°C for 5 hours. TmLac_CBM5-ROS was induced at 37°C for 3 hours and TmLac_CBM2-ROS at 16°C for 16 hours. Bands corresponding to full-length recombinant proteins are marked in yellow.

4.3.2 Study of carbohydrate binding modules as immobilization tags

Substrate affinity (for cellulose or chitin) and binding specificity was monitored for the TmLac thermoresistant hybrids. Preliminary assays indicated that binding to chitin at pH 6.5 occurred with high affinity (12 mg/g of support) both for wild-type and hybrid enzymes, indicating that this represented non-specific binding (not shown). Likely this results from ionic interactions between partially deacetylated, positively charged, chitin and proteins with negative charge at this pH. Therefore, for further studies, binding to chitin at pH 6.5 was not considered. Incubation with cellulose or chitin at pH 6.5 or 8.5 was carried out in parallel to control samples without polysaccharides. After incubation of the hybrids with the support, bound enzyme was subjected to several washing steps and finally analyzed by SDS-PAGE in parallel to control samples. The sample containing enzyme bound to the support (B, in Figure 4.8) was compared to that of total enzyme at the control (C, in Figure 4.8). Binding of TmLac_CBM5 to cellulose or chitin was similar to that of the wild-type enzyme, therefore representing unspecific adsorption (i.e independent of the CBM). This construct was therefore not included in further studies. In contrast, the fractions of Tmlac_CBM9 and TmLac_CBM2 bound to cellulose were significantly higher than that of the wild-type, both at pH 6.5 and 8.5. This was also the case for the binding of TmLac_CBM2 to chitin. In all these instances binding of the hybrid enzyme was higher than for endogenous *E. coli* proteins. This is indicative of the binding specificity of the support for chimeric enzymes, suggesting that protein tagging with CBM2 or CBM9 may be used for one-step immobilization and purification on cellulose or chitin at selected pH values. To carry out a more quantitative evaluation of the binding

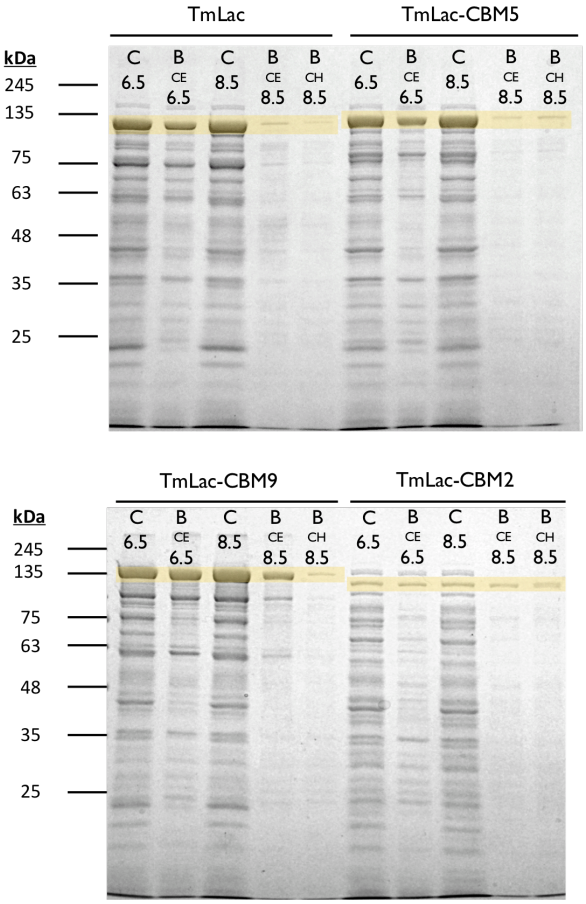


Figure 4.8: Binding affinity of the hybrid and wild-type enzymes specificity analysed by SDS-PAGE. Protein extracts were incubated during 4 hours with acid-washed cellulose (CE) or chitin (CH) in parallel to control samples without support (C) at pH 6.5 or pH 8.5. Bands corresponding to full-length recombinant proteins are marked in yellow.

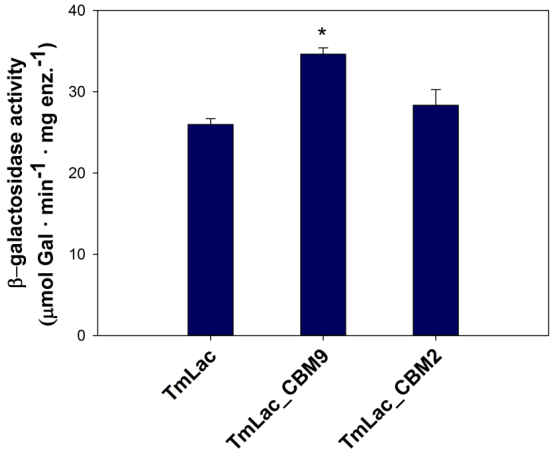


Figure 4.9: β-Galactosidase activity of wild-type and Tmlac_CBM hybrids with 5 mM pNP-β-Gal. *Error bars* indicate standard deviation of triplicates. Asterisk indicate significant difference with wild-type enzyme ($p < 0.05$).

properties of TmLac_CBM2 and TmLac_CBM9, the hybrids and the wild-type enzyme were purified. The specific activity of these enzymes was measured using pNP-β-Gal as substrate (Figure 4.9). TmLac_CBM2 had the same specific activity as wild-type TmLac, whereas Tmlac_CBM9 showed to be around 30% more active than the wild-type enzyme.

To compare the binding efficiency of the different enzymes equal amounts of protein ($\sim$1 μM) were incubated with cellulose or chitin. Tmlac_CBM9 did not bind significantly to chitin at pH 8.5, whereas Tmlac_CBM2 showed a binding efficiency of around 70% of the total enzyme. Immobilization of both hybrids to cellulose at pH 6.5 was around 60-70%, representing 8 mg enzyme/g of support. This binding was ca. 2-fold more efficient than at pH 8.5 (Figure 4.10).

Binding stability of the hybrid enzymes to the support was studied under conditions that mimic those employed for the enzymatic assays with TmLac

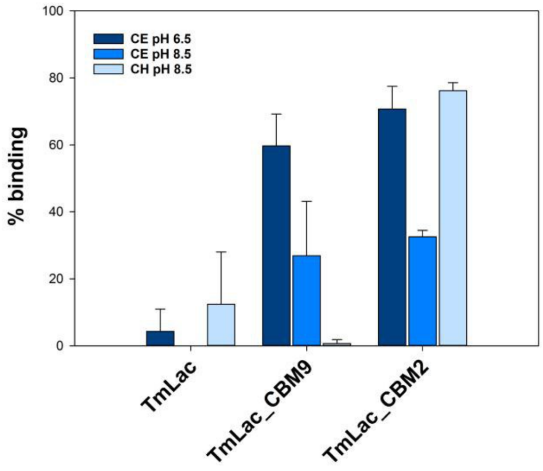


Figure 4.10: Binding efficiency of different TmLac hybrids to acid-washed cellulose (CE) or chitin (CH) at different pH values, compared to wild-type enzyme. *Error bars* indicate standard deviation of triplicates.

(pH 6.5 and 75°C). Moreover, we tested whether sugars like glucose or galactose released from lactose hydrolysis, may interfere with CBM mediated binding. The bound enzyme was incubated for 30 minutes in the presence or absence of a glucose-galactose mix. Afterwards the protein that remained attached to the support and that released to the soluble fraction was analyzed by SDS-PAGE (Figure 4.11). Binding of TmLac_CBM2 was stable in all conditions tested. In the case of TmLac_CBM9 the treatment with glucose and galactose caused detachment of the protein from cellulose, but not from chitin. These sugars seem to displace cellulose from the CBM9 binding site promoting the release of the hybrid enzyme. Low affinity of TmLac_CBM9 to chitin, which is similar to the wild-type enzyme, is probably not mediated by the CBM and therefore it was not influenced by sugar addition.

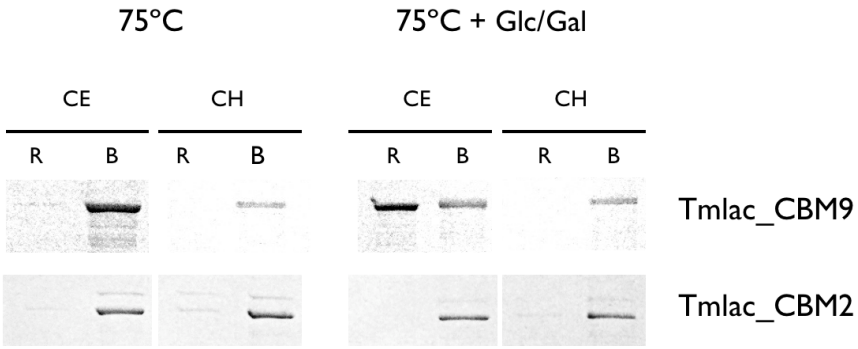


Figure 4.11: Analysis of the binding stability of TmLac hybrids to cellulose or chitin. Hybrid enzymes were immobilized on cellulose (CE) at pH 6.5 or chitin (CH) at pH 8.5. Bound enzyme was incubated at 75°C for 30' in assay buffer (left panel) or Glu/Gal mix (150mM each) in the same buffer (right panel). Released (R) and bound (B) fractions after this treatment were analyzed by SDS-PAGE and Coomasie blue staining.

4.4 Milk lactose hydrolysis

The most common industrial method to remove lactose from milk is the enzymatic hydrolysis to galactose and glucose. This reaction is achieved using lactase, generally β-galactosidase from the yeast *Kluyveromyces lactis*. We have analyzed three commercial lactose-free milk brands marketed in Spain to check whether the enzyme is present in the final product. Samples of lactose-free or regular milk were analyzed by SDS-PAGE after treatment with rennin to remove casein, the major milk protein (Figure 4.12). A differential band was detected in lactose-free milk products with the same size as the major band of *K. lactis* commercial lactase, with an expected molecular weight of 124 kDa. Moreover, β-galactosidase activity was detected in lactose-free milk but not in regular milk using pNP-β-Gal (not shown).

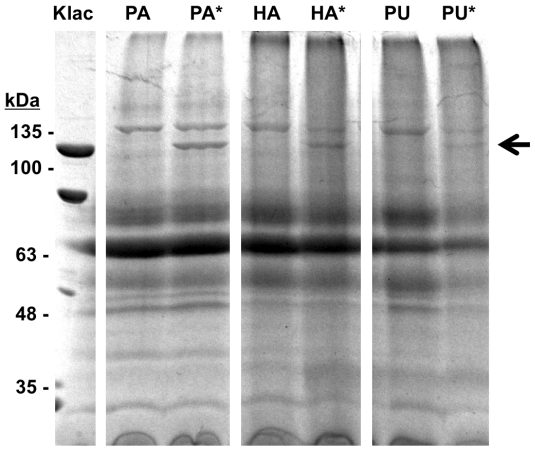


Figure 4.12: SDS-PAGE analysis of semi-skimmed milk after an aspartic protease treatment (1.52 μg/ml) (Rennet (SIGMA)) and 4 mM of $CaCl_2$ during 1 hour at 30°C. The product was centrifuged at 15500 g during 20', and concentrated 5-fold by ultrafiltration through a pore size of 20 kDa (Pierce). The three brands analyzed were: PA: Pascual; HA: Hacendado and PU: Puleva. Asterisk indicate lactose free milk products. Klac corresponds to *Kluyveromyces lactis* lactase from SIGMA (cat. G3665). The differential band found in lactose-free milk is indicated with an arrow.

We have tested the suitability of free and immobilized TmLac to be used for milk lactose hydrolysis. In the assays, lactose at a concentration of 5%, similar to that found in milk, was analyzed. The kinetics with 0.06 mg/ml free wild-type enzyme was similar to that obtained with the covalently immobilized version on epoxy-activated beads at the same concentration (Figure 4.13 A). In both cases 100% hydrolysis was obtained after 3 hours incubation at 75°C. Subsequently, we determined the number of times that the immobilized enzyme could be recycled to hydrolyze 5% of lactose in batches of 3 hours. The performance of the enzyme was not seriosuly compromised after 4 cycles of reutilization, hydrolyzing more than 90% of the initial sugar amount (Figure 4.13 B). Even after 8 cycles of reutilization, the enzyme was able to hydrolyze more than 60%.

A similar study was carried out with non-covalently immobilized enzyme though CBMs. The hybrid TmLac_CBM9 was discarded for this application since binding to cellulose was shown to be unstable in the presence of glucose and galactose concentrations in the range of those released after 5% lactose hydrolysis (Figure 4.11). In contrast TmLac_CBM2 had shown an efficient and stable binding to cellulose or chitin (Figures 4.10, 4.11). TmLac_CBM2 (0.08 mg/ml) immobilized either on cellulose or chitin was able to complete hydrolysis of lactose in 3.5 hours at a similar rate (Figure 4.14 A). Consecutive batches of 5% lactose were incubated 3.5 hours with the recycled enzyme for and the extent of lactose hydrolysis was determined after each batch. The efficiency of the recycled enzyme immobilized either on cellulose or chitin decayed progressively after the first batch and reached 30% of the original activity in the 6th batch (Figure 4.14 B).

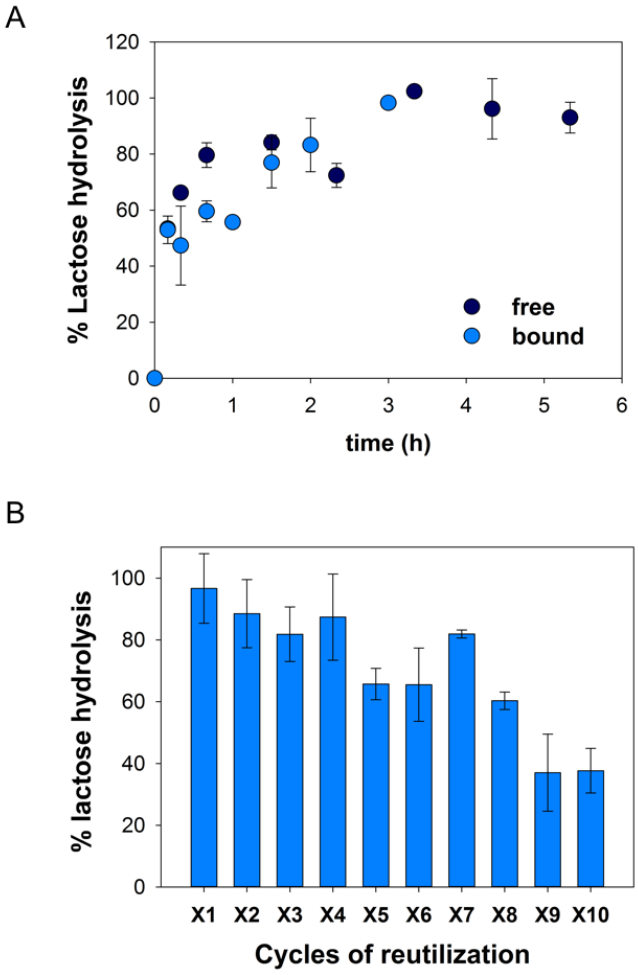


Figure 4.13: Lactose hydrolysis with epoxy-immobilized TmLac. **A** Free and bound TmLac (0.06 mg/ml) were incubated with 5% lactose for the indicated times at 75°C. **B** Bound TmLac was incubated with 5% lactose during 3h at 75°C in 10 repeated cycles. The extent of hydrolysis was calculated from the molar ratio of released glucose to initial lactose. *Error bars* indicate standard deviation of triplicates.

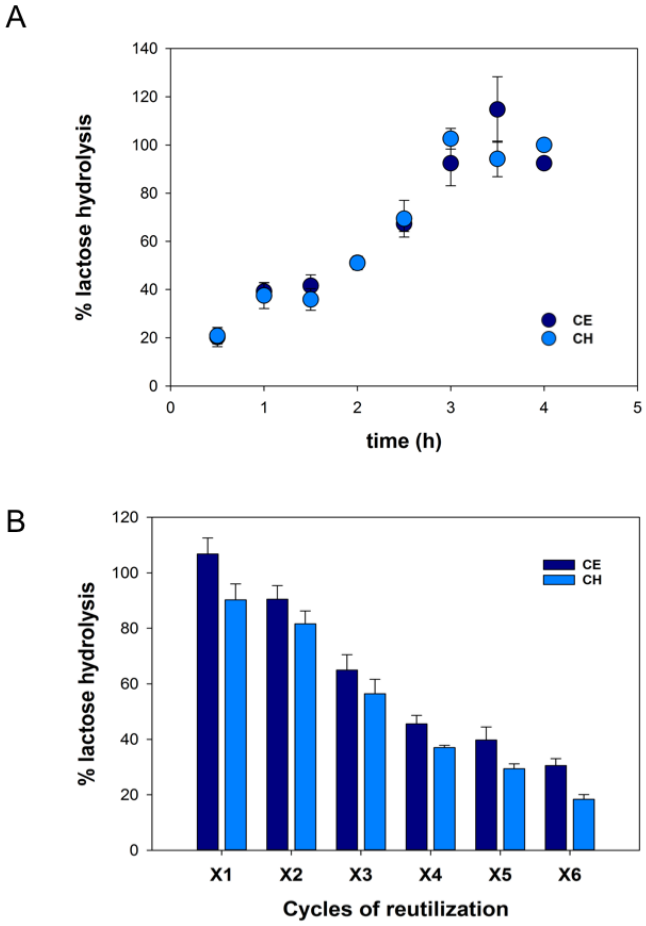


Figure 4.14: Lactose hydrolysis with hybrid TmLac_CBM2 immobilized on cellulose or chitin. **A** Bound TmLac (0.06 mg/ml) was incubated with 5% lactose for the indicated times at 75°C. **B** Bound TmLac was incubated with 5% lactose during 3.5 hours at 75°C in 6 repeated cycles. The extent of hydrolysis was calculated from the molar ratio of released glucose to initial lactose. *Error bars* indicate standard deviation of triplicates.

4.5 Structural analysis of TmLac

4.5.1 Homology based model of TmLac

Since the structure of TmLac has not been fully solved yet, we constructed a structural model of TmLac to facilitate the design of mutants that could improve the transglycosylation capability of the enzyme (Section 4.6). The β-galactosidase from *E. coli*, EcLac (PDB code 3CZJ) which has the highest sequence identity (ca. 40%) with TmLac was used as template (Figure 4.15 A (left)). The resulting structure was formed by an α/β barrel, surrounded by four β-sandwich domains (two N-terminal and two C-terminal) in agreement with the typical architecture of β-galactosidases from the GH2 family. TmLac shows an extra C-terminal extension of ca. 100 residues (Figure 4.16) with a predicted β-sandwich fold (Figure 4.15 B), present in β-galactosidases related with *Thermotoga* but not in other homologous enzymes. This domain was modelled using as template a CARDB-like domain from *Pyrococcus furiosus* which shares 20% of sequence identity (Figure 4.15 A (right)). A detailed study of the catalytic site revealed that residues from loops involved in catalysis or substrate binding in EcLac are highly conserved (70% identity) in TmLac, suggesting a good quality of the model in this region (Table 4.1).

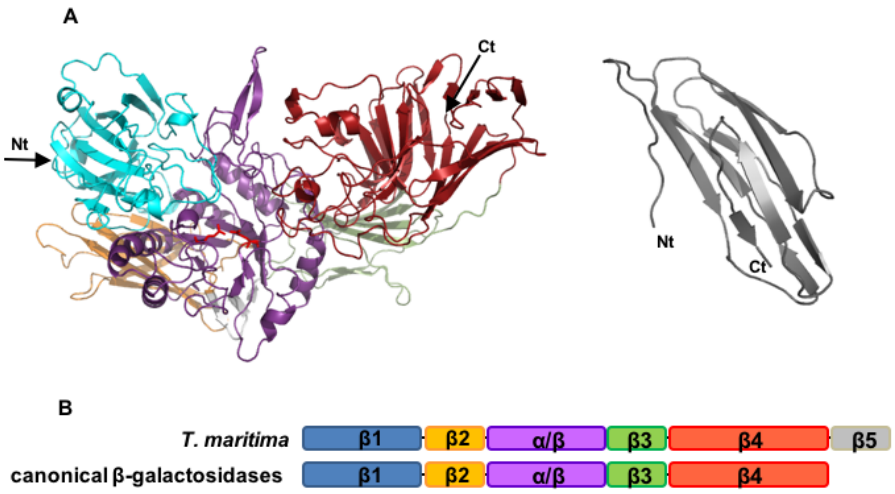


Figure 4.15: Structural modeling of TmLac. **A**. Molecular models of the main body of TmLac (residues 1-983) (left panel) and C-terminal domain (residues 984-1084) (right panel). N-terminal and C-terminal ends of each fragment are indicated. The α/β-barrel surrounding the catalytic residues is coloured in violet, whereas the accompanying β-sandwich domains, homologous to other GH2 enzymes, are coloured in blue, orange, green and red. The β-sandwich specifically found in TmLac and other β-galactosidases from the *Thermotoga* genus is depicted in grey **B**. Schematic representation of the domain architecture of TmLac in comparison with canonical β-galactosidases from the GH2 family.

Table 4.1: Sequence alignment of loops involved in the active site of EcLac. Differential residues between EcLac and Tmlac are underlined.

EcLac	101 YTNVT	200 DQDMWR	356 NRH	389 RCSH
TmLac	90 YTNVV	189 DQDMWW	333 NRH	366 RTSH
EcLac	417 ETHGMVPM	461 NESG	502 PMYARVDEDQ	536 LCEYAHAMGNS
TmLac	394 ESHGIDWD	440 NEAG	483 LMYPKMDILL	483 MCEYAHAMGNS
EcLac	567 FVWDWVDQ	599 DRQFCMNG	793 DIGVSEA	996 GDDSWS
TmLac	536 CIWDWVDQ	568 DGNFCING	759 DIGNRMP	955 GDDSWG

Figure 4.16: Alignment of β-galactosidases from *T. maritima* (TmLac) and *E. coli* (EcLac) (Uniprot codes Q56307 and B8LFD6, respectively). A coloured line on top of TmLac sequence indicates the extension of the catalytic domain (green), the β-sandwiches common to other GH2 enzymes (blue, orange, purple, red) and the additional β-sandwich found in TmLac (grey).

4.5.2 Crystallographic analysis

Crystallographic resolution of TmLac was addressed by the group of Dr. Julia Sanz Aparicio from the "Instituto de Química Física Rocasolano" (CSIC, Madrid). Initial attempts to diffract TmLac crystals, gave a fairly good resolution around the active site region. However, the resolution was poor at the loops that interconect β-sandwich domains, possibly as consequence of the high flexibility of these regions. A possible solution in some instances is the introduction of point mutations that reduce the number of conformations of the loops. Using the Surface Entropy Reduction prediction (SERp) server (Goldschmidt et al., 2007) three different groups of mutations were suggested: 1) E242A/E243A/E244A; 2) E697A/K698A/K699A; 3) K1029A/E1030A/K1031A, located at different regions, all of them distant from the catalytic pocket (Figure 4.17). Thermal stability of each mutant was analyzed by SDS-PAGE. Soluble protein remaining after heat treatment was similar in mutants and wild-type (Figure 4.18), indicating that mutations do not affect thermal stability of the protein. Nevertheless, after obtaining crystals with different crystallization patterns the structure with the best resolution (3.5 Å) was still about the same obtained with the wild-type TmLac (Figure 4.19 A). Diffraction spectra obtained so far do not allow a detailed resolution of the secondary and tertiary structure. However, they are good enough to provide a picture of the overall octameric quaternary structure of the protein (Figure 4.19 B).

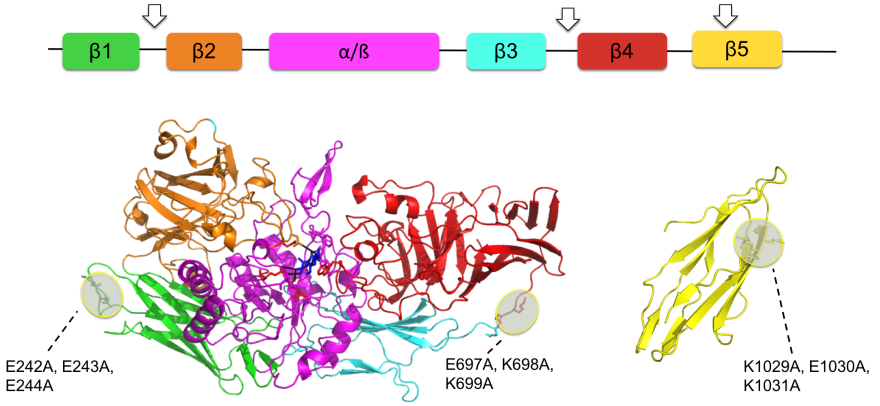


Figure 4.17: Locations of the mutations carried out to reduce the surface entropy in the TmLac model. The relative position of the groups of mutated residues on the protein domains is indicated by arrows in the scheme shown on the top and marked by yellow contoured areas in the structural model.

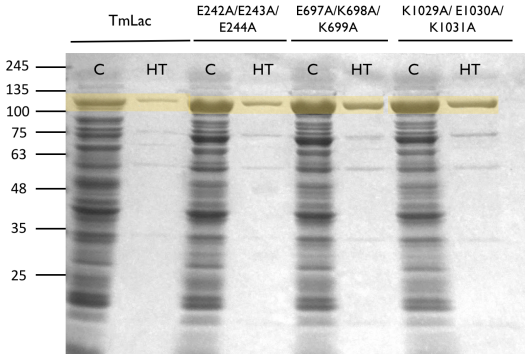


Figure 4.18: Soluble protein from *E. coli* extracts expressing wild-type and mutant enzymes designed by the SERp server. SDS-PAGE and Coomassie blue staining of samples before (C) and after heat treatment at 85°C during 5 minutes (HT). Heterologous gene expression was induced at 37°C for 3 hours in all cases. Bands corresponding to full-length recombinant proteins are marked in yellow.

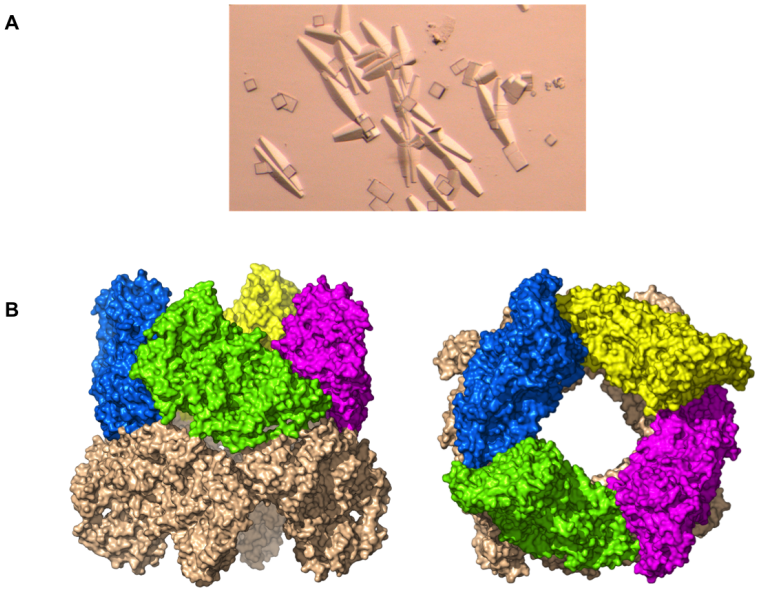


Figure 4.19: Crystallographic study of TmLac. **(A)** Crystal of wild-type TmLac. **(B)** Two views of the TmLac octamer in solvent accessible surface at 3.5 Å resolution.

4.6 Improvement of the tranglycosylating activity of TmLac

4.6.1 Analysis of GOS synthesis by wild-type TmLac

Synthesis of GOS by TmLac was assayed using 25% of lactose as substrate since high lactose concentration favour the transglycosylation reaction (Gänzle, 2012). The performance of both free and covalently bound enzyme to an epoxy activated resin was compared. As a consequence of the retaining mechanism of the enzyme (Figure 1.4), both hydrolysis and transgalactosylation of lactose proceed through a covalent intermediate of the enzyme linked to the galactosyl moiety. In both reactions a glucose molecule is released in the

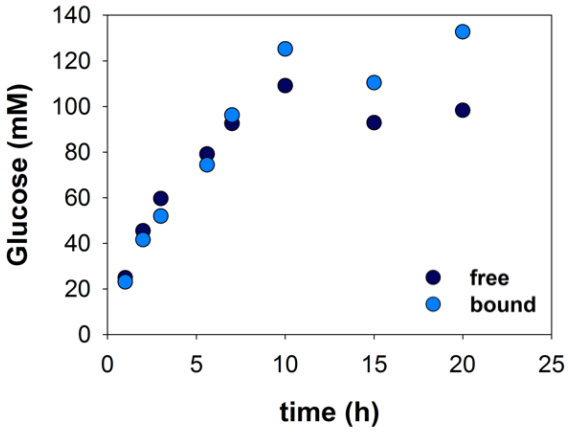


Figure 4.20: Glucose release by free and covalently bound TmLac (3 μg/ml). The enzyme was incubated with 25% lactose at 75°C with continuous shaking.

first step. Therefore, the kinetic of glucose production can be taken as an estimation of the total activity (hydrolysis plus transglycosylation) of the enzyme. The rate of glucose release was similar for free and bound enzyme forms, reaching saturation at a concentration around 110 mM, which represents the consumption of around 16% of the initial lactose present in the medium (Figure 4.20). This enzymatic arrest is likely due to product inhibition as it has been described for other β-galactosidases (Boon et al., 1999; Vera et al., 2011).

Reaction products of free and bound TmLac obtained after 5 hours of reaction were compared after analysis by anion exchange chromatography. Both enzyme forms yielded the same elution profile with two major transglycosylation products with retention times of 5.45 and 7.23 min. These GOS were identified as β-6'-galactosyl-lactose and β-3'-galactosyl-lactose, respectively, by comparison with retention times of commercial standards (Carbosynth, UK). The chemical structure of these trisaccharides was confirmed by NMR carried out

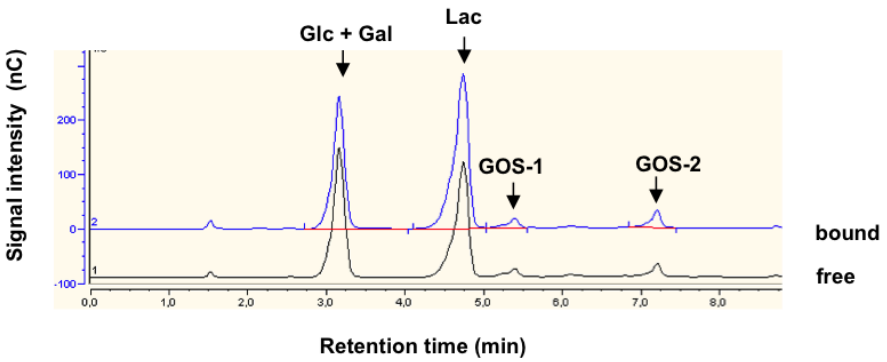


Figure 4.21: Chromatogram of the products obtained after 5 hours incubation of lactose (25%) with free or covalently bound enzyme. Peaks corresponding to glucose and galactose (Glc + Gal), lactose (Lac), and two galacto-oligosaccharides (GOS-1: β-6'-galactosyl-lactose; GOS-2: β-3'-galactosyl-lactose) are indicated.

by the group of Prof. FJ Cañada from "Centro de Investigaciones Biológicas, CIB-CSIC". Quantification of chromatographic peak areas revealed that wild-type enzyme produced β-3'-galactosyl-lactose and β-6'-galactosyl-lactose with a molar ratio of 3:1 (Figure 4.21).

4.6.2 Mutant design

A set of mutations directed to the active site of TmLac was designed with the aim of increasing GOS yield, changing the profile of transglycosylation products or exploring the structural basis of the intrinsic transglycosylating properties of the enzyme. Rational design was based on a structural model of TmLac docked with a molecule of β-3'-galactosyl-lactose, the major transglycosylation product of the wild-type enzyme (Subsection 4.6.1). This complex (Figure 4.22) can be considered an analogue of the reaction intermediate, where the enzyme is covalently linked to a galactosyl moiety and a new lactose molecule

binds to the active site to act as the galactosyl acceptor.

The docking analysis showed a set of residues that may form hydrogen bonds with the galactosyl moiety at the non-reducing end of β-3'-galactosyl-lactose, namely D191, H369, N440, E441, E507, H510, W538 and N574. These residues would conform the -1 subsite, using the nomenclature stablished by Davies et al., 1997 and represent the binding scaffold of the galactosyl moiety covalently bound to the enzyme in the reaction intermediate (Figure 4.22). Most of these residues, considered to be essential for enzyme function, are highly conserved among GH2 β-galactosidases. E507 and E441 (homologous to EcLac E537 and E461), would be the putative catalytic pair (nucleophile and acid/base respectively) (Juers et al., 2001). In EcLac, D191 is involved in Na^+/K^+ binding (described as β-galactosidase activators), H369, N440 and H510 stabilize the transition-state, and W538 binds lactose in the so-called deep mode (Roth and Huber, 1996; Huber et al., 2001; Juers et al., 2001; Xu et al., 2004; Juers et al., 2012; Wheatley et al., 2012). The substitution of many of these residues in EcLac rendered an inactive enzyme and therefore they were discarded as candidates for mutagenesis. Only residue N574 showed some sequence divergence among GH2 enzymes with Asp at an equivalent position in some cases (Figure 4.22 B). Substitution N574D was designed to study whether this change had a functional role. Mutants N574A and N574S were generated in order to disrupt the putative hydrogen bond between the galactosyl moiety covalently bound to the enzyme in the reaction intermediate and the N574 residue. The rationale behind this was to increase the rotational freedom of this group facilitating the coupling to the lactose molecule that acts as galactosyl acceptor.

Subsites +1 and +2, that bind the galactosyl and glucosyl moiety of the acceptor lactose, respectively, would involve fewer residues than subsite -1

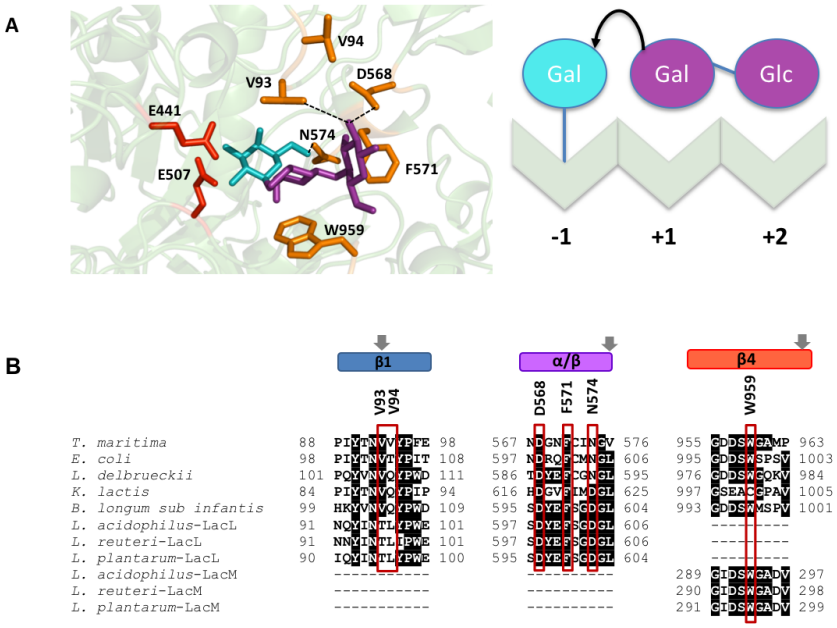


Figure 4.22: Design of TmLac mutants. **A**. Left panel. Detail of TmLac active site highlighting the catalytic residues (red) and the residues targeted for site-directed mutagenesis (orange). β-3'-galactosyl-lactose is depicted with the galactosyl moiety at the non-reducing end in blue. The rest of the molecule, coloured in violet, represents the lactosyl moiety acting as acceptor of the galactosyl group during transgalactosylation. Right panel. Schematic representation of the reaction intermediate in transglycosylation reaction. The galactosyl moiety covalently linked to the -1 subsite is depicted in blue. The acceptor lactose molecule bound to +1/+2 is coloured in violet. **B**. Sequence alignment of TmLac with other GH2 β-galactosidases with transgalactosylating activity. Sequences around the residues targeted for site-directed mutagenesis are shown. The relative position of those sequences within specific TmLac domains is schematically represented above. Colour codes for the different domains are the same as those indicated in Figure 4.15. Genebank codes for the proteins are: *T. maritima*: AAD36268.1, *E. coli*: AAA24053.1, *L. delbrueckii*: CAI98003.1, *K. lactis*: AAA35265.1, *B. longum_sub_infantis*: AAL02052.1, *L. acidophilus*_LacL: ABK59934.1, *L. acidophilus*_LacM: ABK59935.1, *L. reuteri*_LacL: ABF72116.1, *L. reuterii*_LacM: ABF72117.2, *L. plantarum*_LacL: CAZ66936.1, *L. plantarum*_LacM: CAD65570.1.

and should not be as critical for enzyme activity. W959 and N574 are in contact to both galactosyl and glucosyl groups of the lactose acceptor. The W959 residue is conserved in bicistronic β-galactosidases, despite the fact that this residue is located in a different subunit. In the case of *Kluyveromyces lactis* β-galactosidase the equivalent position is occupied by a cysteine residue (Figure 4.22 B). Subsite +2 may be conformed by F571 and D568 residues. The functionality of W959, F571 and D568 in the transglycosylation reaction was tested with the mutants W569C, W959F, W959A, F571L, D568A and D568S. Finally, V93 and V94 residues, located close (around 4 and 7 Å respectively) to the terminal glucose moiety were mutated to Thr or Gln in order to mimic β-galactosidases with high transgalactosylating activity (Figure 4.22 B). These substitutions may stabilize the binding of a more extended galactosyl acceptor, for example a trisaccharide, to yield a tetrasaccharide. Mutants V93T, V94T and V94Q were generated to test this hypothesis.

4.6.3 Analysis of mutant enzymes

Production of mutant enzymes by *E. coli* compared to the wild-type was evaluated by SDS-PAGE analysis of soluble extracts of enzyme-producing transformant clones. As shown in Figure 4.23, TmLac mutants had expression patterns similar to wild-type enzyme, except for N574S and W959F which were produced in significantly lower amounts. However, none of the introduced mutations seem to change the thermal stability of the enzyme, since the fraction of soluble enzyme recovered after heat treatment is similar to that obtained with the wild-type Figure 4.23.

As previously mentioned, since glucose is released both for hydrolysis and transgalactosylation, initial velocity of glucose production can be used to

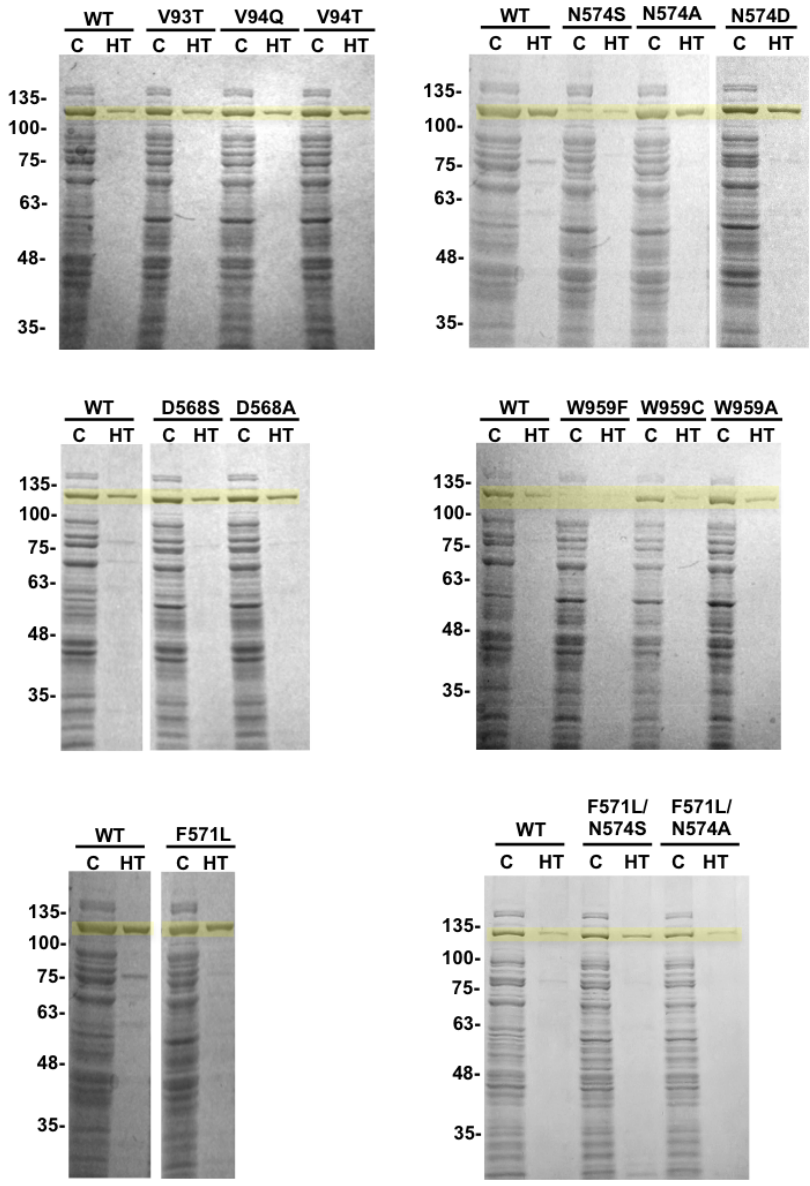


Figure 4.23: Expression of TmLac mutants in *E. coli*. SDS-PAGE analysis of crude soluble extracts (C) and soluble extract after heat treatment (HT) at 85°C for 5 minutes. Wild-type and mutant TmLac versions are marked in yellow.

Table 4.2: Total activity of the different. Relative values to wild-type enzyme are shown in parenthesis (* $p < 0.01$).

Enzyme	Total activity (μmol Glc $\cdot$ min^{-1} $\cdot\mu$g protein^{-1})
wild-type	4.4 ± 0.9 (100%)
V93T	4.15 ± 0.08 (94%)
V94T	3.8 ± 0.3 (86%)
V94Q	4.3 ± 0.7 (98%)
D568S	4.8 ± 0.2 (110%)
D568A	5.17 ± 0.17 (120%)
F571L	3.1 ± 0.2 (70%)
N574A	1.4 ± 0.3 * (31%)
N574S	0.36 ± 0.06 * (8%)
N574D	4.3 ± 0.3 (97%)
W959F	0.22 ± 0.06 * (5%)
W959C	0.94 ± 0.05 * (21%)
W959A	2.0 ± 0.2 * (45%)
F571L/N574S	0.41 ± 0.4 * (9%)
F571L/N574A	0.30 ± 0.3 * (7%)

calculate the total activity (hydrolysis plus transglycosylation) of the enzyme (Table 4.2). Transformants with TmLac substitutions N574A and N574S showed a drastic reduction of activity (down to 31% and 8% of the wild-type, respectively) whereas N574D substitution had no significant effect on total activity. The reduction of activity for N574S can be correlated with the decreased enzyme synthesis (Figure 4.23). The lower activity observed for W959F (5% of the wild-type) may be also explained by decreased expression (Figure 4.23). However, substitution of W959 by non-aromatic residues (W959C and W959A) caused a significant reduction of the total activity (to 21% and 45% of the wild-type, respectively) despite the fact that the enzyme

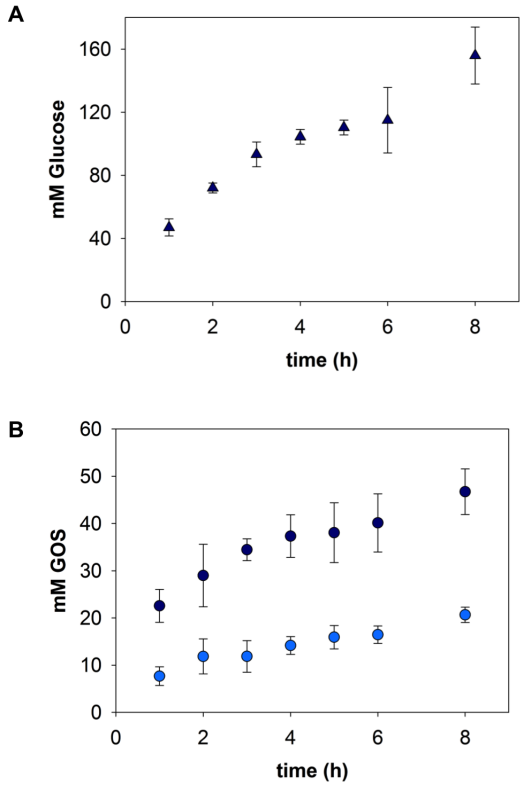


Figure 4.24: Kinetics of wild-type TmLac. **A**. Glucose release. **B**. Galacto-oligosaccharide production. Symbols: closed triangles: glucose; light blue dots: β-6 '-galactosyl-lactose; dark blue dots: β-3 '-galactosyl-lactose. *Error bars* indicate standard deviation of triplicates.

amount was not altered (Figure 4.23). Total activity of the rest of mutants (V93T, V94T, V94Q, D568S, D568A, F571L) was similar to that of the wild-type. Therefore, only mutations W959C and W959A caused a significant decrease in TmLac activity per unit of enzyme mass.

Kinetics of β-3'-galactosyl-lactose, β-6'-galactosyl-lactose production, and glucose release were analyzed for the wild-type enzyme (Figure 4.24). Glucose release kinetics revealed that TmLac is progressively inactivated. After 3 hours of reaction, the rate of glucose production was about 20% of the initial velocity,

probably as a consequence of enzyme inhibition by product accumulation, as reported for other β-galactosidases (Boon et al., 1999; Vera et al., 2011). GOS synthesis also slowed down as the reaction progressed, with a lower rate after 3 hours of reaction for both β-3'-galactosyl-lactose and β-6'-galactosyl-lactose (Figure 4.24 B).

To compare GOS yield between different enzyme variants and the wild-type, the same amount of enzyme (activity units) was used (calculated as initial glucose release per unit of time) and the reactions were carried out for 5 hours. Kinetics of glucose release was monitored in parallel as a control (Figure 4.25 and 4.26).

None of the enzymatic versions synthesized novel GOS types compared to the wild-type. Mutations N574S and N574A, but not N574D, yielded an increase in the synthesis of β-3'-galactosyl-lactose, around 30% and 40%, respectively, whereas β-6'-galactosyl-lactose synthesis was not affected significantly (Figure 4.25 A). Similary, β-3'-galactosyl-lactose synthesis was increased around 40% with mutant F571L (Figure 4.25 B). Substitutions at residues D568, V93 and V94 did not affect transglycosylation efficiency nor the product profile of TmLac (Figure 4.26A and 4.26 C). Substitution W959F did not cause a significant change in the transglycosylating properties of the enzyme either. However, mutations W959C and W959A reduced drastically (around 80%) the synthesis of β-3'-galactosyl-lactose but not so remarkably that of β-6'-galactosyl-lactose (Figure 4.26 B). Consequently, the ratio of β-3'-galactosyl-lactose/β-6'-galactosyl-lactose which is 2.6 for the wild-type enzyme, was reduced to 1 and 0.7, for W959C and W959A, respectively.

Substitutions with higher GOS production were combined, generating mutants F571L/N574S and F571L/N574A. The resulting enzymes showed a

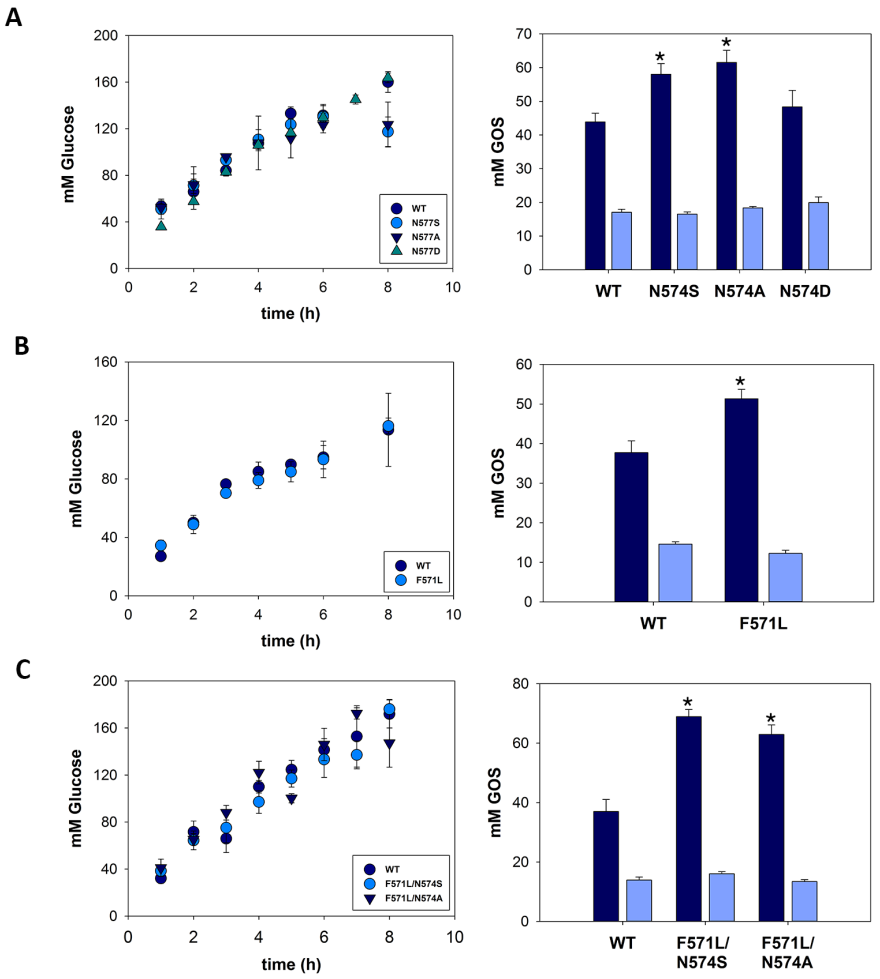


Figure 4.25: Effect of substitution of residues F571 and N574 on the synthesis of galacto-oligosaccharides (right panels): β-3'-galactosyl-lactose (dark blue bars) and β-6'-galactosyl-lactose (light blue bars). *Error bars* indicate standard deviation of triplicates. Asterisks indicate significant differences ($p < 0.01$) compared to wild-type data. Kinetics of glucose release is shown on the left panels as a control.

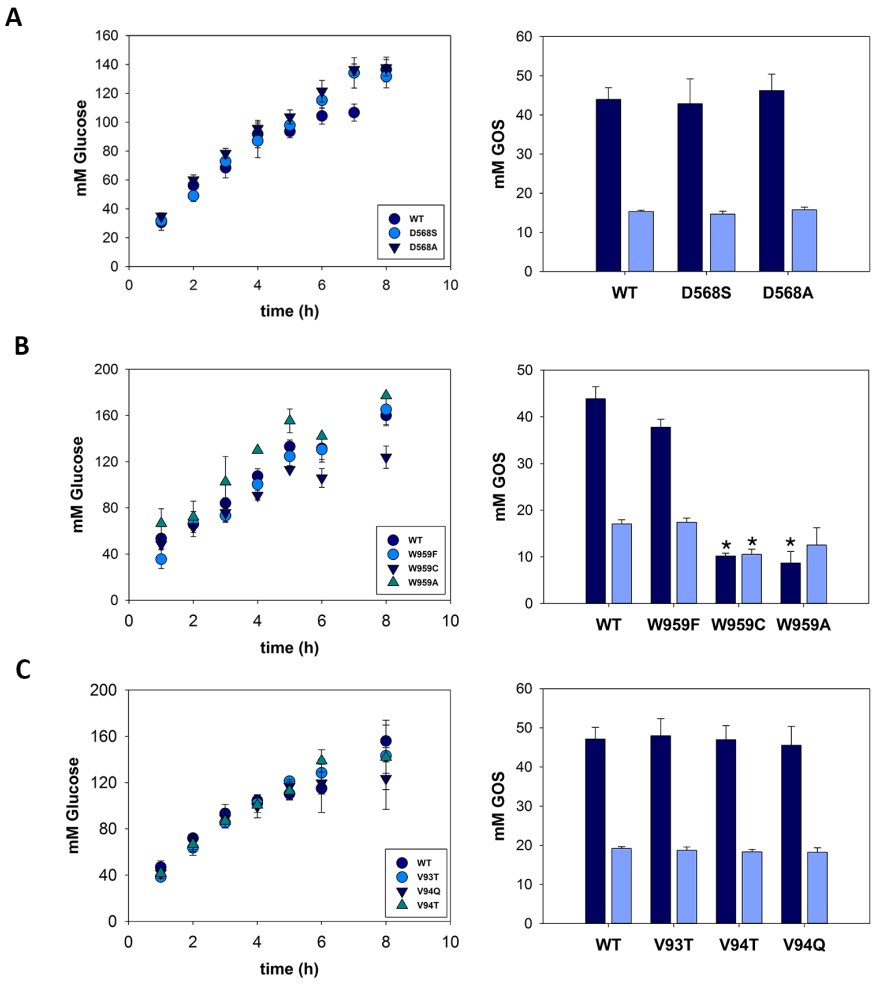


Figure 4.26: Effect of substitution of residues D568, W959, V93, and V94 on the synthesis of galacto-oligosaccharides (right panels): β-3'-galactosyl-lactose (dark blue bars) and β-6'-galactosyl-lactose (light blue bars). *Error bars* indicate standard deviation of triplicates. Asterisks indicate significant differences ($p < 0.01$) compared to wild-type data. Kinetics of glucose release is shown on the left panels as a control.

significant reduction in activity compared to the wild-type (Table 4.2). Both double-mutants yielded an increased synthesis of β-3'-galactosyl-lactose over the wild-type (1.9 and 1.7 fold respectively) representing also a significant increment over the single mutants F571L and N574S (Figure 4.25 C).

4.7 Phylogenetic analysis of GH2 enzymes

4.7.1 GH2 Domain Architectures (DAs)

A phylogenetic study of the GH2 family to which TmLac belongs, was carried out with proteins containing the canonical catalytic domain Glyco_hydro_2_C (GH2C) defined in Pfam database. Sequences in which this domain was absent related with enzymatic activities other than β-galactosidases (e.g. mannosidases and arabinosidases), were excluded. Also, hybrid enzymes with additional catalytic modules (e.g. glycosyl transferases, lipases, deacetylases, kinases, hexoaminidase, etc.) were not considered. All the selected sequences contain two β-sandwich domains at the N-terminal side of GH2C, identified in Pfam as Glyco_hydro_2_N (GH2N) and Glyco_hydro_2 (GH2) (Figure 4.27). These domains correspond to the N-terminal modules described by Juers et al., (1999) as characteristic of GH2 enzymes. In some cases, the GH2 domain is absent (DA type 3, Figure 4.27). However, these proteins contained instead a sequence of similar length to GH2 that likely represents an equivalent domain not defined in Pfam. Exceptionally, additional N-terminal domains related to carbohydrate recognition (Ricin B lectin domain) (Rutenber et al., 1987) or a cell surface adhesion signal (YSIRK signal) (Bae and Schneewind, 2003; DeDent et al., 2008; Tettelin et al., 2001) were found (not shown), suggesting that these sequences correspond to surface anchored

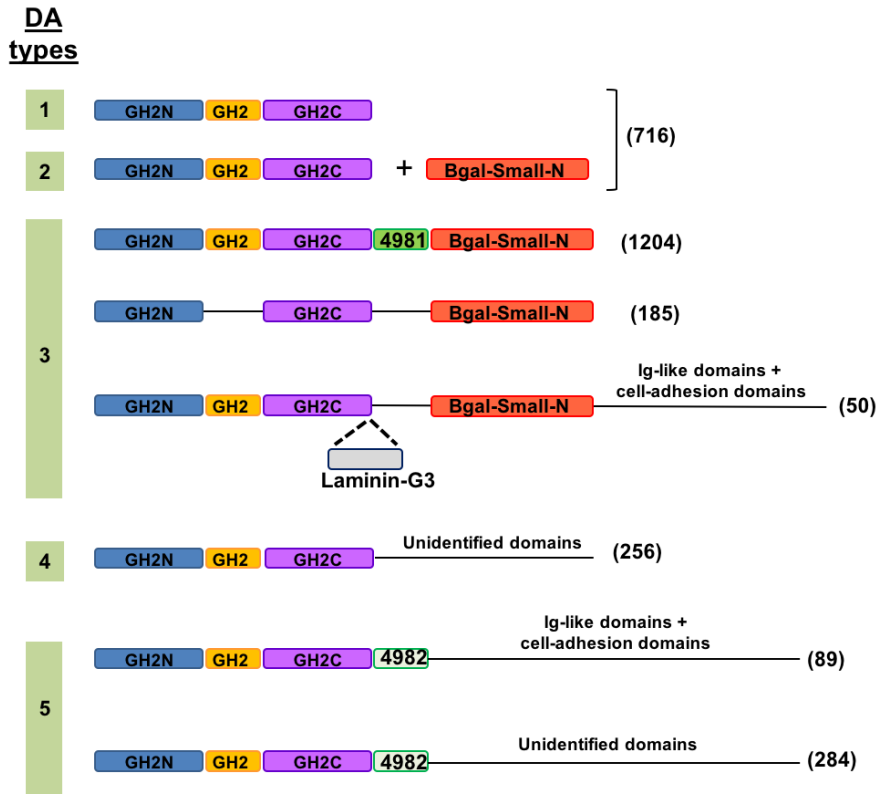


Figure 4.27: Domain architectures of GH2 family. Numbers in parentheses indicate the number of sequences representative of each DA type, included the Pfam database.

galactosidases. The classification of DAs rendered five groups (DA types) shown in Figure 4.27. One of the protein groups with higher number of representants (DA type 1 with ca. 700 sequences) showed a DA constituted by the characteristic core of β-galactosidases (GH2N+GH2+GH2C) without any additional domain. According to GenBank annotations, glucuronidases are included in this group. A variation of this DA is represented by the large subunit (LacL) of bicistronic β-galactosidases (DA type 2). These enzymes have a heterodimeric constitution, formed by the interaction of LacL with a smaller subunit LacM (78 entries in Pfam, labelled Bgal_Small_N).

The most frequent topology (DA type 3) contains the Bgal_Small_N domain at the C-terminal end of the protein. Three different subtypes can be discerned. In most cases (1204 sequences) Bgal_Small_N is linked to the GH2C through a β-sandwich domain, of around 100 amino acids, identified in Pfam as DUF4981 (Domain of Unknown Function). In other cases (185 sequences) either DUF4981 or GH2 or both were absent, being replaced by sequences of similar length. In a small set of sequences (50 entries) probably corresponding to secreted enzymes, there are additional domains appended. For instance, in some proteins a Laminin G3 domain, related with cell adhesion, is inserted between the GH2C and the DUF4981 domains. Some sequences contain modular domains characteristic of extracellular proteins, such as NPCBM or F5/F8, downstream the Bgal_Small_N domain (Rigden, 2005).

DA type 4 includes 256 proteins with a C-terminal (Ct) extension of 60-1053 residues without identified domains. We have classified these sequences according to the length and sequence similarity of the Ct extension (Table 4.3).

DA type 5 (373 entries) includes proteins with the GH2C followed by the DUF4982 domain and additional C-terminal extensions (Figure 4.27).

Table 4.3: Cluster and subcluster classification of DA type 4 proteins with unidentified C-terminal extensions.

C-terminal signature	Length	Number of sequences	Number of sub-clusters
Ct1	60-100	47	Ct1-1, Ct1-2, Ct1-3, Ct1-4
Ct2	100-150	30	Ct2-1, Ct2-2, Ct2-3, Ct2-4
Ct3	150-200	21	Ct3-1, Ct3-2, Ct3-3
Ct4	200-250	55	Ct4-1, Ct4-2, Ct4-3, Ct4-4, Ct4-5, Ct4-6, Ct4-7
Ct5	250-300	35	Ct5-1, Ct5-2, Ct5-3, Ct5-4
Ct6	300-350	15	Ct6-1, Ct6-2, Ct6-3
Ct7	350-400	19	Ct7-1, Ct7-2, Ct7-3
Ct8	400-450	21	Ct8-1, Ct8-2, Ct8-3, Ct8-4
Ct9	450-500	2	-
Ct10	500-550	2	-
Ct11	800-850	7	-

DUF4982 probably acts as a linker between GH2C and the C-terminal domains downstream, in a similar way that DUF4981 does in DA type 3 proteins. Most proteins of this group (284 sequences) contain C-terminal regions (Cter) of 100 to 550 amino acids with unidentified domains. We have classified them according to the C-terminal length and sequence identity (Table 4.4). We considered that DA type 5 proteins deserved special attention because of the high sequence variability at their C-terminal extension. Therefore, detailed analysis of this group was carried out including domains described in Interpro (Mitchell et al., 2015) and not identified in Pfam. Results are shown in Figure 4.28. In a subset of these proteins sequences, different combinations of Ig-like domains mostly occurring in cell-wall anchored proteins and other modules involved in cell adhesion were detected. Many of these sequences show a BIG1-like domain downstream DUF4982.

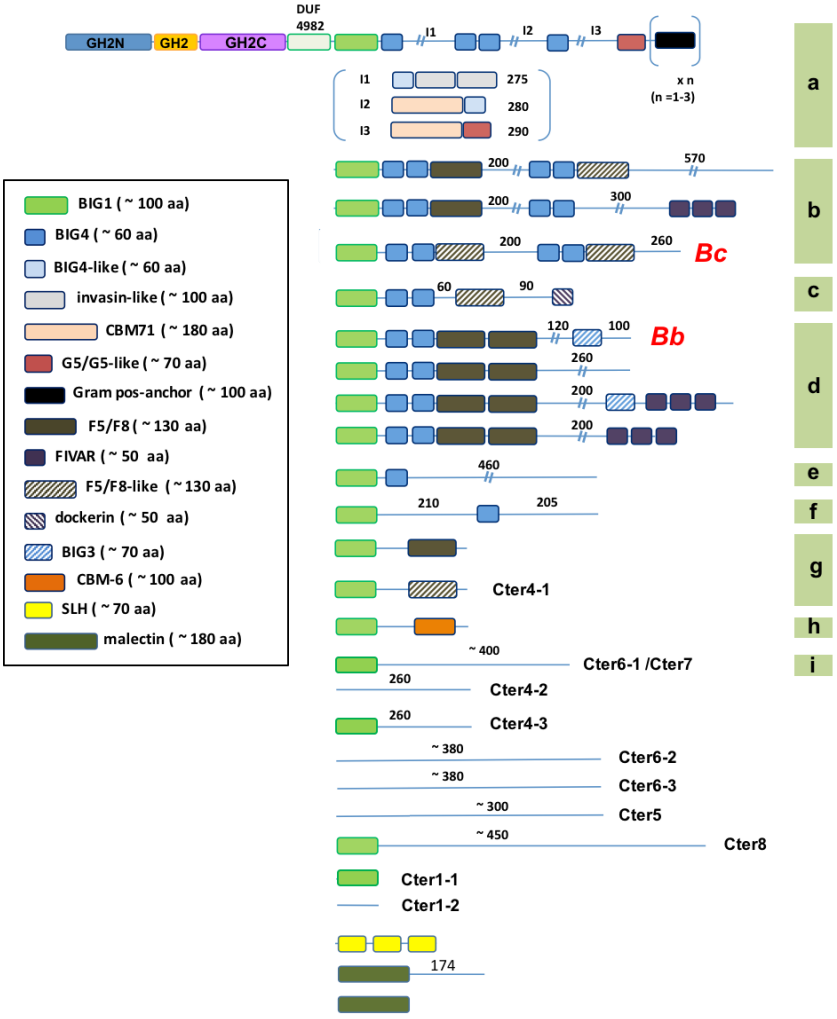


Figure 4.28: Domain architecture of DA type 5 sequences. Coloured and stripped boxes correspond to domains identified by Pfam and Interpro databases, respectively. Modules with more than 40% sequence identity compared to BIG1 domains identified by Interpro, with a coverage higher than 60%, were tagged as BIG1. Letters (a-i) on the right edge of the figure group sequences with similar DAs. Numbers on top of non-identified regions indicate approximate number of residues. The architectures corresponding to *B. circulans* ***(Bc)*** and *B. bifidum* ***(Bb)*** are indicated in red.

Table 4.4: Cluster and subcluster classification of DA type 5 proteins with unidentified C-terminal extensions.

C-terminal signature	Length	Number of sequences	Number of sub-clusters
Cter1	100-150	217	Cter1-1, Cter1-2
Cter2	150-200	2	Cter2-1, Cter2-2
Cter3	200-250	2	Cter3-1, Cter3-2
Cter4	250-300	11	Cter4-1, Cter4-2, Cter4-3
Cter5	300-350	4	-
Cter6	350-400	47	Cter6-1, Cter6-2, Cter6-3
Cter7	400-450	5	-
Cter8	500-550	6	-

4.7.2 Phylogenetic analysis of the GH2 catalytic domain

Figure 4.29 shows a phylogenetic tree constructed using the GH2C modules of representative sequences from each DA. All the sequences corresponding to canonical β-galactosidases (DA type 3) clustered in a single group. This cluster includes subtrees corresponding to bicistronic galactosidases classified as DA type 2. Interestingly, a subset of type 4 (labelled as Ct_8_3) also clusters within canonical galactosidases. A search in the Conserved Domain Database from NCBI (Marchler-Bauer et al., 2015) using the C-terminal region of these sequences as input identified a sequence stretch of around 50-100 residues homologous to the central region of Bgal_Small_N domain. The remaining type 4 sequences cluster in four different nodes. Glucuronidases (DA type 1) are dispersed in three different nodes. DA type 5 is divided in three blocks. Two of them cluster with DA type 4 whereas the other one, labelled 5* in Figure 4.29, is described below. The evolutionary analysis may offer interesting clues

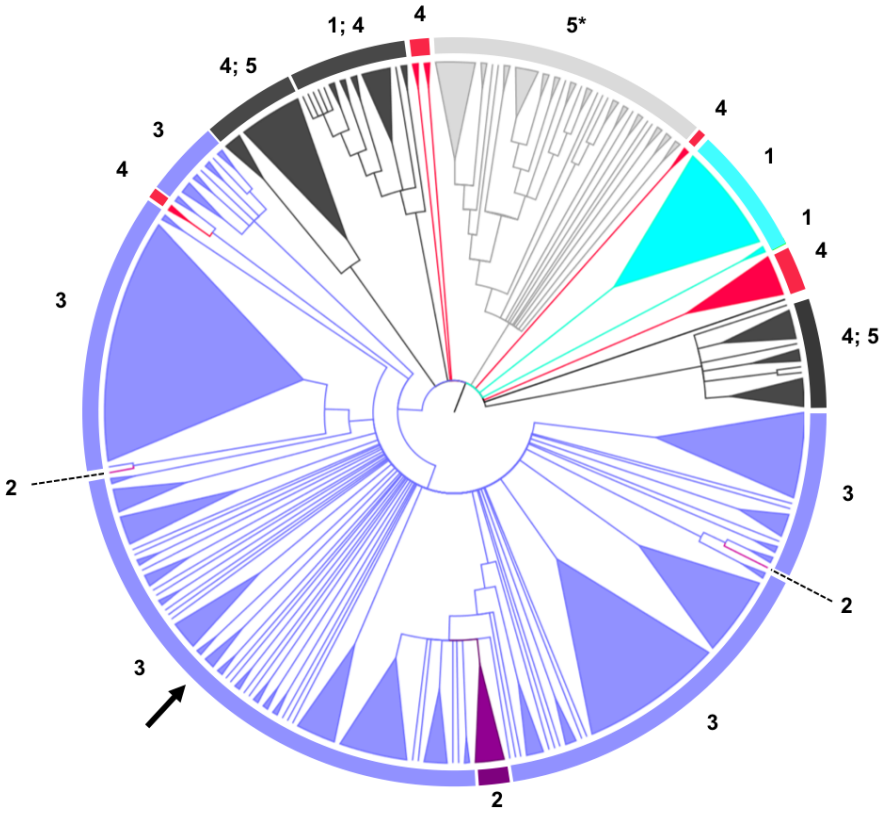


Figure 4.29: Phylogenetic analysis of the GH2C domain. The tree was calculated by Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992) condensed at $< 50\%$ bootstrap support. The analysis involved a selection of 380 amino acid sequences for the different DAs. Values indicate the DA type. The tree was drawn using FigTree software (http://tree.bio.ed.ac.uk/software/figtree/). The asterisk mark the subtree further analysed in Figure 4.30. Arrow indicates the position of the TmLac. Clusters are coloured by DA, dark grey indicate clusters containing more than one DA.

about uncharacterized enzymes with potential industrial value. For instance, enzymes with DA types 2 and 3 show high transglycosylating activity being able to synyhetise β-(1,3) and β-(1,6) GOS (Crittenden and Playne, 1996). Commercial synthesis of β-(1,4) GOS is carried out with a β-galactosidase from *Bacillus circulans* (Yanahira et al., 1995). A related β-galactosidase from *Bifidobacterium bifidum* also shows high transglycosylating efficiency (Goulas et al., 2009). Both enzymes belong to DA type 5. They carry a BIG1-like domain downstream the DUF4982 module (Figure 4.28) and their catalytic domains cluster under the same node (labelled 5* in Figure 4.29 and shown in more detail in Figure 4.30).

In order to relate the transglycosylating capability of industrially interesting enzymes with their phylogenetic position, a detailed study of DA type 5 was carried out. Amino acid residues within the catalytic pocket, potentially involved in transglycosylation, were identified by a docking study of the trisaccharide β-(1,4)-galactosyl-lactose into the *B. circulans* β-galactosidase structure (Figure 4.31). According to this analysis, one of these residues is located within the GH2N domain, whereas the other 13 reside in the GH2C domain. These include the catalytic triad formed by Glu 447 (acid base), Glu 532 (nucleophile) and Tyr 511 (Ishikawa et al., 2015). Figure 4.32 shows a sequence alignment of DA type 5* enzymes. Residues potentially involved in transglycosylation (marked by purple squares) are largely conserved among sequences containing the BIG1 domain. Exceptions are Tyr 449, which is poorly conserved and Asp 481 that changes to Asn in many instances, including the β-galactosidase from *B. bifidum*. Some DA type 5 β-galactosidases lacking the BIG1 domain, labelled Cter6-2, also show high identity at residues potentially involved in transglycosylation.

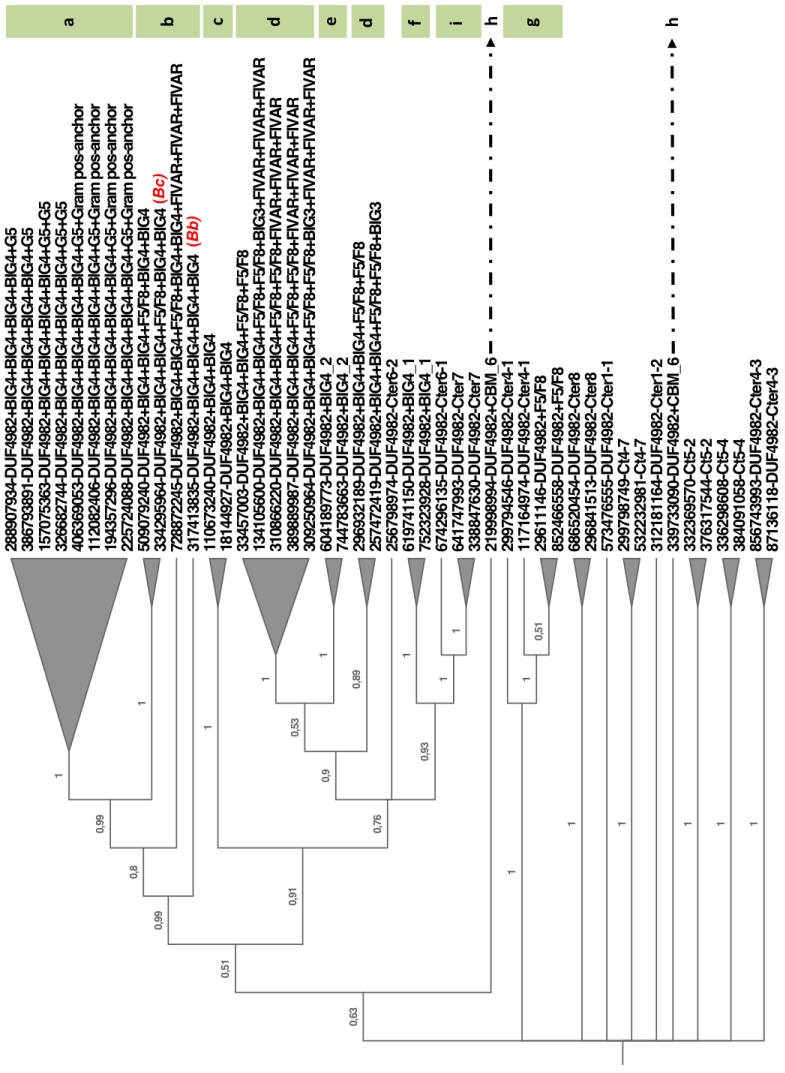


Figure 4.30: Phylogenetic subtree corresponding to the region marked 5* in Figure 4.29. Letters on the right edge of the figure group sequences with similar DAs in Figure 4.28. Branch numbers indicate bootstrap values. Bc and Bf correspond to *Bacillus circulans* and *Bifidobacterium bifidum* β-galactosidases respectively.

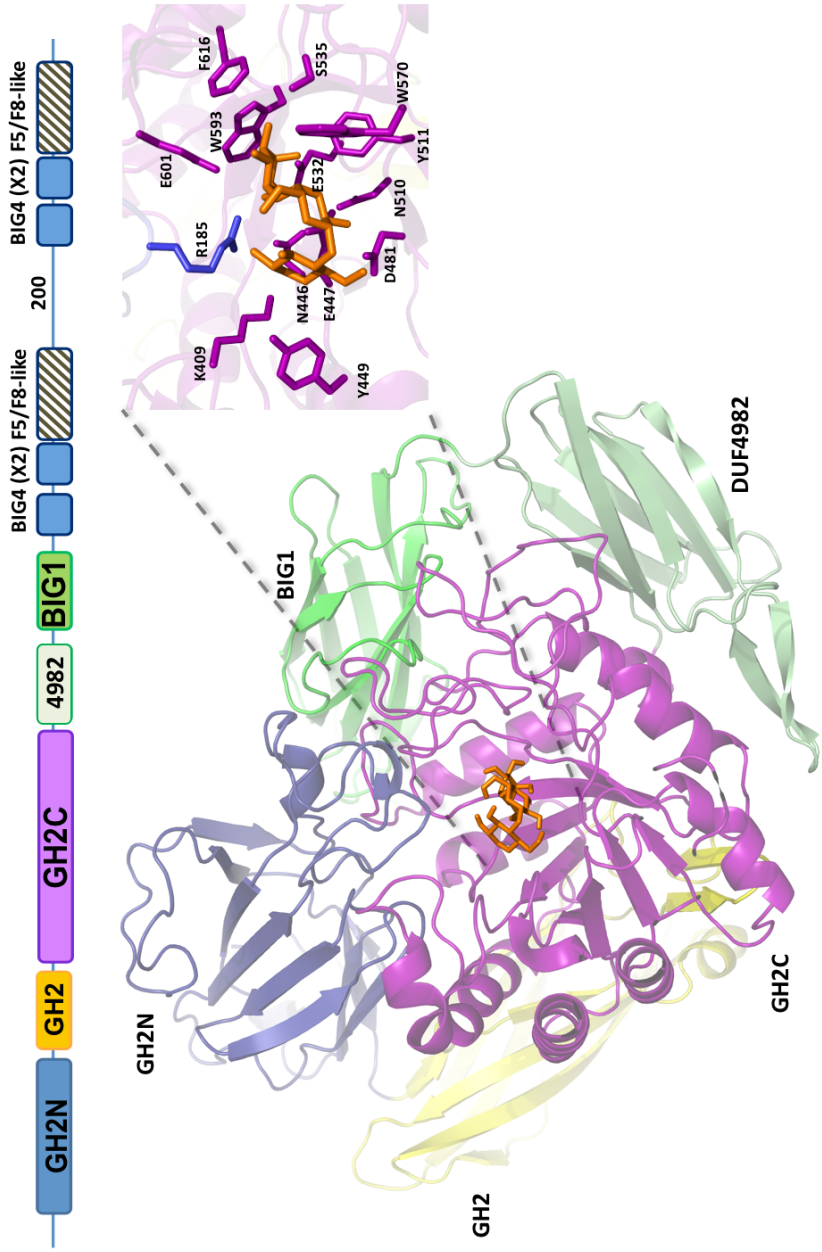


Figure 4.31: Docking of *Bacilus circulans* β-galactosidase (PDB: 4YPJ) with β-4'-galactosyl-lactose.

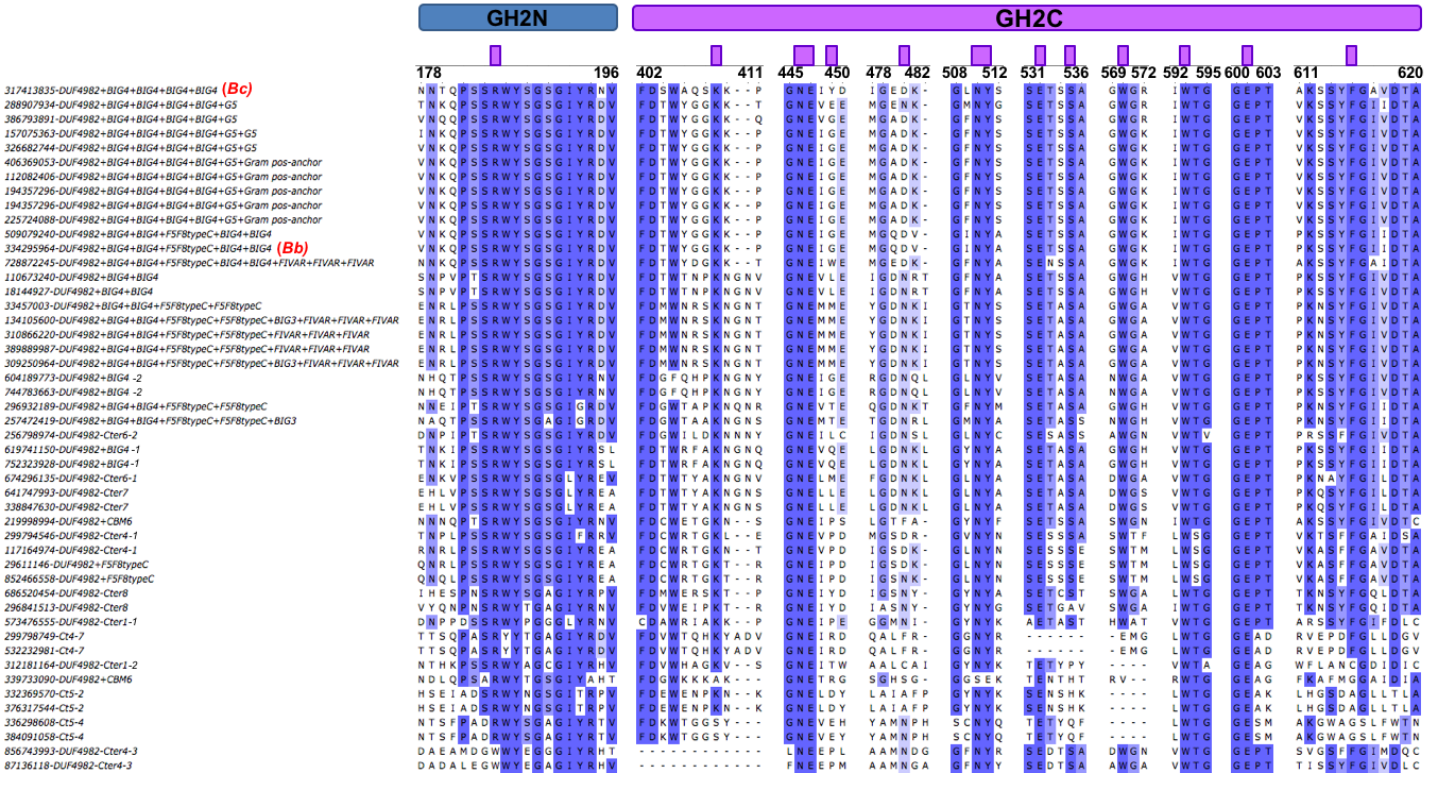


Figure 4.32: Sequence alignment around the putative catalytic site of proteins analyzed in Figure 4.28. Purple boxes indicate residues potentially involved in catalysis. Positions with more than 50% identity are coloured in light blue and those with 100% identity in dark blue colour. Bc and Bf correspond to *Bacillus circulans* and *Bifidobacterium bifidum* β-galactosidases, respectively.

Chapter 5

Discussion

Lactase is among the most relevant enzymes used in the food industry, mostly because its lactose-splitting hydrolytic activity, but also because of its transglycosylating capability which is becoming increasingly important. Currently used industrial lactases belong to *Kluyveromyces lactis, Aspergillus niger,* and *Bacillus circulans*, which work optimally under mesophilic conditions. There is a growing interest to develop new industrial enzymes that operate at cold temperatures (Nakagawa et al., 2006; Pawlak-Szukalska et al., 2014; Fan et al., 2015) or high temperatures (Akiyama et al., 2001; Placier et al., 2009; Torres et al., 2010). The β-galactosidase from the thermophilic bacteria *Thermotoga maritima* was selected for this project to offer an alternative method for lactose hydrolysis and to improve its potential for GOS synthesis. The thermostability of the enzyme is a great advantage to purify the protein overproduced in mesophilic hosts. According to our results, a simple heat treatment allows the purification of the enzyme produced in *E. coli* to more than 95% of the total protein (Figure 4.2). This represents a significant reduction in the time and cost of purification for industrial applications.

The main results achieved in this project are discussed below.

5.1 Immobilization

Enzyme immobilization is a useful methodology for industrial applications since it allows enzyme recycling and working in continuous systems. Stable attachment of the enzyme to a support also prevents its release to the final product. This is particularly interesting for food applications, since current market tendencies favour products without any type of additives. In the case of lactose-free milk our results show that different Spanish commercial brands include significant amounts of lactase (Figure 4.12). Another positive

aspect of immobilization, specifically when multi-point attachment occurs, is associated to an increased stability. These features have been studied using two surface binding methods, covalent cross-linking to epoxy-activated beads or non-covalent adhesion based on hybrid proteins.

5.1.1 Covalent immobilization

Effective covalent immobilization on epoxy-activated resins is highly dependent on the previous adsorption of the enzyme to the support, to bring reactive groups of the protein (mainly lysine residues) in close contact to epoxy residues (Mateo et al., 2007a). Efficient immobilization normally requires salt addition at moderate concentrations. This effect varies, depending on the protein material and the epoxy resin (e.g. polystyrene, methacrylate, polyvinyl alcohol, polyacrylamide or silica) (Wheatley and Schmidt Jr, 1999; Mateo et al., 2000; Torres et al., 2003; Bozhinova et al., 2004). In the case of resins of hydrophobic nature (e.g. polystyrene, methacrylate), the effect of the salt has been attributed to salt-induced hydrophobic interactions between the protein and the resin, that facilitate the initial adsorption. The support used in this work was polyvinyl alcohol (PVA) activated with epoxy groups (M-PVA-E02 from Chemagen). Although PVA is generally considered as a hydrophilic polymer, published results and our own experience indicate that addition of salt at high concentrations yields a more efficient protein immobilization. A possible explanation is that the spacer arms used for the functionalization of the support partially mask the hydrophilic nature of the resin, and may also serve as hydrophobic adhesion points for the protein upon salt addition. Another factor which probably determines immobilization on these supports is the hydrophobicity of the protein. The tendency of TmLac to aggregate at high salt con-

centrations (Figure 4.3) suggests the presence of extended hydrophobic patches on its surface, that may be involved in salt-induced adhesion to the support. The pH used during the incubation process with the support is another factor that may determine binding efficiency. This is because the nucleophilic character of the lysine residues is enhanced at pH values above their pK_a, at which they are mostly in their deprotonated form. TmLac has the advantage of being stable after incubation at high pH values up to 11, which allows the use of high pH values for binding. Under our experimental conditions, 20 mg of the enzyme were bound per gram of resin. This is probably below the loading capacity of the support since no enzyme was detected in the unbound fraction after incubation. Another key aspect for enzyme immobilization is that the bound enzyme may show different properties, related with activity or stability, compared with the free form. Immobilization may cause changes in the catalytic properties due to conformational changes forced by the covalent binding or by diffusion limitations for the enzyme, the substrate or the products. Covalent immobilization of *K. lactis* β-galactosidase to glutaraldehyde modified supports, such as graphite or silica, resulted in reduced activity (Zhou and Dong Chen, 2001) or reduced affinity for the substrate (Verma et al., 2012). In some cases immobilization of the same enzyme on different supports showed contrasting effects. *A. niger* β-galactosidase was completely inactivated when bound to Eupergit C (epoxy-polyacylamide) (Mateo et al., 2000) but retained more activity in epoxy-Sepabeads (polystyrene-polyvinylbenzene) showing that there is no fixed relationship between the funcional group and the enzyme, and other factors such as the composition of the support may influence the behaviour of immobilized enzyme. In the case of TmLac, initial activity of the enzyme bound to epoxy-PVA was increased between 1.4-2.8 fold compared to that of the free form, with *p*NP-β-Gal or lactose as substrates (Figure 4.5).

In contrast, binding to epoxy-polystyrene (Dynabeads) did not change TmLac activity (not shown). Despite of the increased initial activity of the enzyme bound to epoxy-PVA supports, the kinetic behaviour was similar to the free enzyme during long-term incubations (Figure 4.13). This could be explained by differences in product inhibition or by an inefficient agitation system. The transglycosylation efficiency of the enzyme was not altered either upon covalent binding (Figure 4.21).

Enzyme stabilization upon covalent immobilization has been extensively reported (Mateo et al., 2000; Cowan and Fernandez-Lafuente, 2011). Multipoint attachment due to covalent immobilization rigidifies the structure and contributes to stabilize a particular conformation. In the case of TmLac, after SDS treatment of the immobilized enzyme, 60% of the bound enzyme was released (Figure 4.4). Since, according to crystallographic data (Figure 4.19) TmLac shows an octameric structure, this result would be in agreement with an average of 4.8 monomers ($= 0.6 \cdot 8$) per holoenzyme not covalently attached to the support. Consequently TmLac is covalently attached through at least 3.2 linkages per holoenzyme. This may be explain the higher stability shown by the bound enzyme at low temperatures (Figure 4.6). The fact that the stability of the soluble TmLac at low temperatures is mainly compromised when the enzyme is diluted may reflect an unstabilization of the oligomeric structure in these conditions. This may be explained if hydrophobic interactions, which are weaker at low temperatures, make a significant contribution to the inter-subunit surfaces.

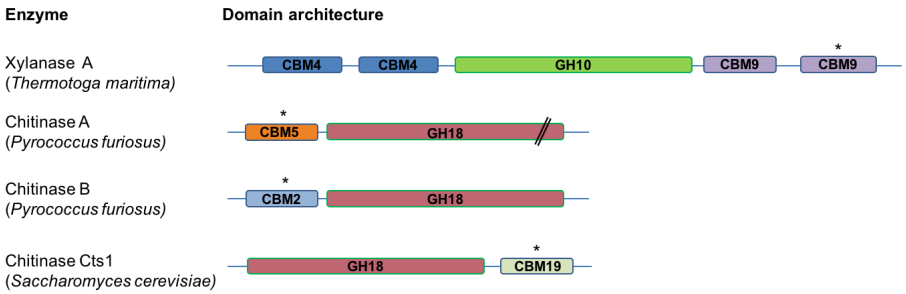


Figure 5.1: Domain architectures of the enzymes containing the CBMs used in the current study. Asterisks mark the specific CBMs used in this work.

5.1.2 CBM-mediated non-covalent immobilization

Cellulose and chitin are the most abundant natural biopolymers and therefore they are attractive candidates to be used as enzyme immobilization supports. Protein adsorption through specific affinity interactions is an interesting strategy to achieve purification and immobilization in a single step. In this study we have analyzed CBMs from four different families to test their binding affinity and specificity to cellulose or chitin, and the thermostability of the hybrids generated by fusion with TmLac. The selected CBMs belong to enzymes with xylanase or chitinase activities (Figure 5.1). However, in nature, these CBMs are also associated to other activities in some cases, as shown in Table 5.1. CBM9 is found in xylanases, α-amylases and in a few number of cellulases. CBM2 is spread among xylanases, chitinases and cellulases. In contrast, CBM5 and CBM19 have been found associated to chitinases only.

CBM5 and CBM2 selected for this study correspond to the CBMs of the Chitinase A (ChiA) and Chitinase B (ChiB) from *Pyrococcus furiosus*, respectively. They are located in two continuous ORFs but, the presence of a STOP codon in the catalytic domain of ChiA yields a truncated, inactive

Table 5.1: Activities associated to the CBMs used in this work. [a] The number of UniProt entries containing each combination of catalytic module and CBM is indicated in brackets.

Domain family	**Associated activities** [a]
CBM9	Uncharacterized proteins [2.153], GH10-like (xylanases)[171], histidin kinase [95], GH57 (α-amylase) [58] , GH (Glycosyl Hydrolases uncharacterized) [52], GH13 (α-amylase) [43], GH97 (glucoamylase, α-glucosidase) [35], α/β hydrolase fold (esterase) [18], GH5 (cellulase) [8]
CBM2	Uncharacterized proteins [793], GH18 (chitinases) [618], GH5 (cellulase, mananase, xylanase) [583], six hairpin glycosidase (cellobiosidase/endoglucanase) [331], concavalin lectin/glucanase A (xylanase) [199], GH10 (xylanase) [190], α/β hydrolase fold (esterases related with polysaccharide hydrolysis) [165], RlpA-like (glicosyltransferase) [75], SGNH hydrolase/esterase domain [85], GH (Glycosyl Hydrolases uncharacterized) [43], GH2 (β-glycosidase) [17]
CBM5	GH18 (chitinases) [1.765], chitin binding proteins [984], uncharacterized proteins [881], GH19 (chitinases) [530], peptidase [346], GH (Glycosyl Hydrolases uncharacterized) [295]
CBM19	Uncharacterized proteins [28], GH18 (chitinases) [163], GH (Glycosyl Hydrolases uncharacterized) [4]

chitinase. Still, CBM5 in this protein shows affinity for chitin (Mine et al., 2014). Interestingly, even though this CBM5 is mostly associated to chitinases (Table 5.1) a non thermophilic CBM5 from *Bacillus circulans* has been used to drive affinity purification to both, cellulose or chitin (Chong et al., 1997; Hong et al., 2008). Hybrid TmLac_CBM5 did not show a significant affinity for chitin or cellulose under our experimental conditions. This may be an intrinsic property of this particular CBM but we cannot rule out that the interaction with TmLac somehow hinders polysaccharide binding.

The hybrid of TmLac with CBM19 from *S.cerevisiae* Cts1 chitinase was discarded because of its low thermostability. Although this CBM derives from a mesophilic microorganism, we wanted to test if its stability was increased within the hybrid. However, heating at 80°C caused a massive precipitation of the hybrid protein, probably triggered by the denaturation of the C-terminal CBM19. Nonetheless, hybrids with CBM19 may be useful for the immobilization of mesophilic enzymes. Indeed, the TmLac_CBM19 fusion showed some affinity for chitin at pH 8.5, although binding to cellulose was less efficient.

Hybrids constructed using either CBM2 or CBM9 showed similar affinity for cellulose but only CBM2 was able to bind chitin (Figure 4.10). This broader binding specificity of CBM2 is suggested by the natural occurrence of this module, which is associated to chitinases, cellulases and xylanases (Table 5.1). Although CBM9 association to cellulases is not very frequent (Table 5.1) the ability of this CBM to bind cellulose, as well as xylan, has been previously shown (Boraston et al., 2001). Moreover, carbohydrate binding modules belonging to the family 9 have been used to create new hybrid proteins with different applications. The fusion of the CBM9 from *Thermotoga maritima* Xyn10A to *Clostridium thermocellum* cellodextrin phosphorylase (CtCDP)

resulted in a enhanced specific activity of the enzyme towards cellulose (Ye et al., 2011). CBM9-tagging was also used to purify proteins by cellulose affinity chromatography (Kavoosi et al., 2007) or to immobilize antibodies to cellulose filters for pathogen biosensing (Hussack et al., 2009).

The differences in polysaccharide specificity of CBM9 and CBM2 are explained by the structural differences in the binding site of these CBMs (Figure 5.2). CBM2 is classified as a type A CBM, characterized by a planar interaction surface with carbohydrates. Linearly arranged aromatic residues (W274, W308 and W326) act as the binding platform of polysaccharides (Nakamura et al., 2008). Previous studies suggest that E279 may form a negatively charged patch responsible for a higher substrate specificity for chitin compared to that for cellulose (Nakamura et al., 2008; Hanazono et al., 2016). In this study we have observed that CBM2 affinity for cellulose increases at pH 6.5 compared to pH 8.5 (Figure 4.10). Interestingly, among protonable residues in CBM2, only E279 shows a predicted pKa with the pH range 6-9 (Table 5.2). This suggests that the negative charge of this residue at pH 8.5 may interfere with cellulose binding. Since carboxyl groups are generated during cellulose isolation and purification (Hirosawa et al., 2001) a possible explanation for the lower affinity of TmLac_CBM2 at pH 8.5 may be the electrostatic repulsion between E279 and the negative charge of cellulose fibres. Apart from its pH dependence, the planar architecture of the CBM2 binding site seems suitable to interact with both cellulose or chitin. In contrast, type C CBM9 seems to be more substrate restrictive (Figure 5.2 B). The pocket-shaped binding site allows the interaction with two consecutive moieties of the polysaccharide, being more adapted to bind terminal ends. As show in the Figure 5.2 B a cellobiose molecule is attached to the CBM9 through two confronting aromatic residues, W71 and W175, interacting with opposite faces of the disaccharide (Notenboom

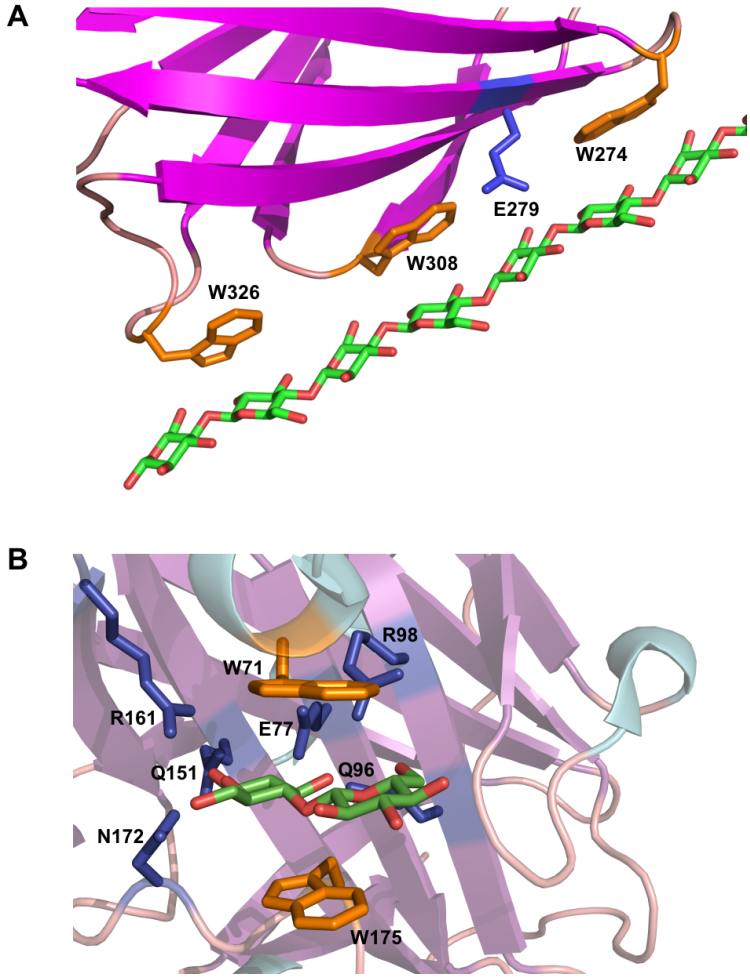


Figure 5.2: Binding surface of CBM2 and CBM9. **A** CBM2 structure from *Pyrococcus furiosus* (PDB 2CWR) with a cellulose fragment (green) docked manually. **B** *Thermotoga maritima* CBM9 structure from xylanase 10A in complex with cellobiose (green) (PDB 1I82). Aromatic residues are coloured in orange. Residues forming polar contacts with the corresponding substrates are coloured in blue.

Table 5.2: CBM2 and CBM9 residues with predicted pKa values between 6-9 calculated by PROPKA (Olsson et al., 2011).

	Residue	**pKa**	**Structural location**
CBM2	E279	6.0	Strand
CBM9	H85	7.3	Loop (Calcium binding site 1)
	E130	6.0	Strand (Calcium binding site 1)
	E26	6.6	Loop
	D74	8.7	Strand (Calcium binding site 2)
	D94	6.6	Strand (Calcium binding site 3)

et al., 2001). Additionally, a network of polar contacts involving multiple residues (E77, Q96, R98, Q151, R161, N172) further stabilizes binding and may explain the high specificity for cellulose as substrate vs. chitin observed in TmLac_CBM9. Indeed, the N-acetyl-glucosamine substitution present in chitin would preclude binding by imposing sterical impediments. In this case, pH dependence is also observed for cellulose binding (Figure 4.10). This may be also a result of protonation/deprotonation of critical residues in the binding site (e.g. E77, R98, R161) although the predicted pKa of these residues is out of the pH range 6-9. Alternatively, conformational changes at the binding site may arise from long-distance effects. Indeed, four residues with predicted pKa between 6 and 9 are involved in three different calcium binding sites (Table 5.2), with a stabilizing role (Notenboom et al., 2001).

The different behaviour of TmLac_CBM9 and TmLac_CBM2 in response to glucose and galactose addition regarding cellulose binding (Figure 4.11) may be explained as well by the architecture of their corresponding binding sites. Contact between the carbohydrate polymers and CBM2 occurs by π-C-H interactions between three linearly arranged tryptophan residues and

a polysaccharide stretch, involving at least five moieties (Figure 5.2 A), suggesting a low binding affinity for monosaccharides. In contrast, the small pocket-shaped binding site of the CBM9, with the two tryptophan residues arranged at opposite faces of substrate and including multiple polar contacts, makes the competition of glucose or galactose more likely, releasing the enzyme from the cellulose support.

Hybrid TmLac_CBM2 retained 100% of activity after binding to cellulose or chitin. However, reutilization assays for lactose hydrolysis in serial batches showed a significant decay of the activity after 7 hours of incubation (Figure 4.14). Further experiments are required to test whether this activity loss is due to enzyme inactivation or enzyme release from the support. In the latter case, immobilization may be further stabilized by covalent crosslinking. TmLac_CBM2 showed the best performance among the CBMs tested, with high thermal stability. The CBM2 module seems to be a promising tag to drive protein immobilization. CBM2 was able to bind both cellulose and chitin, providing a higher versatility. Moreover, binding to the support was firm, compared to that of CBM9. Possible applications include enzyme packing in a column to be used as continuous-flow systems (Rodriguez et al., 2004; Velikodvorskaya et al., 2010; Tsukimoto et al., 2010).

5.2 *In silico* structural analysis and mutant design

Increased GOS yield by genetic manipulation of β-galactosidases has been reported for β-glycosidases belonging to GH1 and GH42 families but not for GH2 enzymes. In the reported cases, mutations yielding successful results

were located within the -1 or +1 catalytic subsites (Placier et al., 2009; Wu et al., 2013). In this study, rational design of TmLac mutants was based on sequence similarity and docking analysis of a structural model with β-3'-galactosyl-lactose, the major GOS produced by TmLac (See section 4.6.1). This complex was assumed to mimic the structure of the transgalactosylating reaction intermediate, where the galactosyl moiety is covalently bound to E441 and a molecule of lactose is in the active site acting as acceptor of this group. Mutations were addressed to putative subsites -1 (N574), +1/+2 (F571, W959, D568) and +3 (V93, V94) (see Figure 4.22).

Mutations carried out at the -1 subsite, which may disrupt the hydrogen bond with the galactosyl moiety (N574S and N574A) reduced significantly the global activity of the enzyme (Table 4.2). However the transglycosylation capacity was increased between 30% and 40% with a specific increment of β-3'-galactosyl-lactose production (Figure 4.25 A). Disruption of the hydrogen bond may decrease the affinity of the enzyme for lactose in the first stage of the reaction, explaining the decrease of the global activity. On the other hand, the galactosyl moiety in the covalent intermediate may have higher rotational freedom in the mutant enzyme, favouring its transfer to the lactose molecule acting as acceptor in the second step of the reaction. Our results also suggest that this higher flexibility would specifically facilitate coupling to the acceptor through a β-(1,3) linkage. Mutation N574D did not cause any significant change in global activity (Table 4.2), enzyme accumulation (Figure 4.23) or GOS synthesis (Figure 4.25 A). This is in agreement with the structural analysis that predicts that the substitution of an amide group by a carboxy group with the same side chain length still allows the formation of a hydrogen bond with the galactosyl moiety (Figure 5.3). Our study also suggests that sequence divergence between Asn/Asp residues observed at this position among different

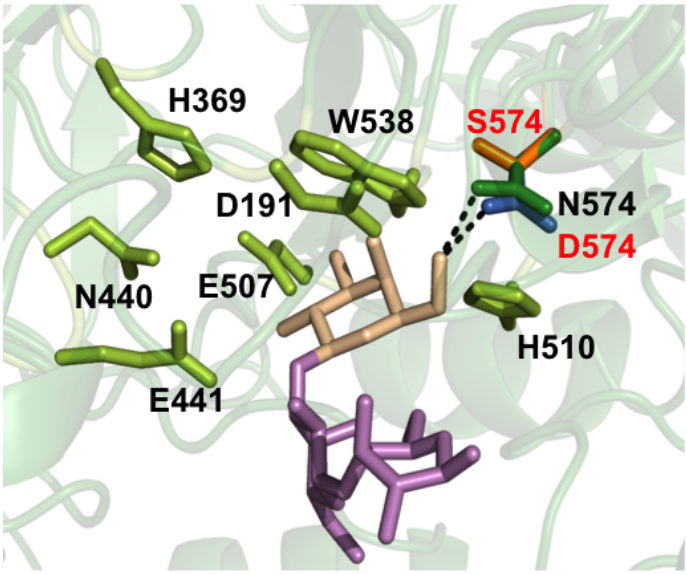


Figure 5.3: Structural detail of putative subsite -1. β-3'-galactosyl-lactose is depicted with the galactosyl moiety at the non-reducing end in light brown, and the lactosyl group acting as acceptor coloured in violet. Residues interacting with the galactosyl moiety at the non-reducing end are coloured in green. Residues substituting N574 in the N574D and N574S are overlapped and depicted in blue and orange, respectively. Dashed lines indicate putative hydrogen bonds at this specific position.

β-galactosidases is not associated to any functional adaptation.

According to docking studies, residues F571 and W959 conform the aromatic binding platform for the lactose molecule acting as galactosyl acceptor (Figure 4.22). In agreement with this, substitutions of residue W959 by non-aromatic residues resulted in a significant negative impact in transglycosylation (Figure 4.26 B). The homologous residue in *E.coli* galactosidase (W999) has been assigned a double role in the catalysis of the enzyme: binding lactose in the so-called "shallow" mode, and binding glucose in the galactosyl covalent intermediate during allolactose synthesis (Huber et al., 2003; Wheatley et al., 2013; Jacobson et al., 1994). Our current work suggests that this residue is also involved in lactose binding when the molecule acts as an acceptor during transgalactosylation. The effect of W959 substitutions was specially remarkable in the synthesis of β-3'-galactosyl-lactose (Figure 4.26 B). The β-(1,3) to β-(1,6) ratio decreased from 2.6 in the wild-type TmLac to 1 and 0.7 in W9859C and W959A, respectively. Interestingly, β-galactosidase from *K. lactis* with a Cys residue in the homologous position synthetizes preferentially β-6'-galactosyl-lactose (Rodriguez-Colinas et al., 2011) suggesting that this residue may be critical for product specificity.

Substitution of highly conserved F571 by a non-aromatic residue (F571L), resulted in a significant increase in the synthesis of β-3'-galactosyl-lactose (Figure 4.25 B). Modelling studies suggest that this substitution may allow the rotation of the residue W959, which is not possible in the wild-type enzyme by steric impediment (Figure 5.4). The reorientation of this residue, may yield a better approach of the acceptor lactose to synthesize β-3'-galactosyl-lactose. The combination of single mutations yielding highest GOS yield (F571L/N574A and F571L/N574S) further increased the transgalactosylating efficiency, up to two-fold compared with the wild-type TmLac. Additive effect of the two

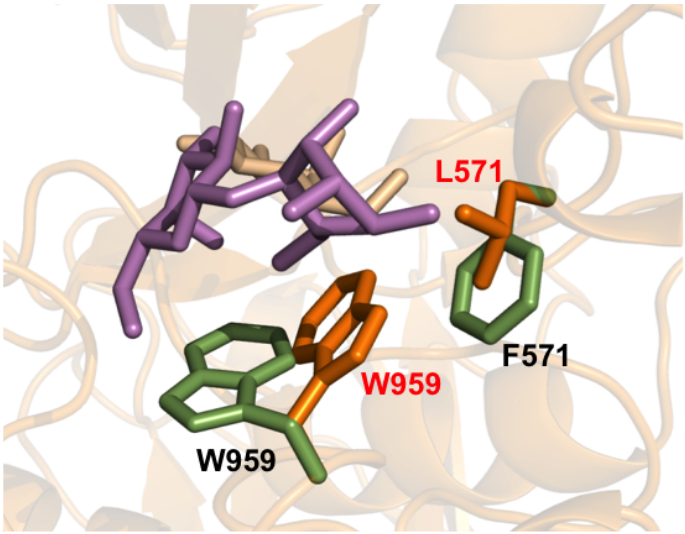


Figure 5.4: Structural model of wild-type TmLac and F571L mutant. β-3'-galactosyl-lactose is depicted with the galactosyl moiety at the non reducing end in light brown, and the lactosyl group acting as acceptor coloured in violet. The position of W959 and F571 in the wild-type enzyme is shown in green. The structure of the F571L mutant (orange) is overlapped and shows that the side chain of W959 is not sterically impeded to rotate to a different position.

single mutations would produce higher flexibility of the galactosyl moiety in the covalent intermediate and a more favorable docking of the acceptor lactose due to the W959 reorientation. This may result in a better coupling of the two reacting molecules. Since both N574 and F571 are highly conserved among β-galactosidases of the GH2 family (Figure 4.22), these results may be extrapolated to other enzymes, including those with different product specificity. Equivalent mutations may increase the yield of different GOS products.

5.2.1 Analysis of domain architecture and phylogenetics of GH2 β-glycosidases

Analysis of GH2 domain architectures (DAs) was carried out using Pfam and Interpro databases. Proteins have been classified in 5 DA types (Figure 4.27) according to C-terminal variable topologies. TmLac is included in type 3, representing canonical β-galactosidases, except for an additional terminal domain downstream Bgal_Small_N (Figure 4.15). The three-dimensional structure of several enzymes of DA types 1 and 3 is known, such as the β-glucuronidases from *E. coli*, *C. perfringens* or *H. sapiens* (Jain et al., 1996; Wallace et al., 2010, 2015), or the β-galactosidases from *Arthrobacter* sp., *E. coli* or *K. lactis* (Jacobson et al., 1994; Juers et al., 2001; Skalova et al., 2005; Pereira-Rodriguez et al., 2012), as well as part of the structures of the β-galactosidases from *B. circulans* (Ishikawa et al., 2015) and *S. pneumoniae* (Cheng et al., 2012) (DA type 5). These structures reveal that GH2C interacts with domains GH2N and with either Bgal_Small_N in DA type 3 proteins or BIG1 in DA type 5, through loops that shape the catalytic site. In the case of TmLac, W959, which has been involved in binding lactose both as

donor and acceptor of the galactosyl moiety, resides in the Bgal_Small_N. The close resemblance of Bgal_Small_N domains in monocistronic (DA type 3) and bicistronic (DA type 2) proteins indicates that these two types evolved from a common ancestor, either by gene fusion of LacL/LacM modules or by gene disruption from the canonical type. Phylogenetic analysis of GH2C domains (Figure 4.29) shows that bicistronic proteins emerge from branches of DA type 3, suggesting that they evolved from monocistronic sequences by gene disruption. The fact that bicistronic enzymes are dispersed in three different nodes may indicate that this was not a single evolutionary event. This gene disruption allowed the loss of the linker domain (primitive DUF4981).

Results shown in Figure 4.29 support the idea that the C-terminal domains directed evolutionary diversification of GH2 β-galactosidases. The signature of DA type 5* proteins, whose node is represented in Figure 4.30, is the presence of a BIG1 domain whereas the signature of the canonical DA type 3 β-galactosidases is the Bgal_Small_N domain. Indeed, the DA type 3 cluster includes not only canonical and bicistronic galactosidases but also a subset from type 4 carrying a C-terminal domain of around 400 amino acids (Ct_8_3) which is also related to Bgal_Small_N. This is in agreement with a close structural relationship between the C-terminal domain and the active site within the GH2C domain, that has been maintained through evolution. In contrast, β-glucuronidases (DA type 1) lacking C-terminal domains are dispersed in different nodes.

We propose a model for the evolution of the different β-galactosidases DAs in family GH2, based on the results of sequence alignment and phylogenetic analysis, which is presented in Figure 5.5.

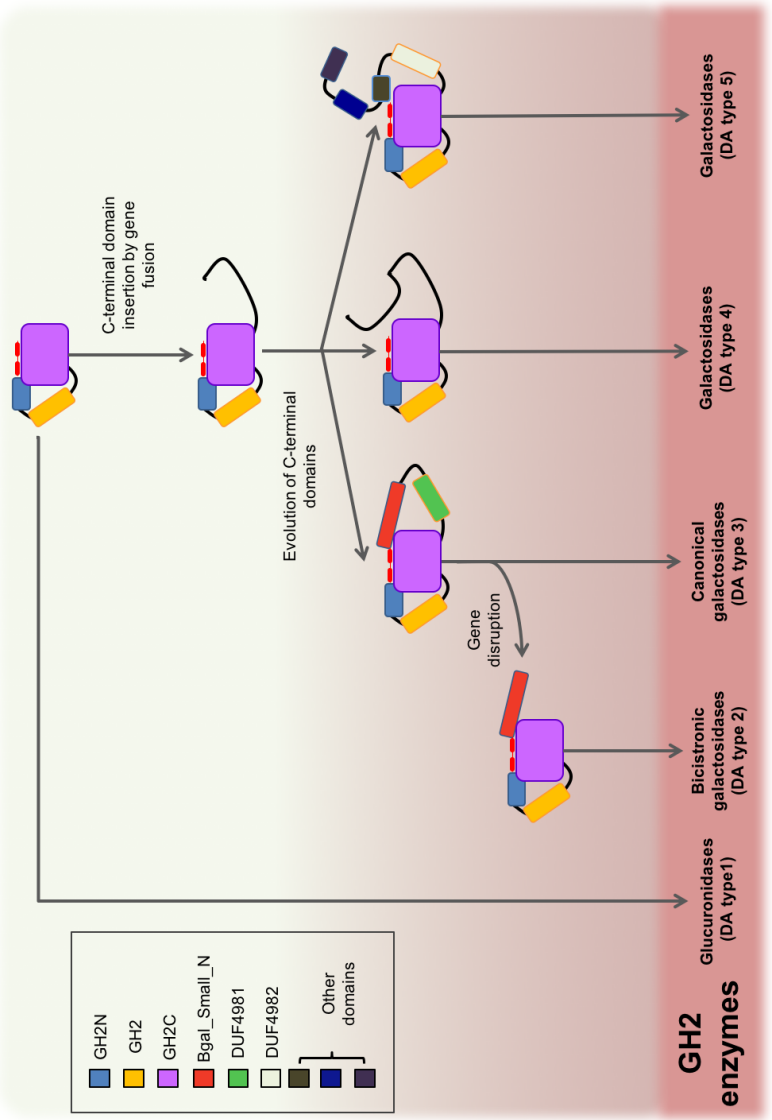


Figure 5.5: Evolutionary model proposed from the results extracted in this work.

Primitive GH2 enzymes probably had the architecture GH2N+GH2+GH2C from which current glucuronidases derive (DA type 1). Subsequently, one or more C-terminal domains were added by gene fusion rendering types 3, 4 and 5. A group of type 3 β-galactosidases gave rise to bicistronic β-galactosidases by gene disruption. Likely, the addition of some of these C-terminal domains such as BIG1 or Bgal_Small_N gave shape to the active site and determined the specialization of GH2 enzymes in the hydrolysis of different galactosides. In other cases, C-terminal domains seem to confer cell adhesion properties. For example, a module (CBM71) of the β-galactosidase from *Streptococcus pneumoniae* has been shown to be involved in host-pathogen recognition (Singh et al., 2014).

This study provides new insights that will help to screen and select novel β-galactosidases suited for specific GOS synthesis. Whereas β-(1,3) and β-(1,6) GOS are the main products of widely characterized β-galactosidases (classified in types 2 and 3), β-(1,4) GOS production has been described for the *B. circulans* β-galactosidase (type 5) (Torres et al., 2010). This suggests that the divergent domain architecture of β-galactosidases may be related to the different product specificities of the transglycosylation reaction. Sequence alignment of residues potentially involved in the catalytic site of *B. circulans* β-galactosidase suggests that other DA type 5 enzymes containing a BIG1 domain may be good candidates for β-(1,4) GOS production, despite high divergence in C-terminal DA. Our analysis also shows the existence of a heterogeneous, poorly characterized group of proteins (DA type 4) with divergent GH2C sequences. Biochemical characterization of enzymes from this group may be reveal novel properties of industrial interest.

Chapter 6

Conclusions

The results obtained in this doctoral thesis are summarized in the following conclusions:

1. Recombinant β-galactosidase from *Thermotoga maritima* (TmLac) produced in *E. coli*, has been purified and immobilized on a epoxy-activated polyvinyl alcohol resin. The resulting covalent, multi-point attachment of the protein to the support improves the enzyme stability and allows its use for lactose hydrolysis in repeated batches.

2. Hybrid enzymes composed of *Thermotoga maritima* β-galactosidase fused to diverse carbohydrate binding modules, have been constructed. The best results were achieved with a hybrid containing a CBM2 from *Pyrococcus furiosus* chitinase A. The hybrid is thermostable and can be attached under controlled conditions to cellulose and chitin.

3. With lactose as the substrate, TmLac carries out transglycosylation of a galactosyl residue to a lactose acceptor molecule, giving rise to two different trisaccharides: β-3'-galactosyl-lactose and β-6'-galactosyl-lactose at a 3:1 molar ratio.

4. β-3'-Galactosyl-lactose synthesis has been improved up to 40% by mutating the Asp 574 from TmLac structure to Ser or Ala. The observed effect may be due to increased rotational freedom of the galactosyl residue being transferred to the acceptor lactose.

5. The Trp at position 959 in the TmLac structure is actively involved in transglycosylation. Substitution of W959 by non-aromatic residues (such as Cys or Ala) decreases β-3'-galactosyl-lactose by about 80%, whereas change to another aromatic residue (Phe) does not cause any effect. We

suggest that this aromatic residue acts as a binding platform for the lactose molecule acting as galactosyl acceptor.

6. Change of Phe at position 571 to Leu increases β-3'-galactosyl-lactose synthesis by about 40%. This mutation may modify the orientation of the lactose binding platform to a more favourable position.

7. Several of the isolated mutations that increase the tranglycosylating activity of TmLac have an additive effect. So far a 2-fold increase of the transglycosylating activity of the wild-type has been achieved with mutants F571L/N574A and F571L/N574S.

8. Protein sequences from glycosyl hydrolases family 2, to which TmLac belongs, have been classified in five types, based on their domain architecture. Proteins of the family show a conserved structure at the N-terminal region of their sequence, but have a high versatility at the C-terminal region.

9. Phylogenetic analysis shows a close relationship between the evolution of the GH2C catalytic domain and the presence of specific C-terminal modules. A model has been proposed to explain the evolution of the different domain architectures existing in the family.

Bibliography

Adam, A. C., Rubio-Texeira, M., and Polaina, J. (2004). Lactose: the milk sugar from a biotechnological perspective. *Crit Rev Food Sci Nutr*, 44:553–557.

Akiyama, K., Takase, M., Horikoshi, K., and Okonogi, S. (2001). Production of galactooligosaccharides from lactose using a beta-glucosidase from thermus sp. z-1. *Biosci Biotechnol Biochem*, 65:438–441.

Ammam, M. and Fransaer, J. (2010). Two-enzyme lactose biosensor based on β-galactosidase and glucose oxidase deposited by ac-electrophoresis: Characteristics and performance for lactose determination in milk. *Sens Actuators B Chem*, 148:583–589.

Andrade, F. K., Moreira, S. M. G., Domingues, L., and Gama, F. M. P. (2010). Improving the affinity of fibroblasts for bacterial cellulose using carbohydrate-binding modules fused to rgd. *J Biomed Mater Res A*, 92:9–17.

Bae, T. and Schneewind, O. (2003). The YSIRK-G/S motif of staphylococcal protein a and its role in efficiency of signal peptide processing. *J Bacteriol*, 185:2910–2919.

Bakker, M., van De Velde, F., van Rantwijk, F., and Sheldon, R. A. (2000).

Highly efficient immobilization of glycosylated enzymes into polyurethane foams. *Biotechnol Bioeng*, 70:342–348.

Ball, J. C., Puckett, L. G., and Bachas, L. G. (2003). Covalent immobilization of beta-galactosidase onto a gold-coated magnetoelastic transducer via a self-assembled monolayer: toward a magnetoelastic biosensor. *Anal Chem*, 75:6932–6937.

Beijerinck, M. (1889). Die lactase, ein neues enzym. *Zentralbl Bakteriol*, 6:44–48.

Bimboim, H. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7:1513–1523.

Bode, L. (2009). Human milk oligosaccharides: prebiotics and beyond. *Nutr Rev*, 67 Suppl 2:S183–91.

Boon, Janssen, and van der Padt, A. (1999). Modelling and parameter estimation of the enzymatic synthesis of oligosaccharides by beta-galactosidase from *Bacillus circulans*. *Biotechnol Bioeng*, 64:558–567.

Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J*, 382:769–781.

Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A., and Kilburn, D. G. (2001). Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. *Biochemistry*, 40:6240–6247.

Bozhinova, D., Galunsky, B., Yueping, G., Franzreb, M., Koster, R., and Kasche, V. (2004). Evaluation of magnetic polymer micro-beads as carriers

of immobilised biocatalysts for selective and stereoselective transformations. *Biotechnol Lett*, 26:343–350.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72:248–254.

Bruins, M. E., Strubel, M., van Lieshout, J. F. T., Janssen, A. E. M., and Boom, R. M. (2003). Oligosaccharide synthesis by the hyperthermostable β-glucosidase from *Pyrococcus furiosus*: kinetics and modelling. *Enzyme Microb Technol*, 33:3–11.

Bruno-Barcena, J. M. and Azcarate-Peril, M. A. (2015). Galacto-oligosaccharides and colorectal cancer: Feeding our intestinal probiome. *J Funct Foods*, 12:92–108.

Cardelle-Cobas, A., Corzo, N., Olano, A., Pelaez, C., Requena, T., and Avila, M. (2011). Galactooligosaccharides derived from lactose and lactulose: influence of structure on *Lactobacillus*, *Streptococcus* and *Bifidobacterium* growth. *Int J Food Microbiol*, 149:81–87.

Cheng, W., Wang, L., Jiang, Y.-L., Bai, X.-H., Chu, J., Li, Q., Yu, G., Liang, Q.-L., Zhou, C.-Z., and Chen, Y. (2012). Structural insights into the substrate specificity of *Streptococcus pneumoniae* β(1,3)-galactosidase BgaC. *J Biol Chem*, 287:22910–22918.

Cherrington, C. A., Hinton, M., Pearson, G. R., and Chopra, I. (1991). Short-chain organic acids at ph 5.0 kill *Escherichia coli* and *Salmonella spp.* without causing membrane perturbation. *J Appl Bacteriol*, 70:161–165.

Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier,

J. J., Paulus, H., and Xu, M. Q. (1997). Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, 192:271–281.

Closa-Monasterolo, R., Gispert-Llaurado, M., Luque, V., Ferre, N., Rubio-Torrents, C., Zaragoza-Jordana, M., and Escribano, J. (2013). Safety and efficacy of inulin and oligofructose supplementation in infant formula: results from a randomized clinical trial. *Clin Nutr*, 32:918–927.

Consortium, T. U. (2015). Uniprot: a hub for protein information. *Nucleic Acids Res*, 43:D204–D212.

Courtois, J. (2009). Oligosaccharides from land plants and algae: production and applications in therapeutics and biotechnology. *Curr Opin Microbiol*, 12:261–273.

Cowan, D. A. and Fernandez-Lafuente, R. (2011). Enhancing the functional properties of thermophilic enzymes by chemical modification and immobilization. *Enzyme Microb Technol*, 49:326–346.

Crittenden, R. G. and Playne, M. J. (1996). Production, properties and applications of food-grade oligosaccharides. *Trends Food Sci Technol*, 7:353–361.

Davies, G. J., Wilson, K. S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J*, 321:557–559.

DeDent, A., Bae, T., Missiakas, D. M., and Schneewind, O. (2008). Signal peptides direct surface proteins to two distinct envelope locations of *Staphylococcus aureus*. *EMBO J*, 27:2656–2668.

den Besten, G., van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D.-J., and Bakker, B. M. (2013). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res*, 54:2325–2340.

Duncan, S. H., Louis, P., Thomson, J. M., and Flint, H. J. (2009). The role of pH in determining the species composition of the human colonic microbiota. *Environ Microbiol*, 11:2112–2122.

Fan, Y., Hua, X., Zhang, Y., Feng, Y., Shen, Q., Dong, J., Zhao, W., Zhang, W., Jin, Z., and Yang, R. (2015). Cloning, expression and structural stability of a cold-adapted β-galactosidase from *Rahnella sp.* r3. *Protein Expr Purif*, 115:158–164.

Feng, W. and Ji, P. (2011). Enzymes immobilized on carbon nanotubes. *Biotechnol Adv*, 29:889–895.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., and Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Res*, 42:D222–D230.

Gänzle, M. G. (2012). Enzymatic synthesis of galacto-oligosaccharides and other lactose derivatives (hetero-oligosaccharides) from lactose. *Int Dairy J*, 22:116–122.

Gibson, G. R. and Roberfroid, M. B. (1995). Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *J Nutr*, 125:1401–1412.

Goffin, D., Delzenne, N., Blecker, C., Hanon, E., Deroanne, C., and Paquot, M. (2011). Will isomalto-oligosaccharides, a well-established functional food

in Asia, break through the European and American market? The status of knowledge on these prebiotics. *Crit Rev Food Sci Nutr*, 51:394–409.

Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., and Eisenberg, D. (2007). Toward rational protein crystallization: A web server for the design of crystallizable protein variants. *Protein Sci*, 16:1569–1576.

Goulas, T., Goulas, A., Tzortzis, G., and Gibson, G. R. (2009). Expression of four beta-galactosidases from *Bifidobacterium bifidum* NCIMB41171 and their contribution on the hydrolysis and synthesis of galactooligosaccharides. *Appl Microbiol Biotechnol*, 84:899–907.

Green, Michael R. (Michael Richard), ., Sambrook, J., and Laboratory, C. S. H. (2012). *Molecular cloning : a laboratory manual / Michael R. Green, Joseph Sambrook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Greenberg, N. A. and Mahoney, R. R. (1983). Formation of oligosaccharides by β-galactosidase from *Streptococcus thermophilus*. *Food Chem*, 10:195–204.

Guillén, D., Sánchez, S., and Rodríguez-Sanoja, R. (2010). Carbohydrate-binding domains: multiplicity of biological roles. *Appl Microbiol Biotechnol*, 85:1241–1249.

Guimarães, P. M. R., Teixeira, J. A., and Domingues, L. (2010). Fermentation of lactose to bio-ethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. *Biotechnol Adv*, 28:375–384.

Haimovitz, R., Barak, Y., Morag, E., Voronov-Goldman, M., Shoham, Y., Lamed, R., and Bayer, E. A. (2008). Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules. *Proteomics*, 8:968–979.

Hall, T. (1999). Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser (Oxf)*, 41:95–98.

Hanazono, Y., Takeda, K., Niwa, S., Hibi, M., Takahashi, N., Kanai, T., Atomi, H., and Miki, K. (2016). Crystal structures of chitin binding domains of chitinase from *Thermococcus kodakarensis* KOD1. *FEBS Lett*, 590:298–304.

Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., and Galas, D. J. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res*, 17:6545–6551.

Hernandez-Hernandez, O., Marin-Manzano, M. C., Rubio, L. A., Moreno, F. J., Sanz, M. L., and Clemente, A. (2012). Monomer and linkage type of galacto-oligosaccharides affect their resistance to ileal digestion and prebiotic properties in rats. *J Nutr*, 142:1232–1239.

Hernot, D. C., Boileau, T. W., Bauer, L. L., Middelbos, I. S., Murphy, M. R., Swanson, K. S., and Fahey, G. C. J. (2009). In vitro fermentation profiles, gas production rates, and microbiota modulation as affected by certain fructans, galactooligosaccharides, and polydextrose. *J Agric Food Chem*, 57:1354–1361.

Hirosawa, S., Minato, K., and Nakatsubo, F. (2001). Influence of carboxyl group on the acid hydrolysis of cellulose. *J Wood Sci*, 47:141–144.

Homaei, A. A., Sariri, R., Vianello, F., and Stevanato, R. (2013). Enzyme immobilization: an update. *J Chem Biol*, 6:185–205.

Hong, J., Ye, X., Wang, Y., and Zhang, Y.-H. P. (2008). Bioseparation of recombinant cellulose-binding module-proteins by affinity adsorption on an ultra-high-capacity cellulosic adsorbent. *Anal Chim Acta*, 621:193–199.

Huber, R. E., Hakda, S., Cheng, C., Cupples, C. G., and Edwards, R. A. (2003). Trp-999 of beta-galactosidase (*Escherichia coli*) is a key residue for binding, catalysis, and synthesis of allolactose, the natural *lac* operon inducer. *Biochemistry*, 42:1796–1803.

Huber, R. E., Hlede, I. Y., Roth, N. J., McKenzie, K. C., and Ghumman, K. K. (2001). His-391 of beta-galactosidase (*Escherichia coli*) promotes catalyses by strong interactions with the transition state. *Biochem Cell Biol*, 79:183–193.

Husain, Q. (2010). Beta galactosidases and their potential applications: a review. *Crit Rev Biotechnol*, 30:41–62.

Huson, D. H. and Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol*, 61:1061–1067.

Hussack, G., Luo, Y., Veldhuis, L., Hall, J. C., Tanha, J., and MacKenzie, R. (2009). Multivalent anchoring and oriented display of single-domain antibodies on cellulose. *Sensors*, 9:5351.

Ingram, C. J. E., Mulcare, C. A., Itan, Y., Thomas, M. G., and Swallow, D. M. (2009). Lactose digestion and the evolutionary genetics of lactase persistence. *Hum Genet*, 124:579–591.

Ishikawa, K., Kataoka, M., Yanamoto, T., Nakabayashi, M., Watanabe, M., Ishihara, S., and Yamaguchi, S. (2015). Crystal structure of beta-galactosidase from *Bacillus circulans* ATCC 31382 (BgaD) and the construction of the thermophilic mutants. *FEBS J*, 282:2540–2552.

Itan, Y., Jones, B., Ingram, C., Swallow, D., and Thomas, M. (2010). A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evolutionary Biology*, 10:1–11.

Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994). Three-dimensional structure of beta-galactosidase from e. coli. *Nature*, 369:761–766.

Jain, S., Drendel, W. B., Chen, Z. W., Mathews, F. S., Sly, W. S., and Grubb, J. H. (1996). Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. *Nat Struct Biol*, 3:375–381.

Ji, E.-S., Park, N.-H., and Oh, D.-K. (2005). Galacto-oligosaccharide production by a thermostable recombinant β-galactosidase from thermotoga maritima. *World Journal of Microbiology and Biotechnology*, 21:759–764.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*, 8:275–282.

Juers, D. H., Heightman, T. D., Vasella, A., McCarter, J. D., Mackenzie, L., Withers, S. G., and Matthews, B. W. (2001). A structural view of the action of *Escherichia coli* (LacZ) beta-galactosidase. *Biochemistry*, 40:14781–14794.

Juers, D. H., Huber, R. E., and Matthews, B. W. (1999). Structural comparisons of tim barrel proteins suggest functional and evolutionary relationships between beta-galactosidase and other glycohydrolases. *Protein Sci*, 8:122–136.

Juers, D. H., Matthews, B. W., and Huber, R. E. (2012). LacZ beta-galactosidase: structure and function of an enzyme of historical and molecular biological importance. *Protein Sci*, 21:1792–1807.

Kavoosi, M., Lam, D., Bryan, J., Kilburn, D. G., and Haynes, C. A. (2007). Mechanically stable porous cellulose media for affinity purification of family 9

cellulose-binding module-tagged fusion proteins. *J Chromatogr A*, 1175:187–196.

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008). GLYCAM06: a generalizable biomolecular force field. carbohydrates. *J Comput Chem*, 29:622–655.

Kunz, C., Rudloff, S., Baier, W., Klein, N., and Strobel, S. (2000). Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annu Rev Nutr*, 20:699–722.

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem*, 77:521–555.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). Clustal W and clustal X version 2.0. *Bioinformatics*, 23:2947–2948.

Latorre-García, L., del Castillo-Agudo, L., and Polaina, J. (2007). Taxonomical classification of yeasts isolated from kefir based on the sequence of their ribosomal RNA genes. *World J Microbiol Biotechnol*, 23:785–791.

Lebeer, S., Vanderleyden, J., and De Keersmaecker, S. C. J. (2010). Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. *Nat Rev Microbiol*, 8:171–184.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*, 42:490–495.

Lomer, M. C. E., Parkes, G. C., and Sanderson, J. D. (2008). Review article: lactose intolerance in clinical practice - myths and realities. *Aliment Pharmacol Ther*, 27:93–103.

Lukacheva, L. V., Zakemovskaya, A. A., Karyakina, E. E., Zorov, I. N., Sinitsyn, A. P., Sukhacheva, M. V., Netrusov, A. I., and Karyakin, A. A. (2007). Determination of glucose and lactose in food products with the use of biosensors based on Berlin blue. *Fresenius J Anal Chem*, 62:388–393.

Macfarlane, G. T., Steed, H., and Macfarlane, S. (2008). Bacterial metabolism and health-related effects of galacto-oligosaccharides and other prebiotics. *J Appl Microbiol*, 104:305–344.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., and Bryant, S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res*, 43:D222–D226.

Mateo, C., Fernandez-Lorente, G., Abian, O., Fernandez-Lafuente, R., and Guisan, J. M. (2000). Multifunctional epoxy supports: a new tool to improve the covalent immobilization of proteins. The promotion of physical adsorptions of proteins on the supports before their covalent linkage. *Biomacromolecules*, 1:739–745.

Mateo, C., Grazu, V., Palomo, J. M., Lopez-Gallego, F., Fernandez-Lafuente, R., and Guisan, J. M. (2007a). Immobilization of enzymes on heterofunctional epoxy supports. *Nat Protoc*, 2:1022–1033.

Mateo, C., Palomo, J. M., Fernandez-Lorente, G., Guisan, J. M., and Fernandez-Lafuente, R. (2007b). Improvement of enzyme activity, stability

and selectivity via immobilization techniques. *Enzyme Microb Technol*, 40:1451–1463.

Mattar, R., de Campos Mazo, D. F., and Carrilho, F. J. (2012). Lactose intolerance: diagnosis, genetic, and clinical factors. *Clin Exp Gastroenterol*, 5:113–121.

Millqvist-Fureby, A., MacManus, D. A., Davies, S., and Vulfson, E. N. (1998). Enzymatic transformations in supersaturated substrate solutions: Ii. Synthesis of disaccharides via transglycosylation. *Biotechnol Bioeng*, 60:197–203.

Mine, S., Nakamura, T., Sato, T., Ikegami, T., and Uegaki, K. (2014). Solution structure of the chitin-binding domain 1 (chBD1) of a hyperthermophilic chitinase from *Pyrococcus furiosus*. *J Biochem*, 155:115–122.

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S.-Y., Bateman, A., Punta, M., Attwood, T. K., Sigrist, C. J. A., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D. A., Wu, C. H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P. D., and Finn, R. D. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res*, 43:D213–D221.

Monsan, P. F. and Ouarné, F. (2009). *Oligosaccharides Derived from Sucrose*, chapter Oligosaccharides Derived from Sucrose, pages 293–336. Springer New York.

Montanier, C., van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firbank, S. J., Lewis, R. J., Grondin, G. G., Ghinet, M. G.,

Gloster, T. M., Herve, C., Knox, J. P., Talbot, B. G., Turkenburg, J. P., Kerovuo, J., Brzezinski, R., Fontes, C. M. G. A., Davies, G. J., Boraston, A. B., and Gilbert, H. J. (2009). Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *PNAS*, 106:3065–3070.

Moore, J. B., Markiewicz, P., and Miller, J. H. (1994). Identification and sequencing of the *Thermotoga maritima* lacZ gene, part of a divergently transcribed operon. *Gene*, 147:101–106.

Moreno, F. J., Montilla, A., Villamiel, M., Corzo, N., and Olano, A. (2014). Analysis, structural characterization, and bioactivity of oligosaccharides derived from lactose. *Electrophoresis*, 35:1519–1534.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem*, 30:2785–2791.

Mu, W., Chen, Q., Wang, X., Zhang, T., and Jiang, B. (2013). Current studies on physiological functions and biological production of lactosucrose. *Appl Microbiol Biotechnol*, 97:7073–7080.

Myung, S., Zhang, X.-Z., and Percival Zhang, Y.-H. (2011). Ultra-stable phosphoglucose isomerase through immobilization of cellulose-binding module-tagged thermophilic enzyme on low-cost high-capacity cellulosic adsorbent. *Biotechnol Progr*, 27:969–975.

Nakagawa, T., Ikehata, R., Uchino, M., Miyaji, T., Takano, K., and Tomizuka, N. (2006). Cold-active acid β-galactosidase activity of isolated psychrophilic-basidiomycetous yeast *Guehomyces pullulans*. *Microbiol Res*, 161:75–79.

Nakamura, T., Mine, S., Hagihara, Y., Ishikawa, K., Ikegami, T., and Uegaki, K. (2008). Tertiary structure and carbohydrate recognition by the chitin-binding domain of a hyperthermophilic chitinase from *Pyrococcus furiosus*. *J Mol Biol*, 381:670–680.

Notenboom, V., Boraston, A. B., Kilburn, D. G., and Rose, D. R. (2001). Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry*, 40:6248–6256.

Oliveira, C., Carvalho, V., Domingues, L., and Gama, F. M. (2015). Recombinant CBM-fusion technology - Applications overview. *Biotechnol Adv*, 33:358–369.

Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput*, 7:525–537.

Panesar, P. S., Panesar, R., Singh, R. S., Kennedy, J. F., and Kumar, H. (2006). Microbial production, immobilization and applications of β-D-galactosidase. *J Chem Technol Biotechnol*, 81:530–543.

Park, A.-R. and Oh, D.-K. (2010). Galacto-oligosaccharide production using microbial β-galactosidase: current state and perspectives. *Appl Microbiol Biotechnol*, 85:1279–1286.

Patel, S. and Goyal, A. (2011). Functional oligosaccharides: production, properties and applications. *World J Microbiol Biotechnol*, 27:1119–1128.

Pawlak-Szukalska, A., Wanarska, M., Popinigis, A. T., and Kur, J. (2014). A novel cold-active β-D-galactosidase with transglycosylation activity from

the Antarctic *Arthrobacter sp.* 32cB - Gene cloning, purification and characterization. *Process Biochem*, 49:2122–2133.

Pereira-Rodriguez, A., Fernandez-Leiro, R., Gonzalez-Siso, M. I., Cerdan, M. E., Becerra, M., and Sanz-Aparicio, J. (2012). Structural basis of specificity in tetrameric *Kluyveromyces lactis* beta-galactosidase. *J Struct Biol*, 177:392–401.

Petzelbauer, I., Reiter, A., Splechtna, B., Kosma, P., and Nidetzky, B. (2000). Transgalactosylation by thermostable beta-glycosidases from *Pyrococcus furiosus* and *Sulfolobus solfataricus*. Binding interactions of nucleophiles with the galactosylated enzyme intermediate make major contributions to the formation of new beta-glycosides during lactose conversion. *Eur J Biochem*, 267:5055–5066.

Placier, G., Watzlawick, H., Rabiller, C., and Mattes, R. (2009). Evolved beta-galactosidases from *Geobacillus stearothermophilus* with improved trans-galactosylation yield for galacto-oligosaccharide production. *Appl Environ Microbiol*, 75:6312–6321.

Polaina, J. and Adam, A. C. (1991). A fast procedure for yeast DNA purification. *Nucleic Acids Res*, 19:5443.

Polaina, J. and MacCabe, A. (2007). *Industrial Enzymes: Structure, Function and Applications*. Springer ebook collection / Chemistry and Materials Science 2005-2008. Springer Netherlands.

Prasad, S., Dhiman, R. K., Duseja, A., Chawla, Y. K., Sharma, A., and Agarwal, R. (2007). Lactulose improves cognitive functions and health-related quality of life in patients with cirrhosis who have minimal hepatic encephalopathy. *Hepatology*, 45:549–559.

Quintero, M., Maldonado, M., Perez-Munoz, M., Jimenez, R., Fangman, T., Rupnow, J., Wittke, A., Russell, M., and Hutkins, R. (2011). Adherence inhibition of *Cronobacter sakazakii* to intestinal epithelial cells by prebiotic oligosaccharides. *Curr Microbiol*, 62:1448–1454.

Raman, R., Raguram, S., Venkataraman, G., Paulson, J. C., and Sasisekharan, R. (2005). Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Methods*, 2:817–824.

Richins, R. D., Mulchandani, A., and Chen, W. (2000). Expression, immobilization, and enzymatic characterization of cellulose-binding domain-organophosphorus hydrolase fusion enzymes. *Biotechnol Bioeng*, 69:591–596.

Rigden, D. J. (2005). Analysis of glycoside hydrolase family 98: catalytic machinery, mechanism and a novel putative carbohydrate binding module. *FEBS Lett*, 579:5466–5472.

Roberfroid, M. (2007). Prebiotics: The concept revisited. *J Nutr*, 137:830S–837S.

Roberfroid, M., Gibson, G. R., Hoyles, L., McCartney, A. L., Rastall, R., Rowland, I., Wolvers, D., Watzl, B., Szajewska, H., Stahl, B., Guarner, F., Respondek, F., Whelan, K., Coxam, V., Davicco, M.-J., Leotoing, L., Wittrant, Y., Delzenne, N. M., Cani, P. D., Neyrinck, A. M., and Meheust, A. (2010). Prebiotic effects: metabolic and health benefits. *Br J Nutr*, 104 Suppl 2:S1–63.

Rodriguez, B., Kavoosi, M., Koska, J., Creagh, A. L., Kilburn, D. G., and Haynes, C. A. (2004). Inexpensive and generic affinity purification of recombinant proteins using a family 2a CBM fusion tag. *Biotechnol Prog*, 20:1479–1489.

Rodriguez-Colinas, B., de Abreu, M. A., Fernandez-Arrojo, L., de Beer, R., Poveda, A., Jimenez-Barbero, J., Haltrich, D., Ballesteros Olmo, A. O., Fernandez-Lobato, M., and Plou, F. J. (2011). Production of Galacto-oligosaccharides by the beta-Galactosidase from *Kluyveromyces lactis* : comparative analysis of permeabilized cells versus soluble enzyme. *J Agric Food Chem*, 59:10477–10484.

Rosado, J. L. (1997). Lactose digestion and maldigestion: implications for dietary habits in developing countries. *Nutr Res Rev*, 10:137–149.

Roth, N. J. and Huber, R. E. (1996). The beta-galactosidase (*Escherichia coli*) reaction is partly facilitated by interactions of His-540 with the C6 hydroxyl of galactose. *J Biol Chem*, 271:14296–14301.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, 5:725–738.

Rutenber, E., Ready, M., and Robertus, J. (1987). Structure and evolution of ricin B chain. *Nature*, 326:624—626.

Sabater-Molina, M., Larque, E., Torrella, F., and Zamora, S. (2009). Dietary fructooligosaccharides and potential benefits on health. *J Physiol Biochem*, 65:315–328.

Schrödinger, LLC (2010). The PyMOL molecular graphics system, version 1.3r1.

Schuster-Wolff-Bühring, R., Fischer, L., and Hinrichs, J. (2010). Production and physiological action of the disaccharide lactulose. *Int Dairy J*, 20:731–741.

Sezginturk, M. K. and Dinckaya, E. (2008). Beta-galactosidase monitoring by a biosensor based on Clark electrode: its optimization, characterization and application. *Biosens Bioelectron*, 23:1799–1804.

Sheldon, R. A. and van Pelt, S. (2013). Enzyme immobilisation in biocatalysis: why, what and how. *Chem Soc Rev*, 42:6223–6235.

Shoaf, K., Mulvey, G. L., Armstrong, G. D., and Hutkins, R. W. (2006). Prebiotic galactooligosaccharides reduce adherence of enteropathogenic *Escherichia coli* to tissue culture cells. *Infect Immun*, 74:6920–6928.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular systems biology*, 7:539.

Sinclair, H. R., de Slegte, J., Gibson, G. R., and Rastall, R. A. (2009). Galactooligosaccharides (GOS) inhibit vibrio cholerae toxin binding to its GM1 receptor. *J Agric Food Chem*, 57:3113–3119.

Singh, A. K., Pluvinage, B., Higgins, M. A., Dalia, A. B., Woodiga, S. A., Flynn, M., Lloyd, A. R., Weiser, J. N., Stubbs, K. A., Boraston, A. B., and King, S. J. (2014). Unravelling the multiple functions of the architecturally intricate *Streptococcus pneumoniae* beta-galactosidase, BgaA. *PLoS Pathog*, 10:e1004364.

Skalova, T., Dohnalek, J., Spiwok, V., Lipovova, P., Vondrackova, E., Petrokova, H., Duskova, J., Strnad, H., Kralova, B., and Hasek, J. (2005). Cold-active beta-galactosidase from *Arthrobacter* sp. C2-2 forms compact 660 kDa hexamers: crystal structure at 1.9 å resolution. *J Mol Biol*, 353:282–294.

Sterchi, E. E., Mills, P. R., Fransen, J. A., Hauri, H. P., Lentze, M. J., Naim, H. Y., Ginsel, L., and Bond, J. (1990). Biogenesis of intestinal lactase-phlorizin hydrolase in adults with lactose intolerance. Evidence for reduced biosynthesis and slowed-down maturation in enterocytes. *J Clin Invest*, 86:1329–1337.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). Mega6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol*, 30:2725–2729.

Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., Durkin, A. S., Gwinn, M., Kolonay, J. F., Nelson, W. C., Peterson, J. D., Umayam, L. A., White, O., Salzberg, S. L., Lewis, M. R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A. M., Utterback, T. R., Hansen, C. L., McDonald, L. A., Feldblyum, T. V., Angiuoli, S., Dickinson, T., Hickey, E. K., Holt, I. E., Loftus, B. J., Yang, F., Smith, H. O., Venter, J. C., Dougherty, B. A., Morrison, D. A., Hollingshead, S. K., and Fraser, C. M. (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science*, 293:498–506.

Torres, D. P. M., Gonçalves, M. d. P. F., Teixeira, J., and Rodrigues, L. R. (2010). Galacto-Oligosaccharides: Production, properties, applications, and significance as prebiotics. *Compr Rev Food Sci Food Saf*, 9:438–454.

Torres, R., Mateo, C., Fernandez-Lorente, G., Ortiz, C., Fuentes, M., Palomo, J. M., Guisan, J. M., and Fernandez-Lafuente, R. (2003). A novel heterofunctional epoxy-amino sepabeads for a new enzyme immobilization protocol: immobilization-stabilization of beta-galactosidase from *Aspergillus oryzae*. *Biotechnol Prog*, 19:1056–1060.

Torres-Salas, P., del Monte-Martinez, A., Cutino-Avila, B., Rodriguez-Colinas, B., Alcalde, M., Ballesteros, A. O., and Plou, F. J. (2011). Immobilized biocatalysts: novel approaches and tools for binding enzymes to supports. *Adv Mater*, 23:5275–5282.

Tsukimoto, K., Takada, R., Araki, Y., Suzuki, K., Karita, S., Wakagi, T., Shoun, H., Watanabe, T., and Fushinobu, S. (2010). Recognition of cellooligosaccharides by a family 28 carbohydrate-binding module. *FEBS Lett*, 584:1205–1211.

van den Heuvel, E. G. H. M., Schoterman, M. H. C., and Muijs, T. (2000). Transgalactooligosaccharides stimulate calcium absorption in postmenopausal women. *J Nutr*, 130:2938–2942.

Velikodvorskaya, G. A., Tikhonova, T. V., Gurvits, I. D., Karyagina, A. S., Lavrova, N. V., Sergienko, O. V., Tashlitskii, V. N., Lunina, N. A., and Lunin, V. G. (2010). Chimeric lactase capable of spontaneous and strong immobilization on cellulose and development of a continuous-flow system for lactose hydrolysis at high temperatures. *Appl Environ Microbiol*, 76:8071–8075.

Vera, C., Guerrero, C., Illanes, A., and Conejeros, R. (2011). A pseudo steady-state model for galacto-oligosaccharides synthesis with beta-galactosidase from *Aspergillus oryzae*. *Biotechnol Bioeng*, 108:2270–2279.

Verheijden, K. A. T., Akbari, P., Willemsen, L. E. M., Kraneveld, A. D., Folkerts, G., Garssen, J., Fink-Gremmels, J., and Braber, S. (2015). Inflammation-induced expression of the alarmin interleukin 33 can be suppressed by galacto-oligosaccharides. *Int Arch Allergy Immunol*, 167:127–136.

Verma, M. L., Barrow, C. J., Kennedy, J. F., and Puri, M. (2012). Immobilization of β-D-galactosidase from *Kluyveromyces lactis* on functionalized silicon dioxide nanoparticles: Characterization and lactose hydrolysis. *Int J of Biol Macromol*, 50:432–437.

Vulevic, J., Drakoularakou, A., Yaqoob, P., Tzortzis, G., and Gibson, G. R. (2008). Modulation of the fecal microflora profile and immune function by a novel trans-galactooligosaccharide mixture (B-GOS) in healthy elderly volunteers. *Am J Clin Nutr*, 88:1438–1446.

Walker, J., Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M., Appel, R., and Bairoch, A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server*, pages 571–607. Humana Press.

Wallace, B. D., Roberts, A. B., Pollet, R. M., Ingle, J. D., Biernat, K. A., Pellock, S. J., Venkatesh, M. K., Guthrie, L., O'Neal, S. K., Robinson, S. J., Dollinger, M., Figueroa, E., McShane, S. R., Cohen, R. D., Jin, J., Frye, S. V., Zamboni, W. C., Pepe-Ranney, C., Mani, S., Kelly, L., and Redinbo, M. R. (2015). Structure and inhibition of microbiome beta-glucuronidases essential to the alleviation of cancer drug toxicity. *Chem Biol*, 22:1238–1249.

Wallace, B. D., Wang, H., Lane, K. T., Scott, J. E., Orans, J., Koo, J. S., Venkatesh, M., Jobin, C., Yeh, L.-A., Mani, S., and Redinbo, M. R. (2010). Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science*, 330:831–835.

Wang, Y., Harvey, C. B., Hollox, E. J., Phillips, A. D., Poulter, M., Clay, P., Walker-Smith, J. A., and Swallow, D. M. (1998). The genetically programmed down-regulation of lactase in children. *Gastroenterology*, 114:1230–1236.

Weijers, C. A. G. M., Franssen, M. C. R., and Visser, G. M. (2008). Glycosyltransferase-catalyzed synthesis of bioactive oligosaccharides. *Biotechnol Adv*, 26:436–456.

Wheatley, J. B. and Schmidt Jr, D. E. (1999). Salt-induced immobilization of affinity ligands onto epoxide-activated supports. *J Chromatogr A*, 849:1–12.

Wheatley, R. W., Kappelhoff, J. C., Hahn, J. N., Dugdale, M. L., Dutkoski, M. J., Tamman, S. D., Fraser, M. E., and Huber, R. E. (2012). Substitution for Asn460 cripples beta-galactosidase (*Escherichia coli*) by increasing substrate affinity and decreasing transition state stability. *Arch Biochem Biophys*, 521:51–61.

Wheatley, R. W., Lo, S., Jancewicz, L. J., Dugdale, M. L., and Huber, R. E. (2013). Structural explanation for allolactose (lac operon inducer) synthesis by lacZ beta-galactosidase and the evolutionary relationship between allolactose synthesis and the lac repressor. *J Biol Chem*, 288:12993–13005.

Whisner, C. M., Martin, B. R., Schoterman, M. H. C., Nakatsu, C. H., McCabe, L. D., McCabe, G. P., Wastney, M. E., van den Heuvel, E. G. H. M., and Weaver, C. M. (2013). Galacto-oligosaccharides increase calcium absorption and gut bifidobacteria in young girls: a double-blind cross-over trial. *Br J Nutr*, 110:1292–1303.

Wierzba, A., Reichl, U., Turner, R. F., Warren, R. A., and Kilburn, D. G. (1995). Production and properties of a bifunctional fusion protein that mediates attachment of vero cells to cellulosic matrices. *Biotechnol Bioeng*, 47:147–154.

Wu, Y., Yuan, S., Chen, S., Wu, D., Chen, J., and Wu, J. (2013). Enhancing

the production of galacto-oligosaccharides by mutagenesis of *Sulfolobus solfataricus* beta-galactosidase. *Food Chem*, 138:1588–1595.

Xu, J., McRae, M. A. A., Harron, S., Rob, B., and Huber, R. E. (2004). A study of the relationships of interactions between Asp-201, Na^+ or K^+, and galactosyl C6 hydroxyl and their effects on binding and reactivity of beta-galactosidase. *Biochem Cell Biol*, 82:275–284.

Xu, Z., Bae, W., Mulchandani, A., Mehra, R. K., and Chen, W. (2002). Heavy metal removal by novel CBD-EC20 sorbents immobilized on cellulose. *Biomacromolecules*, 3:462–465.

Yanahira, S., Kobayashi, T., Suguri, T., Nakakoshi, M., Miura, S., Ishikawa, H., and Nakajima, I. (1995). Formation of oligosaccharides from lactose by *Bacillus circulans* beta-galactosidase. *Biosci Biotechnol Biochem*, 59:1021–1026.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat Meth*, 12:7–8.

Ye, X., Zhu, Z., Zhang, C., and Zhang, Y.-H. P. (2011). Fusion of a family 9 cellulose-binding module improves catalytic potential of *Clostridium thermocellum* cellodextrin phosphorylase on insoluble cellulose. *Appl Microbiol Biotechnol*, 92:551–560.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9:40.

Zhou, Q. Z. K. and Dong Chen, X. (2001). Immobilization of β-galactosidase on graphite surface by glutaraldehyde. *J Food Eng*, 48:69–74.

Appendix

A. Functions developed in PYTHON

To obtain the data used in the DA study of the GH2 family different functions were created using PYTHON language that allowed to obtain the necessary data for the analysis performed in the current work. In this section the different functions that compose the script "Domain_study.py" are described. Note that this code is for particular input files and with a particular data format, because of that it may not work for other files without previous editing.

Domain_seq function

Domain_seq function prints the sequence between two coordinates defined by Pfam domain coverage using as reference the Pfam envelope coordinates.

```
def domain_seq(sequence, coordinate):
        start=int(coordinate[0])-1
        endd = int(coordinate[1])
        seq = sequence[start:endd]
        if not seq.strip():
                print 'sequence is empty', coordinate, start, endd
        return seq
```

Get_sequence_dict function

This function creates a dictionary using the Gene Bank numbers (GI) as keys and the protein sequences as values from FASTA sequences file, indicated as "Sequences.txt' on the code.

```
def get_sequece_dict():
        f2=open('Sequences.txt', 'r')
        fasta=f2.read()
        f2.close()
        fasta_dic={}
        entry=fasta.split(">")[1:]
        for line in entry:
                sp=line.split('\n',1)
                sp1=sp[0]
                sp2=sp[1].replace("\n", "")
                fasta_dic[sp1]=sp2
        return fasta_dic
```

Create_domaindict function

This function creates the dictionary *domaindict* using the GI numbers as keys and a list of Pfam domain names as values. It also, creates the *domaindict2* dictionary which contains the GI numbers as keys and the list of coordinates for all the Pfam domains belonging to GI from the Uniprot file as values, showed in the code as "UNIPROT_INFO.txt".

```
def create_domaindict():
        f = open('UNIPROT_INFO.txt', 'r')
        domaindict=dict()
        domaindict2=dict()
        for line in f:
                l = re.split('/', line)
                if l[0] not in domaindict:
                        li=[]
                        li.append(l[5])
                        domaindict[l[0]]=li
                else:
                        li=domaindict[l[0]]
                        li.append(l[5])
                        domaindict[l[0]]=li
                list1=[l[3],l[4]]
                if l[0] not in domaindict2:
                        li2=[]
                        li2.append(list1)
                        domaindict2[l[0]]=li2
                else:
                        li2=domaindict2[l[0]]
                        li2.append(list1)
                        domaindict2[l[0]]=li2
        f.close()
        return domaindict, domaindict2
```

Create_binary_colnames function

This function creates a dictionary using the GI as keys and the binary vector as value, where each value corresponds to the presence/absence to each Pfam domain.

```
def create_binary_colnames(domaindict):
        pfam_list=[]
        coutunters={}
        for domains in domaindict.values():
                cc=Counter(domains)
                for i in cc:
                        if i not in coutunters.keys():
                                coutunters[i]=[]
                        coutunters[i].append(cc[i])
        pfam_list.extend(domains)
        pfam_list=list(set(pfam_list))
        print pfam_list
        kk=coutunters.keys()
        kk.sort()
        col_names=[]
        for i in kk:
                for j in range(1, max(coutunters[i]) +1):
                        col_names.append(i + '_' + str(j))
        print '\n'.join(col_names)
        print len(col_names)
        return col_names
```

Print functions

These functions were created to print and parse the data in a correct format.

```
def print_bin(gi, t):
        a =  gi.split('|')[1]
        print a, ' '*(10-len(a)), parse_vec(t),


def parse_vec(t):
        return ''.join([str(i) for i in t])
```

Create_tree_dict function

This function creates a diccionary using GI numbers as keys and binary vectors as values. Data obtanied from this script prints the GI, binary vector and the Pfam domain codes which correspond to the domain composition.

```
def create_tree_dict(domaindict, col_names):
        tree_dict={}
        for gi in domaindict.keys():
                t=[0]*len(col_names)
                dd=domaindict[gi]
                cc=Counter(dd)
                for pf in cc:
                        for j in range(1, cc[pf]+1):
                                counted_pf = pf + '_' + str(j)
                                t[col_names.index(counted_pf)]=1
                tree_dict[gi]=tuple(t)
                print_bin(gi, t)
                print " ".join(domaindict[gi])
        return tree_dict
```

Create_binary_unique function

This function creates a dictionary using the binary vector as keys and a list of all the GI numbers with the same vector as values.

```
def create_binary_unique(tree_dict, col_names):
        done={}
        for gi in tree_dict:
                if tree_dict[gi] not in done:
                        done[tree_dict[gi]]=[]
                done[tree_dict[gi]].append(gi)
        return done
```

Stats function

This function allows to calculates the distribution of each domain architecture reporting the binary vector and the GI number that have the same composition.

```
def stats(tree_dict):
        cc=Counter(tree_dict.values())
        kk=cc.keys()
        kk.sort()
        for k in kk:
                print cc[k],';', parse_vec(k)
```

Get_sequence_for_domain_from_all function

This function allows to obtain the sequence for a particular Pfam domain indicated on the main function introducing the Pfam number.

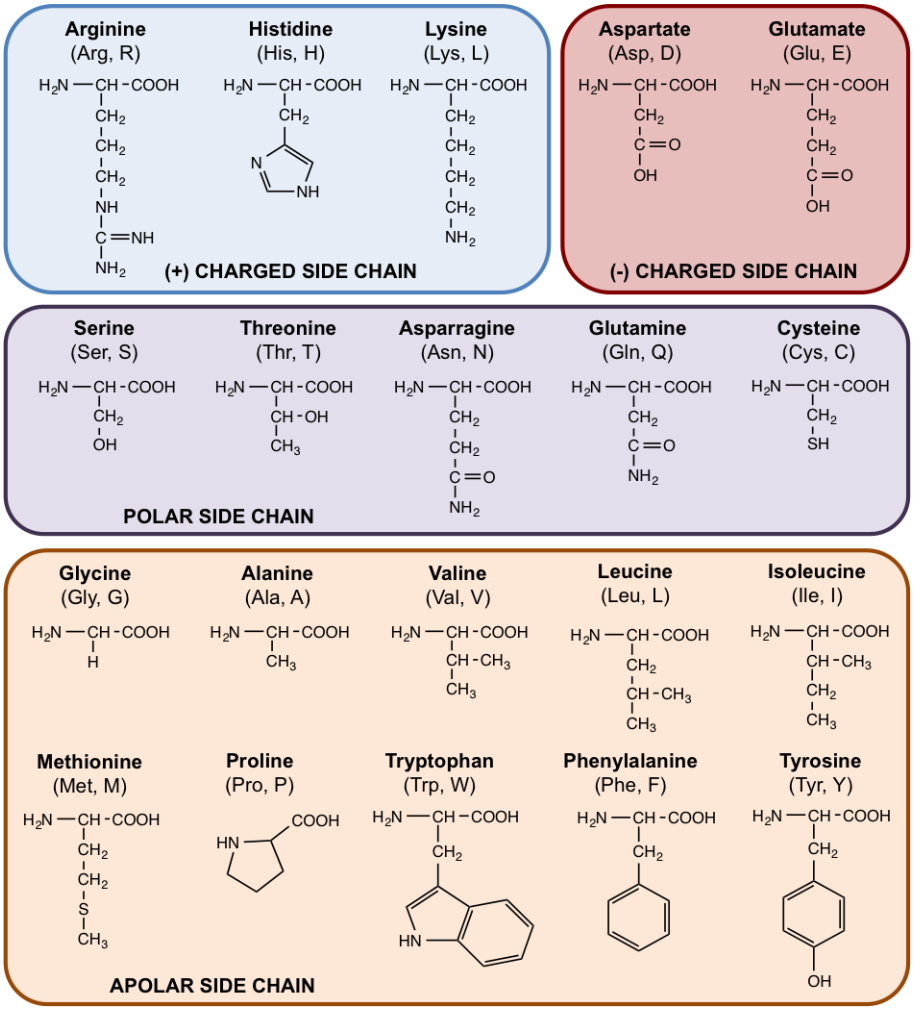
```
def get_sequence_for_domain_from_all(domaindict, domaindict2,
    fasta_dic, binaryToGis, domainName):
        OUT={}
        for vec in binaryToGis.keys():
                print  "\n" , parse_vec(vec)
                for gi in binaryToGis[vec]:
                        for name, pos in zip(domaindict[gi],
                            domaindict2[gi]):
                                if name==domainName:
                                        pos=[int(i) for i in pos]
                                        seq=domain_seq(fasta_dic[
                                            gi], pos)
                                        print ">" + gi + "\n" +
                                            seq
        return OUT
```

Main function

The main function allows to coordinate the above functions and modify certain parameters in order to obtain specific data. In this function the Pfam number of the domain is used to obtain their sequence.

```
def main():
        domaindict, domaindict2=create_domaindict()
        col_names=create_binary_colnames(domaindict)
        tree_dict=create_tree_dict(domaindict, col_names)
        binaryToGis=create_binary_unique(tree_dict, col_names)
        fasta_dic=get_sequece_dict()
        get_sequence_for_domain_from_all(domaindict, domaindict2,
            fasta_dic, binaryToGis , "PFAMnumber")
        stats(tree_dict)
main()
```

B. Chemical structure of amino acids



C. Acronyms and abreviatures

CAZy: Carbohydrate-Active EnZymes.

CE: Cellulose

CBM: Carbohydrate Binding Module.

CH: Chitin.

CLEAs Cross-linked enzymes aggregates.

CLECs Cross-linked enzymes crystals.

CMP Cytidine 5'-Monophosphate.

Ctx: Cholera toxin.

DA: Domain Architecture.

EC: Enzyme Commission.

EcLac: *E. coli* β-galactosidase.

FOS: Fructooligosaccharide.

Gal: Galactose.

GDP Guanidine 5'-Diphosphate.

GH: Glycosyl hydrolase.

GI: Gene Identifier.

Glc: Glucose.

GOS: Galactooligosaccharides.

GTP Guanidine 5'-Triphosphate.

IPTG: Isopropyl β-D-1-thiogalactopyranoside

kDa: kiloDalton.

LNP: Lactase non-persistant.

LP: Lactase persistant.

OD: Optical Density.

PCR: Polymerase chain reaction.

***p*NP-β-Gal:** *p*-nitrophenyl β-D-Galactopyranoside.

SDS-PAGE: Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis.

SCFA: Short Chain Fatty Acid.

TmLac: *Thermotoga maritima* β-galactosidase

UDP: Uridine 5'-Diphosphate.

WT: Wild Type.

D. Publications

Publications related with the thesis:

Talens-Perales D., Górska A., Huson D. H., Polaina J., Marín-Navarro J (2016) Analysis of domain architecture and phylogenetics of GH2 glycosidases: biotechnological implications. *PLoS One* (submited)

Talens-Perales, D., Garrido, D., Almansa, E., Polaina, J.; Marín-Navarro, J. (2016) Fixation of bioactive compounds to the cuticle of Artemia. *PLoS One* (submited).

Talens-Perales, D., Polaina, J., and Marín-Navarro, J., (2016) Enzyme engineering for oligosaccharide biosynthesis. In: Frontier Discoveries and Innovations in Interdisciplinary Microbiology. P. Shukla (ed), pp 9-31. Springer India.

Talens-Perales, D., Polaina, J., Marín-Navarro, J. (2016) Structural dissection of the active site of *Thermotoga maritima* β-galactosidase identifies key residues for transglycosylating activity. *J Agr Food Chem.* 64:2917-2924.

Marín-Navarro, J., Talens-Perales, D., Oude-Vrielink, A., Cañada, FJ., Polaina, J. (2014) Immobilization of thermostable β-galactosidase on epoxy support and its use for lactose hydrolysis and galactooligosaccharides biosynthesis. *World J Microbiol Biotechnol.* 30:989-998.

Other publications:

Talens-Perales, D., Marín-Navarro, J., and Polaina, J. (2016) Functions and Characteristics of Enzymes. In: Encyclopedia of Food and Health. Caballero,

B., Finglas, P., and Toldrá, F. (eds), vol 2, pp.532-538. Elsevier Science, Amsterdam, Netherlands.

Marín-Navarro, J., Roupain, N., Talens-Perales, D., Polaina, J. (2015) Identification and structural analysis of amino acid substitutions that increase the stability and activity of *Aspergillus niger* glucose oxidase. *PLoS One* 7:e0144289.

Marín-Navarro, J., Talens-Perales, D., Polaina, J. (2015) One-pot production of fructooligosaccharides by a *Saccharomyces cerevisiae* strain expressing an engineered invertase. *Appl Microbiol Biotechnol.* 99:2549-2555.