



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**Maestría y Doctorado en Ciencias Bioquímicas**

Interacción entre un dominio de unión al almidón  
de la familia CBM26 con  $\alpha$ -glucanos

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M en C. Silvia Armenta Jaime

TUTOR PRINCIPAL

Dra. Romina Rodríguez Sanoja  
[Instituto de Investigaciones Biomédicas](#)

MIEMBROS DEL COMITÉ TUTOR

Dra. Amelia Farrés González Sarabia  
[Facultad de Química](#)

Dr. Karén Manoutcharian Airapetian  
[Instituto de Investigaciones Biomédicas](#)

Ciudad de México. Febrero, 2019



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



PMDCB/2331/2018

**Armenta Jaime Silvia**  
**Estudiante de Doctorado en Ciencias Bioquímicas**  
**P r e s e n t e**

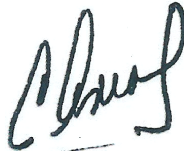
Los miembros del Subcomité Académico en reunión ordinaria del día 17 de septiembre del presente año, conocieron su solicitud de asignación de **JURADO DE EXAMEN** para optar por el grado de **DOCTORADO EN CIENCIAS**, con la réplica de la tesis "**Interacción entre un dominio de unión al almidón de la familia CBM26 con  $\alpha$ -glucanos**", dirigida por el/la Dr(a). **Rodríguez Sanoja Romina María de la Paz**.

De su análisis se acordó nombrar el siguiente jurado integrado por los doctores:

PRESIDENTE	Sánchez Esquivel Sergio
VOCAL	Hernández Santoyo Alejandra
VOCAL	Sosa Peinado Alejandro
VOCAL	Saab Rincón Gloria
SECRETARIO	Coello Coutiño Martha Patricia

Sin otro particular por el momento, aprovecho la ocasión para enviarle un cordial saludo.

**Atentamente**  
**"POR MI RAZA, HABLARÁ EL ESPÍRITU"**  
Cd. Universitaria, Cd. Mx., a 17 de septiembre de 2018.  
**COORDINADORA**



**Dra. ANA BRÍGIDA CLORINDA ARIAS ÁLVAREZ**

contacto: mdcbq@posgrado.unam.mx Tel. 5623 7006

## **FINANCIAMIENTO**

Este trabajo fue financiado por los proyectos UNAM DGAPA-  
PAPIIT IN209316 e IN216419

## **AGRADECIMIENTOS**

A la Universidad Nacional Autónoma de México y al Programa de Doctorado en Ciencias Bioquímicas por el apoyo e infraestructura disponibles para realizar este proyecto.

Agradezco de igual manera al Consejo Nacional de Ciencia y Tecnología (Conacyt) por la beca otorgada.

Agradezco el Programa de Apoyo a los Estudios de Posgrado (PAEP) por el apoyo otorgado para asistir al 13<sup>th</sup> Bratislava Symposium on Saccharides “Recent advances in Glycomics” realizado en Slovakia.

Al apoyo del Laboratorio de Bioquímica y Resonancia Magnética Nuclear de la Universidad Autónoma del Estado de Morelos (LabDP-UAEM) por el acceso al equipo ITC200 (Malvern) y al Instituto de Química de la UNAM por el acceso al espectropolarímetro Jasco J-710 (Jasco Inc., Easton, MD, USA).

Así mismo agradezco al Dr. Daniel Alejandro Guillén Santos por el apoyo técnico y logístico que hizo posible el desarrollo de este proyecto.

Agradezco a la M. en C. Ma Elena Munguía por el soporte técnico para la construcción de la biblioteca combinatoria y el despliegue en fagos.

Agradezco a la Dra. Beatriz Ruiz Villafán por su contribución para el enriquecimiento y desarrollo de este proyecto.

Agradezco a la Dra. Nancy O. Pulido Mayoral por su apoyo en el análisis de resultados obtenidos por ITC.

Finalmente a la M. en C. Patricia de la Torre por el apoyo en la secuenciación de los genes de las proteínas mutantes generadas a lo largo de trabajo.

# ÍNDICE

ABREVIATURAS.....	7
RESUMEN .....	8
ABSTRACT.....	9
1. Introducción.....	10
2. Marco teórico.....	13
2.1 $\alpha$ -glucanos .....	13
2.2. Dominios de unión a carbohidratos .....	16
2.2.1. Importancia Biológica de los CBMs.....	16
2.2.2. Características estructurales .....	18
2.2.3. Interacción de los CBMs con carbohidratos.....	21
2.3. Aplicaciones biotecnológicas y nuevos retos en el uso de CBMs.....	31
2.4. Mutagénesis como estrategia para el estudio de CBMs .....	32
2.4.1. Mutagénesis sitio dirigida .....	32
2.4.2. Mutagénesis aleatoria y procesos de selección .....	32
3. Antecedentes.....	35
4. Justificación .....	37
5. Hipótesis .....	37
6. Objetivos .....	38
6.1. Objetivo general.....	38
6.2. Objetivos específicos.....	38
7. Metodología.....	39
7.1. Estrategia experimental .....	39
7.2. Predicción de un modelo estructural del dominio CBM26 de la $\alpha$ -amilasa de <i>L. amylovorus</i> .....	40
7.3. Mutagénesis sitio dirigida.....	40
7.4. Construcción de una biblioteca combinatoria de un dominio CBM26 del dominio LaCBM26 de la $\alpha$ -amilasa de <i>L. amylovorus</i> .....	41
7.5. Producción de partículas de fagos con variantes del LaCBM26 desplegadas en su superficie .....	43
7.6. Inmovilización de almidón en placas de poliestireno .....	44
7.7. Selección de variantes del dominio LaCBM26 con almidón .....	45
7.8. Clonación, expresión y purificación de las proteínas generadas de forma soluble .....	47

7.9. Electroforesis de afinidad (AGE) .....	48
7.10. Dicroísmo Circular .....	48
7.11. Calorimetría de Titulación Isotérmica (ITC).....	49
8. Resultados.....	50
8.1 Análisis <i>in silico</i> de un módulo CBM26 de la $\alpha$ -amilasa de <i>L. amylovorus</i> .....	50
8.2. Caracterización bioquímica de mutantes generadas mediante mutagénesis sitio-dirigida.....	58
8.3. Caracterización de mutantes del dominio LaCBM26 generadas mediante la construcción de una biblioteca combinatoria .....	64
9. Discusión.....	74
10. Conclusiones.....	82
11. Prospectivas.....	83
12. Referencias.....	84
ANEXOS .....	100

## ABREVIATURAS

CBM	Dominios de Unión a Carbohidratos (Carbohydrate Binding Modules)
SBD	Dominios de Unión al almidón (Starch Binding Domain)
CFS	Sitio de unión en la zona cóncava (Concave Face Site)
VLS	Sitio de unión en la zona variable (Variable Loop Site)
AFM	Microscopía de fuerza atómica (Atomic Force Microscopy)
ITC	Calorimetría de Titulación Isotérmica (Isothermal Titration Calorimetry)
AGE	Electroforesis de Afinidad (Affinity Gel Electrophoresis)
CD	Dicroísmo circular (Circular Dichroism)
RMSD	Raíz de desviación cuadrática media (Root Mean Square Deviation)

## RESUMEN

Los módulos de unión a carbohidratos (CBM, por sus siglas en inglés) unen carbohidratos a través de interacciones CH- $\pi$ , enlaces de puente de hidrógeno e interacciones hidrofóbicas. Se ha establecido que los aminoácidos aromáticos son mediadores de la interacción de dominios de unión al almidón (SBD, por sus siglas en inglés) con el sustrato. Sin embargo, mutaciones puntuales en el sitio de unión de un CBM26 del dominio de unión al almidón de la  $\alpha$ -amilasa de *Lactobacillus amylovorus* (LaCBM26), mostraron que un residuo no aromático conservado también es esencial en la función de reconocimiento del almidón, ya que la mutación de la glutamina 68 por leucina (Q68L) elimina la unión al almidón y a la  $\beta$ -ciclodextrina, incluso en presencia de aminoácidos aromáticos necesarios para la unión del ligando. El análisis de la estructura secundaria de las proteínas mutadas indica que no hay cambios a nivel de estructura secundaria con respecto al dominio silvestre. Por otro lado, mutaciones combinatorias de cinco residuos del sitio de unión (Y18, Y20, Q68, E74 y F77) provocaron la pérdida de la función, lo cual podría estar asociado a cambios a nivel de estructura secundaria. Es probable que las mutaciones en el LaCBM26 sean incompatibles con el plegamiento de estas proteínas, lo que sugiere que el dominio LaCBM26 tiene poca tolerancia al cambio por lo tanto su evolución puede ser limitada.

## ABSTRACT

The carbohydrate-binding modules (CBM) bind carbohydrates through CH- $\pi$  interactions, hydrogen bond interactions and hydrophobic interactions. It has been established that aromatic amino acids are mediators of the interaction of starch binding domains (SBD) with the substrate. However, point mutations in the binding site of a CBM26 of the starch binding domain of the  $\alpha$ -amylase of *Lactobacillus amylovorus* (LaCBM26), showed that a conserved non-aromatic residue is essential in the recognition function of the starch, since the mutation of Glutamine 68 by leucine (Q68L) eliminates binding to starch and  $\beta$ -cyclodextrin, even in the presence of aromatic amino acids necessary for ligand binding. Analysis of the secondary structure of the mutated proteins indicates that there are no changes at the secondary structure level with respect to the wild-type domain. Otherwise, combinatorial mutations of five residues of the binding site (Y18, Y20, Q68, E74 and F77) caused the loss of function, which could be associated with changes at the secondary structure level. It is likely that mutations in LaCBM26 are incompatible with the folding of these proteins, suggesting that the LaCBM26 domain has little tolerance to change and therefore its evolution may be limited.



## 1. Introducción

Los CBMs son dominios no catalíticos, ampliamente distribuidos en bacterias, arqueas y eucariotas. Aunque pueden presentarse como dominios aislados, normalmente forman parte de enzimas multimodulares como glucosiltransferasas y glucósido-hidrolasas, donde su principal función es permitir su interacción con sustratos insolubles, como celulosa, xilano y almidón. Estos dominios han sido clasificados en 88 familias con base en su similitud en secuencia de aminoácidos, de los cuales quince (CBM20, CBM21, CBM25, CBM26, CBM34, CBM41, CBM45, CBM48, CBM53, CBM58, CBM68, CBM69, CBM74, CBM82 y CBM83) se distinguen por su capacidad para reconocer principalmente almidón y/o maltooligosacaridos (Carbohydrate Active Enzymes database: <http://www.cazy.org/>; Lombard et al., 2014).

Estos dominios de unión al almidón (SDBs) se han encontrado en enzimas como las amilasas ya sea de forma aislada o formando tándems, lo cual mejora de forma importante su afinidad por el sustrato al tener un efecto cooperativo (Boraston et al., 2006; Guillén et al., 2007). También se ha asociado la presencia de CBMs en tándems para diversificar la capacidad de reconocer otros sustratos (Motherway et al., 2008) e incluso contribuir e incrementar la degradación de los gránulos de almidón insolubles (Valk et al., 2016, 2015).

Como ejemplo de amilasas bacterianas que tienen SDBs en tándems con más de tres CBMs, se encuentran las  $\alpha$ -amilasas de *Lactobacillus plantarum*, *Lactobacillus manihotivorans* y *Lactobacillus amylovorus*, que presentan cuatro o cinco dominios de la familia CBM26 idénticos en tándem (Giraud and Cuny, 1997). Este arreglo tan

particular no se ha observado en  $\alpha$ -amilasas de otros microorganismos u organismos eucariontes, lo que resulta de interés ya que aún no se conoce la importancia biológica de la presencia de este arreglo en tándems, más allá del incremento de la afinidad al almidón insoluble (Guillén et al., 2007).

Cada módulo CBM26 de la  $\alpha$ -amilasa de *L. amylovorus* (LaCBM26), conformado por cinco repeticiones idénticas en tándem, es capaz de unir almidón insoluble, amilopectina, amilosa, pululano y glucógeno. Por estudios de mutagénesis dirigida se sabe que para la unión al almidón son importantes tres residuos aromáticos, Y18, Y20 y W32; de los cuales el W32 es indispensable para la función, ya que la mutante W32L no es capaz de unir almidón soluble, mientras que las mutantes Y18L y Y20L unen almidón diferencialmente y en menor medida que el dominio silvestre (Rodríguez-Sanoja et al., 2009).

Por otro lado, estudios estructurales indican que seis aminoácidos, W36, Y23, Y25, Q71, G76 y E77 en otro CBM de la familia 26 presente en la amilasa G-6 de *Bacillus halodurans* C125 (BhCBM26) participan en la interacción con maltosa (PDB:2C3H; Boraston et al., 2006). La unión se da, básicamente, a través de interacciones tipo van der Waals con los residuos Y25 y W36, estabilizadas por cinco puentes de hidrógeno formados por residuos aromáticos (Y23), polares (Q71 y E77) y no polares (G76) (Boraston et al., 2006).

En los dominios LaCBM26 y BhCBM26, los aminoácidos de unión relevantes están conservados (a excepción de la G76). Sin embargo, y de forma interesante, existe únicamente un 28% de identidad en la secuencia de aminoácidos de ambos dominios. Esto sugiere que la disposición de los aminoácidos en el sitio de unión es

diferente, y justifica más estudios para determinar el papel que desempeñan los residuos aromáticos y no aromáticos, en la unión al sustrato del dominio LaCBM26. La importancia de los aminoácidos aromáticos como mediadores de la interacción CBM-carbohidrato, ha sido establecida en varias ocasiones (Bharali et al., 2007; Bolam et al., 2001; Cid et al., 2010). Sin embargo, no existe información suficiente que describa el impacto producido durante la sustitución de residuos polares en dominios de unión al almidón.

Por lo tanto, el objetivo de este estudio fue caracterizar la participación de los residuos aromáticos y no aromáticos del dominio LaCBM26 en el reconocimiento de algunos  $\alpha$ -glucanos, mediante análisis termodinámicos de mutantes obtenidas por mutagénesis sitio dirigida y mutagénesis aleatoria.

## 2. Marco teórico

### 2.1 $\alpha$ -glucanos

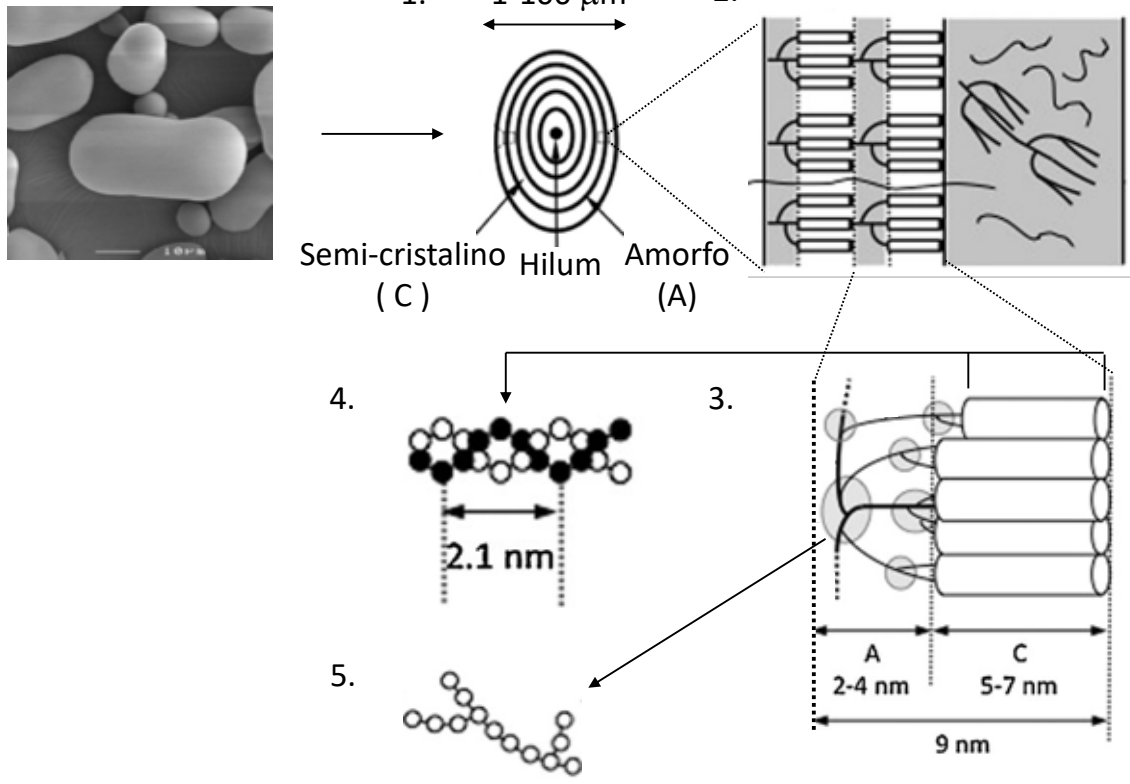
Los carbohidratos están ampliamente distribuidos en la naturaleza. Los polímeros de glucosa se dividen dependiendo del enlace que forman con el oxígeno del carbono anomérico en  $\beta$ -glucanos o  $\alpha$ -glucanos. Entre los  $\alpha$ -glucanos, el almidón es el polisacárido más abundante.

Los gránulos de almidón son insolubles en agua y varían de tamaño (1-100 $\mu$ m), forma y carga, en función de su origen botánico. Éstos tienen de forma alternada zonas amorfas y semi-cristalinas. La estructura de las zonas amorfas es poco conocida, pero se estima que la mayor cantidad de amilosa presente en el gránulo se encuentra en esta zona, además de la amilopectina. Los anillos semicristalinos consisten en varias láminas amorfas y cristalinas alternadas aproximadamente cada 9 nm debido principalmente a la amilopectina, aunque cadenas de amilosa podrían estar involucradas (Figura 1; Vamadevan and Bertoft, 2015).

La amilosa constituye del 15-25% del almidón, son cadenas lineales de glucosas unidas mediante enlaces  $\alpha$ -1,4-glucosídicos, aunque pueden contener hasta 0.1% de ramificaciones presentes como enlaces  $\alpha$ -1,6-glucosídicos. En contraste, la amilopectina está conformada de enlaces  $\alpha$ -1,4-glucosídicos con una proporción del 5-6% de ramificaciones por la presencia de enlaces  $\alpha$ -1,6-glucosídicos.

De forma similar, el glucógeno, la principal forma de almacenamiento de glucosa en mamíferos, está formado por cadenas de glucosa unidas por enlaces glucosídicos  $\alpha$ -1,4- y  $\alpha$ -1,6-, pero a diferencia del almidón, tiene mayor frecuencia de ramificaciones, cerca del 9% (Buléon et al., 1998; Kossmann and Lloyd, 2000).

Gránulo de almidón

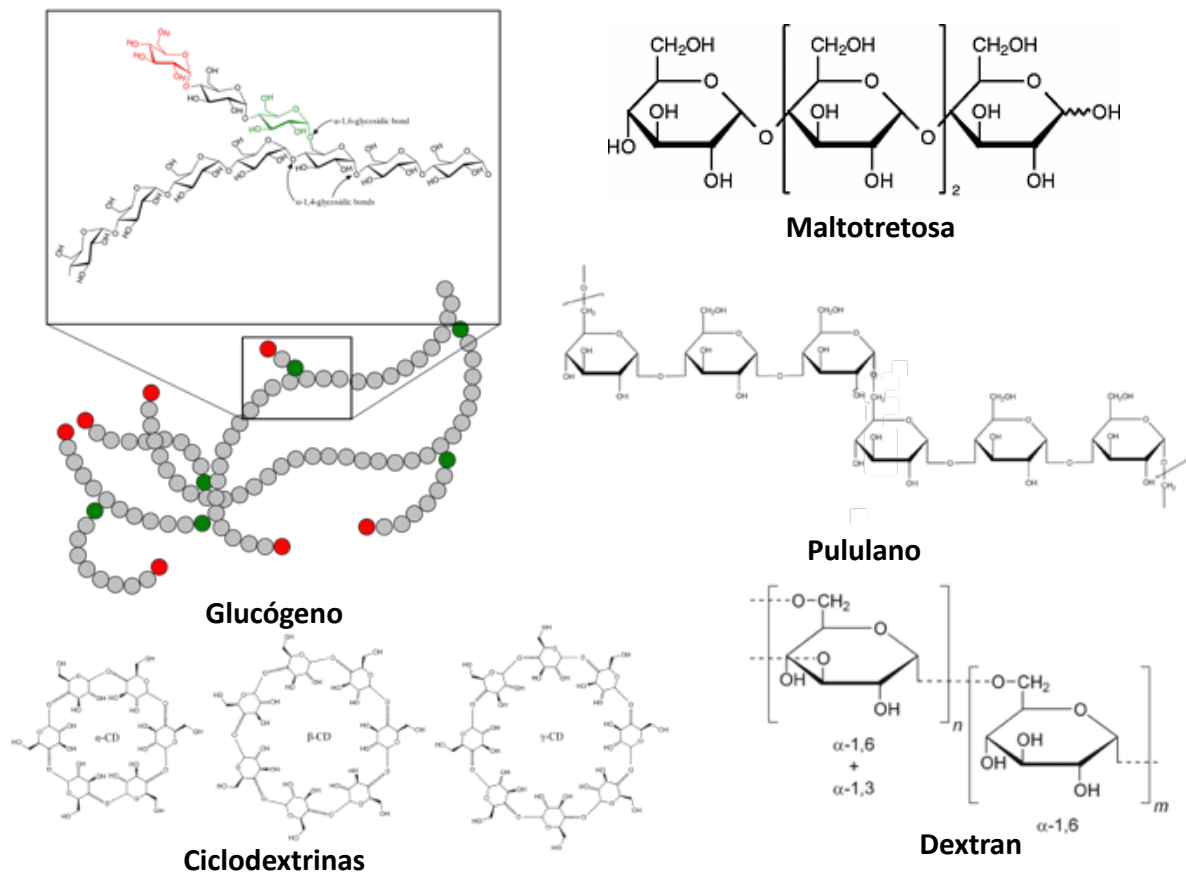


**Figura 1.** Características estructurales de los gránulos de almidón. Micrografía de gránulos de almidón de papa (imagen izquierda). 1-5) Representación esquemática de los diferentes niveles estructurales en un gránulo de almidón. 1) Estructura laminar, indicando como se intercalan las zonas amorfas (A) y semi-cristalinas (C). 2 y 3) Acercamiento de las zonas A y C, que también indica que la última capa del gránulo es principalmente amorfa. 3) Zona semicristalina formada principalmente por amilopectina formando dobles hélices (4), mientras que en la zona amorfa principalmente hay ramificaciones (5). Adaptado de Vamadevan y Bertorft, 2015.

Otros  $\alpha$ -glucanos importantes son el pululano, que está formado por unidades de maltotriosa, tres unidades de glucosa unidas por un enlace  $\alpha$ -1,4-glucosídico, unidas por un enlace  $\alpha$ -1,6-glucosídico; y el dextran, que es un polisacárido de

unidades de glucosa unidas principalmente por enlaces  $\alpha$ -1,6-glucosídico con ramificaciones de enlaces  $\alpha$ -1,2-,  $\alpha$ -1,3- y  $\alpha$ -1,4-glucosídico en proporciones variables (Figura 2; Xu et al., 2014).

Dentro de los  $\alpha$ -glucanos también se encuentran los productos de degradación de los polisacáridos antes mencionados como malto-oligosacáridos de diferentes longitudes y otros sintetizados por conversión enzimática a partir de almidón, como las ciclodextrinas ( $\alpha$ -,  $\beta$ -,  $\gamma$ -,  $\delta$ -ciclodextrina; Figura 2).



**Figura 2.** Estructura de algunos de los principales alfa-glucanos.

Aunque, todos estos carbohidratos están formados básicamente por enlaces  $\alpha$ -1,4- y  $\alpha$ -1,6-glucosídicos, muchas de las propiedades y características funcionales de

los mismos dependen en gran medida de la organización y longitud de las cadenas de glucosa. Por lo que, la estructura y organización tridimensional de estos carbohidratos generalmente es compleja.

## **2.2. Dominios de unión a carbohidratos**

### **2.2.1. Importancia Biológica de los CBMs**

Los módulos de unión a carbohidratos (CBM, por sus siglas en inglés) son dominios no catalíticos, definidos como módulos de anclaje que principalmente forman parte de enzimas multimodulares relacionadas al metabolismo de polisacáridos. Funcionalmente, estos dominios permiten la interacción entre el sustrato insoluble y la enzima en solución, lo que incrementa la concentración del sustrato en el sitio activo del dominio catalítico (Rodríguez-Sanoja et al., 2005), lo que permite aumentar la hidrólisis de sustratos insolubles; mantener el dominio catalítico en proximidad al sustrato; promover la desorganización estructural de carbohidratos complejos; y anclar proteínas en la superficie celular (Guillén et al., 2010).

Generalmente, los CBMs se encuentran en la naturaleza en una amplia diversidad de enzimas como las glucósido-hidrolasas y glucósido-transferasas, entre otras, (CaZy, [www.cazy.org](http://www.cazy.org)) en las cuales, su importancia va más allá de permitir la interacción de la enzima con el sustrato, ya que también se les atribuye la capacidad de modular las propiedades catalíticas de la enzima. Por ejemplo, en la enzima almidón sintasa III (SSIII) de *Arabidopsis thaliana*, la capacidad catalítica de la enzima es modulada por los SBDs (CBMs de la familia 48) en el amino terminal (Valdez et al., 2008).

Por otro lado, la desorganización de sustratos insolubles por CBMs se ha demostrado en quitina, celulosa y almidón. El dominio CBM33 de una quitinasa de *Serratia marcescens* une quitina insoluble, provocando cambios estructurales en el sustrato y aumenta la accesibilidad al mismo (Vaaje-Kolstad et al., 2005).

De forma similar, el dominio de unión a celulosa presente en la endoglucanasa (CenA) de *Cellulomonas fimi*, provoca cambios estructurales en fibras de celulosa (Din et al., 1991). También se demostró por microscopía de fuerza atómica (AFM) que la estructura helicoidal de la amilosa tipo III, se modifica por acción del dominio CBM21 de la glucoamilasa de *Rhizopus oryzae* (RoSBD), el cual tiene dos sitios de unión, mismos que participan en la unión y desorganización estructural de cadenas de amilosa. Los dominios de RoSBD aparentemente gira las hebras de amilosa helicoidal para exponer mayor superficie y aumentar la unión del dominio, y en consecuencia aumentar la susceptibilidad a la hidrólisis por la enzima glucoamilasa (Jiang et al., 2012).

Por otra parte, se ha demostrado que algunos CBMs permiten la adhesión a bacterias, debido a que éstos unen polisacáridos presentes en la pared celular bacteriana. Por ejemplo, se analizaron las propiedades de unión de tres CBMs de la familia 37 de *Ruminococcus albus*, lo cual mostró que el CBM37, que une celulosa, también es capaz de unirse a la superficie de *R. albus* (Ezer et al., 2008). También, Salama-Alber et al., (2013) demostraron que el sistema celulósico de *Ruminococcus flavefaciens* está anclado a la pared celular bacteriana a través de una subunidad de andamio ScaE relativamente pequeña, conformada de una cohesina de tipo IIIe responsable de la unión de dos proteínas de andamio: ScaB y



la proteína de unión a celulosa CttA. Esta última, media la unión al sustrato bacteriano a través de dos módulos de unión a carbohidratos.

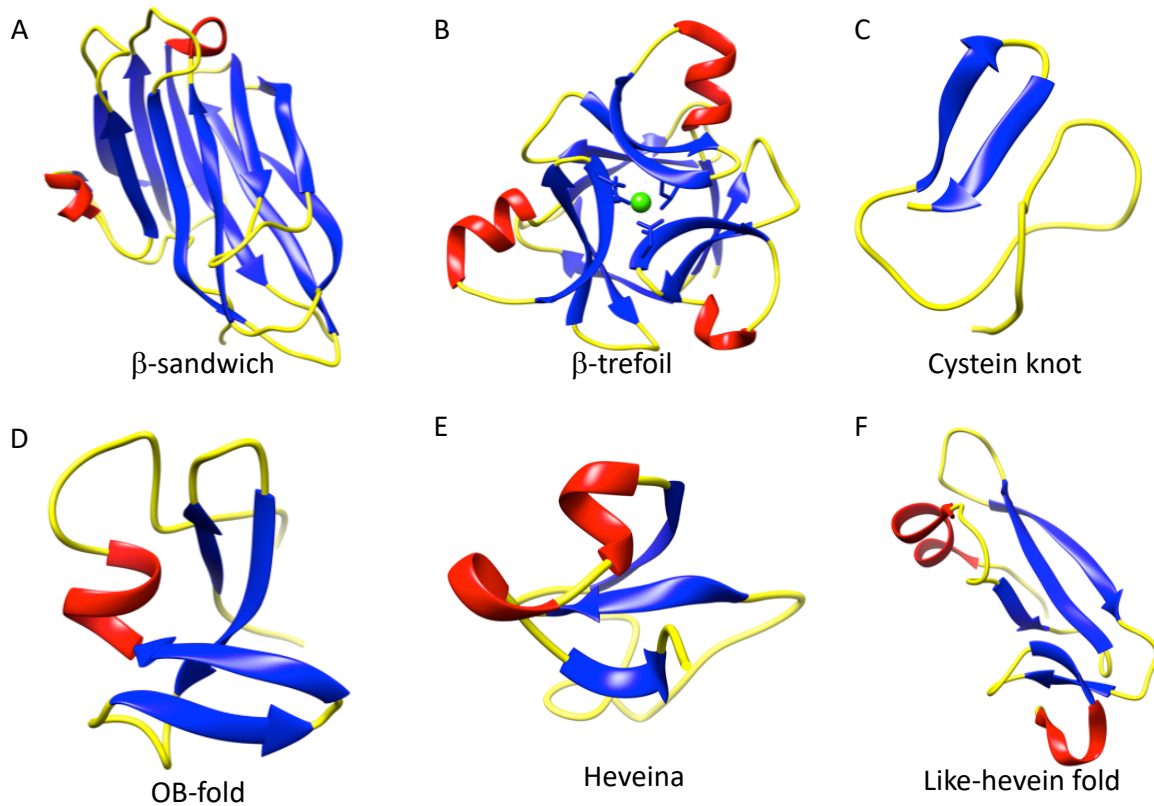
En consecuencia, conocer y comprender las funciones naturales de los CBMs en las interacciones proteína-carbohidrato en los diferentes sistemas biológicos puede permitir desarrollar aplicaciones biotecnológicas, por ejemplo, el desarrollo de biocombustibles, biomateriales, entre otras, y contribuir a la ingeniería molecular de estas proteínas.

### **2.2.2. Características estructurales**

Con base en la similitud de la secuencia de aminoácidos, los CBMs se clasifican en 88 familias dentro de la base de datos CaZy ([www.cazy.org](http://www.cazy.org)), en la cual se incorporan constantemente nuevas secuencias y familias (Lombard et al., 2014). Estas familias engloban dominios con la capacidad de unir a prácticamente todos los carbohidratos existentes en la naturaleza, desde polisacáridos estructurales y de reserva, hasta factores de virulencia (Guillén et al., 2010).

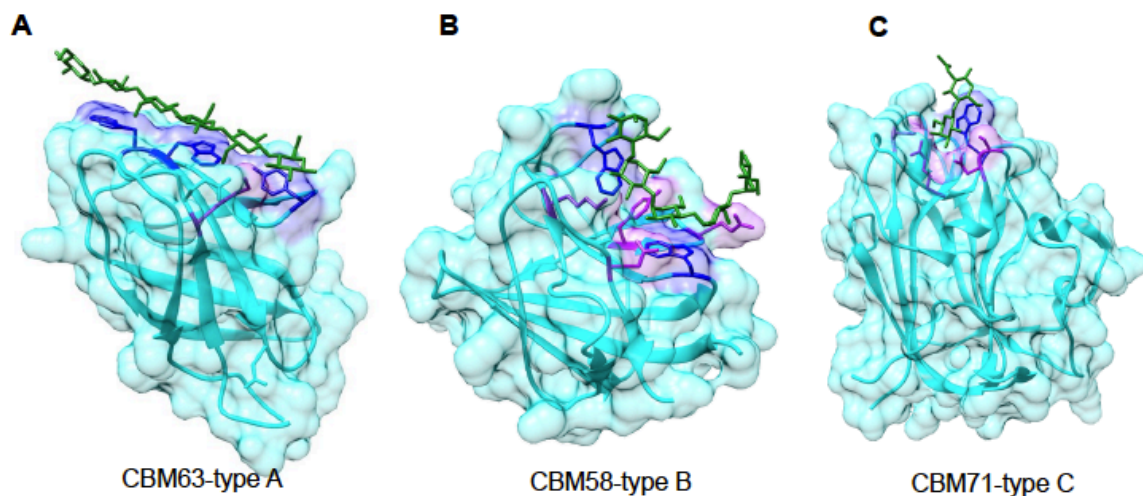
Hasta el momento, son siete diferentes plegamientos los reportados para los CBMs, entre los que destacan el barril beta distorsionado ( $\beta$ -sándwich),  $\beta$ -trefoil, Cysteine knot, OB fold, Hevein fold y like-hevein fold. Sin embargo, la estructura de  $\beta$ -sándwich es la predominante entre los distintos CBMs (Figura 3; Boraston et al., 2004; Hashimoto, 2006). En la mayoría de los CBMs el sitio de unión a carbohidratos está localizado hacia la cara cóncava (CFS, por sus siglas inglés) de la estructura de  $\beta$ -sandwich, aunque también puede ser localizado hacia uno de los extremos del

dominio entre una región predominantemente desordenada, por lo que se conoce como variable (VLS, por sus siglas en inglés; Abbott and van Bueren, 2014).



**Figura 3.** Principales plegamientos que adoptan los CBMs. A) Dominio de unión a xilano (CBM4-2) de la xilanasa 10A de *Rhodothermus marinus* (PDB ID 1K45), B) Dominio CBM42 de *Clostridium thermocellum* (PDB ID 3KMV), C) Dominio de unión a celulosa (CBM1) de la Celobiohidrolasa I de *Trichoderma reesei* (PDB ID 1AZ6), D) Dominio de unión a celulosa (CBM10) de la xilanasa A de *Pseudomonas fluorescens* sp. *cellulosa* (PDB ID 1E8R), E) Heveina (CBM18) (PDB ID 1HEV), F) Proteína de avirulencia 4 (CBM14) de *Pseudocercospora fuligena* (PDB ID 4Z4A). Representación esquemática que muestra los elementos de estructura secundaria: hebras beta (azul), hélices (rojo) y giros (amarillo). Las estructuras fueron generadas con el software Chimera (<https://www.cgl.ucsf.edu/chimera>).

La unión de carbohidratos depende de la apropiada topología que complementa la estructura del sustrato y de la orientación de los aminoácidos conservados localizados en el sitio de unión de los CBMs. La topología del sitio de unión puede ser de tipo planar, conocida como topología A, la cual es característica de los CBMs que tienen la capacidad de unir celulosa (Boraston et al., 2004; Georgelis et al., 2012; Gilbert et al., 2013); también destaca la topología en forma de canal o hendidura, conocida como topología tipo B, característica de los CBMs que unen cadenas de oligosacáridos como xilano o amilosa, entre otros (Boraston et al., 2004; Gilbert et al., 2013; Luís et al., 2013; Simpson et al., 2002); y la topología tipo C que tienen forma de cavidad y que permite la unión a carbohidratos pequeños como mono-, di-, y tri-sacáridos, principalmente (Figura 4; Boraston et al., 2004, 2001; Gilbert et al., 2013). Cambios sutiles en la topología del sitio de unión determinan la especificidad por el ligando y explican por qué los CBMs con aparente estructura similar reconocen diferentes sustratos.



**Figura 4.** Diferentes tipos de topología del sitio de unión en CBMs. A) Topología tipo A: CBM63 de la expansina EXLX1 de *Bacillus subtilis* en complejo con celohexosa (PDB ID

4FER), B) Topología tipo B: CBM58 de la  $\alpha$ -amilasa *Bacteroides thetaiotaomicron* en complejo con maltoheptosa (PDB ID 3K8L), C) Topología C: CBM71 de la  $\beta$ -galactosidasa (BgaA) de *Streptococcus pneumoniae* en complejo con  $\beta$ -D-galactopiranosil-1,4-N-acetil-D-glucosamina (PDB ID 4CUB). Aminoácidos aromáticos que establecen interacciones tipo CH- $\pi$  se muestran en azul, los aminoácidos que establecen enlaces de puente de hidrógeno se muestran en magenta, los sustratos se muestran en verde en todos los casos. Las estructuras fueron generadas con el software Chimera (<https://www.cgl.ucsf.edu/chimera>). Tomada de Armenta et al., 2017.

### **2.2.3. Interacción de los CBMs con carbohidratos**

Los carbohidratos y las proteínas son de las principales biomoléculas que existen en la naturaleza, por lo que su interacción domina una amplia cantidad de funciones esenciales para la vida, por ejemplo, la adhesión celular, procesos de infección, diferentes aspectos de la respuesta inmune, comunicación celular, entre muchas otras (Gabiús et al., 2011).

La unión entre estas biomoléculas está dada principalmente por tres diferentes interacciones: 1) interacciones polares, principalmente enlaces de puente de hidrógeno y 2) interacciones CH- $\pi$ , además de 3) interacciones hidrofóbicas.

Los enlaces de puentes de hidrógeno son una interacción ion-dipolo entre un átomo electronegativo, como el nitrógeno y oxígeno, con un átomo de hidrógeno que requiere aprox. 5-30 kJ/mol de energía. Debido a los numerosos grupos -OH en los polisacáridos, estos enlaces se establecen de manera cooperativa, comprometiendo simultáneamente las propiedades donante/aceptor de un grupo hidroxilo, y las interacciones bidentadas con las cadenas laterales de aminoácidos

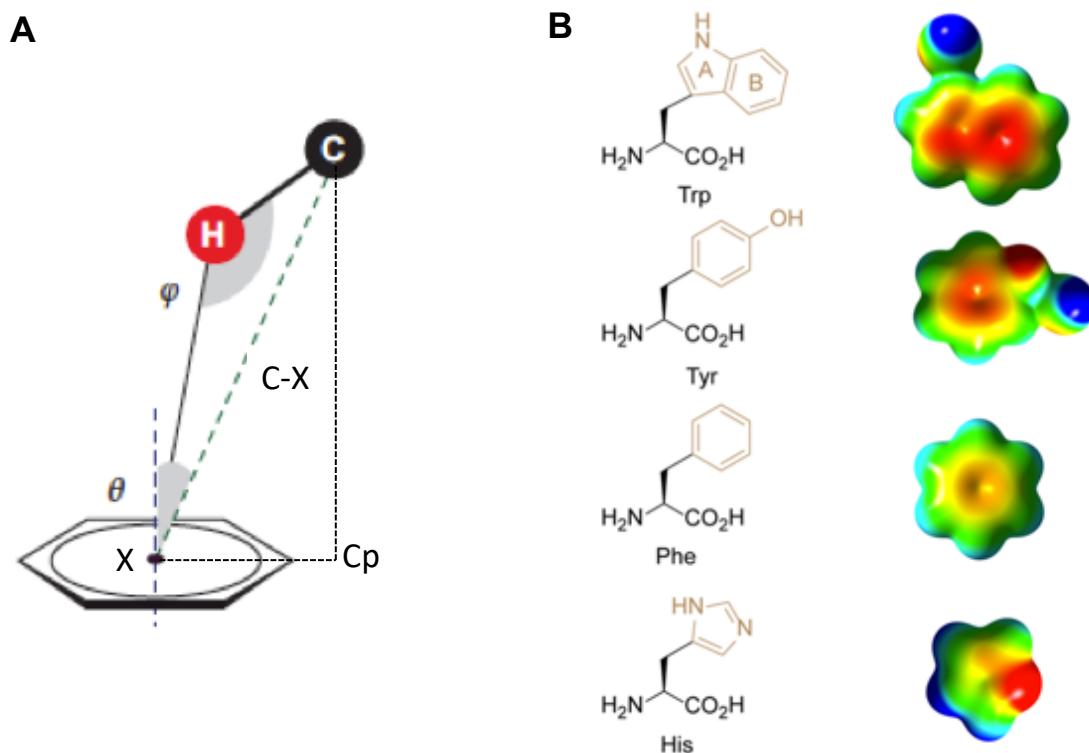
polares como ácido aspártico y glutámico, asparagina, glutamina, arginina y serina, entre otros (Gabiús et al., 2011). Además, los enlaces de puente de hidrógeno también se establecen tanto con los grupos hidroxilo, como con grupos amino y carboxilo de los aminoácidos.

Los enlaces de puentes de hidrógeno son importantes durante la interacción CBM-carbohidrato, debido a su contribución en la afinidad, así como en la selectividad, siendo en muchos casos la disposición estereoquímica característica de los grupos hidroxilo lo que le da a una proteína su especificidad hacia algún tipo de polisacárido (Fernández-Alonso et al., 2012).

La interacción CH- $\pi$ , se puede clasificar entre un enlace de hidrógeno convencional e interacciones más débiles dominadas por fuerzas de dispersión, que dependen de la polaridad de un grupo donador. Por tanto, esta interacción involucra un componente de dispersión o de van der Waals, y un componente electrostático, siendo este último el de menor contribución. En sistemas modelo como benzeno/metano (CH/ $\pi$ ) o agua (OH/ $\pi$ ) se ha predicho que tiene una energía de interacción entre 4-16 kJ/mol (Tsuzuki et al., 2000).

Hudson y col. (2015) establecieron algunas características que definen la interacción CH- $\pi$  en proteínas, tales como tener un ángulo theta ( $\theta$ )  $\leq 40^\circ$ , un ángulo psi ( $\varphi$ ) entre  $120^\circ$  y  $180^\circ$ , una distancia (C-X)  $\leq 4.5 \text{ \AA}$  y una distancia de proyección (Cp-X)  $\leq 1.6\text{-}2.0 \text{ \AA}$  (Figura 5). La interacción más fuerte ocurre cuando el enlace CH-interacciona directo al centro del anillo aromático (es decir,  $\theta = 0^\circ$  y  $\varphi = 180^\circ$ ). Cambiar el valor de  $\varphi$  o el desplazamiento lateral del grupo donante (definido por  $\theta$ ), sólo reduce ligeramente la energía de interacción. Solo cuando aumenta la

distancia donante-aceptor (C-X), la energía de interacción global disminuye. Por lo tanto, la ubicación del carbono determina la energía de interacción en lugar de la ubicación del protón. (Plevin et al., 2010).



**Figura 5.** Características que definen la interacción CH- $\pi$  y los aminoácidos que participan. A) Parámetros para identificar la interacción CH- $\pi$  :  $\theta \leq 40^\circ$ ,  $120^\circ < \varphi < 180^\circ$ ,  $C-X \leq 4.5 \text{ \AA}$  y  $Cp-X \leq 1.6-2.0 \text{ \AA}$ . B) Representación superficial que muestra el potencial electrostático del sistema  $\pi$  de la cadena lateral de los aminoácidos triptófano, tirosina, fenilalanina e histidina. Escala:  $\geq 130 \text{ kJ/mol}$  (electropositivo, azul), neutro (verde) y  $-130 \text{ kJ/mol}$  (electronegativo, rojo). Adaptada de Hudson et al. (2015).

Por otro lado, se han establecido cuatro residuos como importantes en la interacción CH- $\pi$ , que en orden de importancia son: Triptófano>Tirosina>Fenilalanina>Histidina. La principal contribución, y el que más

frecuentemente se encuentra en este tipo de interacción, es el triptófano. El triptófano presenta una superficie de densidad electrónica más electronegativa que la tirosina y la fenilalanina (Figura 5), a través de la cual interacciona con los carbohidratos que presentan elementos CH- formando la interacción CH- $\pi$ . La histidina es complicada por la diferencia de enlaces de hidrógeno y estados de protonación que puede tomar; sin embargo, también puede estar involucrada en la interacción CH- $\pi$  en los complejos proteína-carbohidrato (Hudson et al., 2015).

La contribución de las cadenas laterales de aminoácidos aromáticos ocurre a través de interacciones CH- $\pi$  y el efecto hidrofóbico que se genera durante la interacción donde los residuos aromáticos, frecuentemente, están apilados con el anillo de piranosa del carbohidrato. Tal contacto típicamente involucra dos o tres grupos CH- de las piranosas y la densidad electrónica  $\pi$  del anillo aromático. Estas interacciones pueden dar lugar a una variedad de geometrías, que puede incluir un arreglo paralelo o antiparalelo del aromático y las unidades del carbohidrato (Asensio et al., 2013; Chen et al., 2013).

Por otro lado, las interacciones hidrofóbicas se deben a la agrupación de tres o más grupos CH- adyacentes, producto de la disposición estérica característica de los grupos hidroxilo, lo que crea parches hidrofóbicos en la superficie de los carbohidratos que pueden establecer interacciones hidrofóbicas con aminoácidos no polares tales como leucina, isoleucina, valina, fenilalanina, entre otros. En solventes polares, como el agua, los componentes hidrofóbicos no interaccionan con el solvente, pero provocan que el orden de las moléculas aumente, lo cual disminuye la entropía del sistema, sin que esto represente cambios en la entalpía.

Un segundo componente de las interacciones hidrofóbicas son las fuerzas de dispersión de London, entre las cuales están las fuerzas de van der Waals, en las que pueden aparecer dipolos instantáneos que establecen interacciones que de forma individual son débiles, pero de forma cooperativa pueden ser muy poderosas (Fernández-Alonso et al., 2012).

Cabe señalar que la interacción CBM-carbohidrato ocurre a pesar de no tener características estructurales específicas o únicas, es decir, desde el punto de vista de la proteína, diferentes arquitecturas del sitio de unión suelen presentarse, dependiendo del número y localización de los residuos aromáticos.

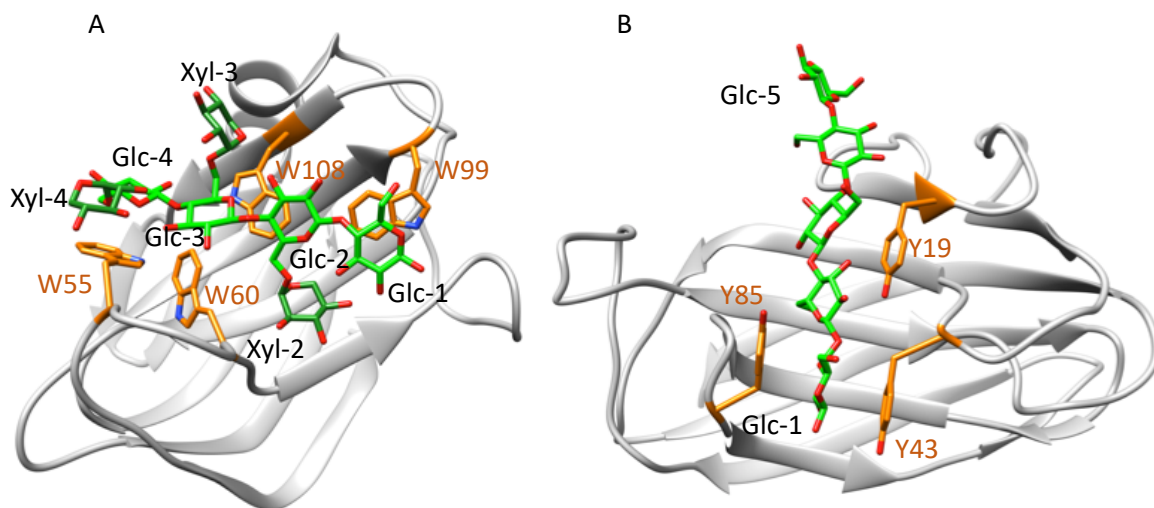
Por lo tanto, se han establecido tres puntos importantes en la interacción CBM-carbohidrato: 1) La presencia de residuos aromáticos es muy importante en el sitio de unión, sin embargo no todos estos aminoácidos son equivalentes; 2) Existe cierta preferencia por residuos polares en el sitio de unión para la formación de enlaces de hidrógeno, siendo los más encontrados el ácido aspártico y la asparagina; 3) La presencia de residuos alifáticos está desfavorecida en los sitios de unión (Gabiús et al., 2011).

#### **2.2.3.1. Papel de los aminoácidos aromáticos en el sitio de unión de los CBMs**

Los aminoácidos aromáticos del sitio de unión son muy importantes en la interacción con carbohidratos, estos forman apilamientos con los sacáridos y establecen interacciones CH- $\pi$  e hidrofóbicas, tipo van der Waals. Varios estudios sugieren que la posición y orientación de estos residuos determina las diferentes afinidades y especificidades entre las distintas familias de CBMs (Boraston et al., 2004).



Por ejemplo, el dominio CBM65A de *Eubacterium cellulosolvens*, tiene un sitio de unión con cuatro residuos aromáticos (W55, W60, W99, W108), cada uno con diferente orientación capaz de participar de forma particular en presencia de sustratos como xiloglucano. Mutagenizar cada uno de estos aminoácidos por alanina disminuyó diferencialmente la afinidad por xiloglucano (W55A,  $k_a$   $9.6 \times 10^4$   $M^{-1}$ ; W60A,  $k_a$   $3.1 \times 10^3$   $M^{-1}$ ; W99A,  $k_a$   $5.0 \times 10^3$   $M^{-1}$ ) e incluso la eliminó (W108A) en comparación con el dominio CBM65A silvestre ( $k_a$   $1.7 \times 10^5$   $M^{-1}$ ). Sin embargo, estos aminoácidos aromáticos pueden adoptar cambios de orientación que permiten la unión de otros sustratos como la celohexosa (Figura 6; Luís et al., 2013).



**Figura 6.** Interacción de aminoácidos aromáticos con polisacáridos. A) CBM65A de *Eubacterium cellulosolvens* en complejo con xiloglucano. B) CBM4-1 de la celulasa (Cel9B) de *Cellulomonas fimi* en complejo con celopentosa. La estructura global de cada dominio se muestra en color gris, los sustratos en color verde y los aminoácidos aromáticos que participan en la interacción en color naranja. Las estructuras fueron generadas con el software Chimera (<https://www.cgl.ucsf.edu/chimera>).

Otro ejemplo es el dominio CBM2b-2 de la xilanasas 11A de *C. fimi*, el cual tiene dos triptófanos, W570 y W602, en orientación perpendicular el uno del otro, ambos indispensables en la unión de xilano ( $5.6 \times 10^4 \text{ M}^{-1}$ ) y celulosa ( $<10^2 \text{ M}^{-1}$ ). De forma interesante, en este dominio cambiar la arginina 573 (adyacente al W570) a glicina alteró la especificidad de unión, ya que de unir xilano unió preferencialmente celulosa (Bolam et al., 2001). Esto debido a que la mutación provocó un cambio en la orientación de W570 adoptando una forma planar, como lo requieren los CBM de topología A que unen celulosa.

Por lo tanto, aminoácidos como el triptófano son importantes en el sitio de unión de los CBMs. Adicionalmente es de importancia la participación de tirosinas en el sitio de unión, como sucede en el dominio CBM4-1 de la celulasas (Cel9B) de *C. fimi*, que tiene tres tirosinas en el sitio de unión (Y19, Y85 y Y43) que establecen interacciones hidrofóbicas. De éstas, las tirosinas 85 y 43 se apilan con la primer glucosa (Glc1) de la celopentosa (Figura 6B; Boraston et al., 2002).

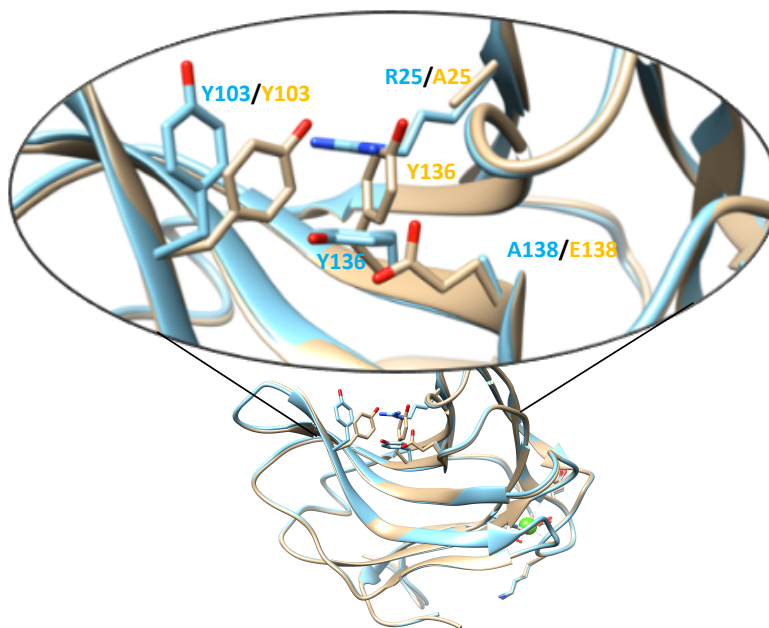
Aunque generalmente el mayor efecto en la afinidad se obtiene al mutar triptófanos del sitio de unión, existen algunas excepciones, donde el mayor efecto en la unión se obtiene al mutar tirosinas, como ocurre con el dominio CBM11 de la celulasas bifuncional del celulosoma de *Clostridium thermocellum* que tiene tres aminoácidos aromáticos, W65, Y22 y Y129. La mutación Y22A eliminó la capacidad de unir glucanos con enlaces  $\beta$ -1,3 y  $\beta$ -1,4; la mutación Y129A por su parte disminuye la afinidad por estos polisacáridos, mientras que el menor efecto lo presenta la mutación W65A (Bharali et al., 2007).

### **2.2.3.2. Papel de los aminoácidos polares y iones en el sitio de unión de los CBMs**

Aunque la orientación y posición de los aminoácidos aromáticos en el sitio de unión de los CBMs muchas veces gobierna la interacción con los carbohidratos, otras interacciones, incluidas los enlaces de puentes de hidrógeno y los enlaces de coordinación mediados por calcio, también tienen un papel relevante en la unión de los sustratos (Abbott and van Bueren, 2014; Armenta et al., 2017; Boraston et al., 2004; Gilbert et al., 2013).

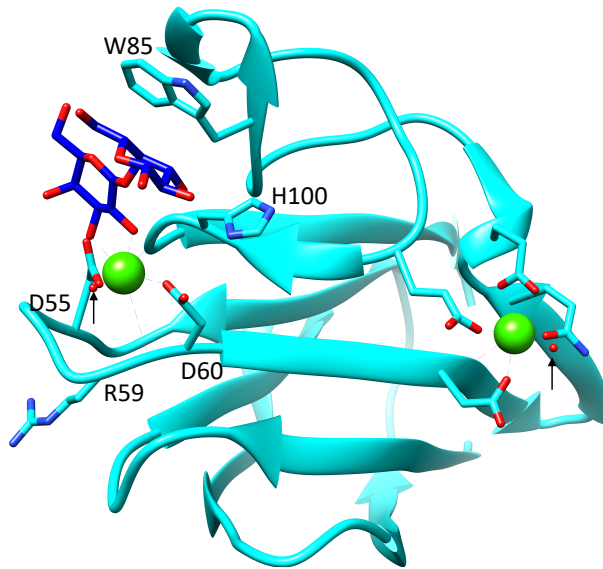
Se ha observado, que los enlaces de puente de hidrógeno establecidos por los aminoácidos polares tienen mayor importancia en función del tipo de topología del sitio de unión. En los CBMs con topología tipo A, mutar aminoácidos polares no representa un efecto importante en la unión de celulosa (McLean et al., 2002).

Por el contrario, en los CBMs con topología tipo B o C, mutar aminoácidos polares del sitio de unión tiene, al parecer, una mayor repercusión en la unión. Por ejemplo, el dominio CBM22-2 de la xilanasasa (Xyn10B) de *C. thermocellum* tiene varios residuos conservados que están implicados en la unión al sustrato. En este dominio, los aminoácidos W53, Y103 y E138 tienen un importante papel en la interacción del CBM22-2 con xilano, mientras que Y136 y R25 además de estar involucrados en la unión, tienen importancia en la integridad estructural del sitio de unión. De forma interesante, la mutación de la R25A provoca que dos aminoácidos aromáticos, Y103 y Y136, cambien su orientación (Figura 7) reduciendo hasta un grado de magnitud ( $\sim 10^2 \text{ M}^{-1}$ ) su afinidad por xilohexosa (Xie et al., 2001).



**Figure 7.** Importancia de aminoácidos polares en la unión de polisacáridos. Alineamiento tridimensional de dos mutantes puntuales [R25A (PDB ID 1H6X, color crema) y E138A (PDB ID 1H6Y, color azul)] del dominio CBM22-2 de la xilanasas 10B de *Clostridium thermocellum*. En la imagen se muestra un acercamiento a los aminoácidos importantes en el sitio de unión. Las estructuras fueron generadas con el software Chimera (<https://www.cgl.ucsf.edu/chimera>).

Por otro lado, en algunas familias de CBMs (ej. CBM36, CBM60, CBM35, CBM6, CBM62) iones como el calcio ( $\text{Ca}^{2+}$ ) tienen importancia estructural, así como en la unión de polisacáridos. Por ejemplo, el CBM60 de la endo- $\beta$ -1,4-xilanasas de *Cellvibrio japonicus* presenta especificidad por celulosa, xilano, y  $\beta$ -1,4-galactano. Durante la unión con  $\beta$ -1,4-galactano, un átomo de  $\text{Ca}^{2+}$  establece enlaces de coordinación con R59 O, H100 O, D55 O $\delta$ 1, D60 O  $\delta$ 2, completando el octaedro de coordinación con el O2 y O3 de una molécula de galactosa y una molécula de agua (Figura 8; Montanier et al., 2010).



**Figura 8.** Importancia del ion  $\text{Ca}^{2+}$  en CBMs. vCBM60 de *Cellvibrio japonicus* en complejo con  $\beta$ -1,4-galactano (PDB ID 2XFE). Se muestran los aminoácidos que participan en la unión formando enlaces de coordinación con un ión  $\text{Ca}^{2+}$ , así como los que participan de forma directa con el sustrato (W85). Con la flecha se indican las moléculas de agua que completan el octahedro del ión  $\text{Ca}^{2+}$  que participa en la unión del sustrato, así como el que es estructural. Las estructuras fueron generadas con el software Chimera (<https://www.cgl.ucsf.edu/chimera>).

De la misma forma, el dominio CBM35 de una ramnogalacturon liasa de *C. thermocellum*, tiene un ion  $\text{Ca}^{2+}$  en el sitio de unión con el cual se establecen enlaces de coordinación al unir carbohidratos. También, tiene iones  $\text{Ca}^{2+}$  involucrados en la dimerización (Dhillon et al., 2018) de forma similar como ocurre en el dominio CBM62 de *C. thermocellum* (Montanier et al., 2011).

### **2.3. Aplicaciones biotecnológicas y nuevos retos en el uso de CBMs**

Los CBMs han sido empleados en novedosas aplicaciones biotecnológicas (Oliveira et al., 2015). Algunas de las cuales son como tallo de purificación a celulosa (Kavoosi et al., 2007; Ramos et al., 2010; Sugimoto et al., 2012) y almidón (Guillén et al., 2013; Lin et al., 2009). Debido a su plegamiento independiente, la fusión de distintas proteínas a algunos CBMs ha permitido expandir las aplicaciones en las que pueden emplearse, por ejemplo, en el desarrollo de tecnologías para el tratamiento de agua contaminada con atrazina (Kauffmann et al., 2000); la fusión al dominio catalítico de una cutinasa termoestable de *Thermobifida fusca*, donde tiene aplicación potencial en el biocultivo del algodón para la industria textil (Zhang et al., 2010); el desarrollo de vacunas (Guillén et al., 2014; Moreno-mendieta et al., 2014); de biopelículas (Rivkin et al., 2015), entre otras aplicaciones.

Adicionalmente, se ha sugerido que los CBMs podrían ser blancos de estrategias de mutagénesis para mejorar algunas de sus propiedades. Por ejemplo, se ha sugerido que estas proteínas podrían funcionar como plataformas para la generación de nuevas proteínas sucesoras que tengan mejor especificidad y/o afinidad por algunos sustratos (Gunnarsson et al., 2004), ya que en su mayoría, tienen una estructura tridimensional altamente conservada en diferentes miembros de diferentes familias, con un sitio de unión al sustrato expuesto en la superficie. Sin embargo, hasta el momento han sido pocos los trabajos en los que esto se ha demostrado experimentalmente (Armenta et al., 2017), en comparación con los miles de proteínas depositadas en la base de datos de CaZy ([www.cazy.org](http://www.cazy.org)). Por lo que es importante llevar a cabo estudios más profundos que permitan entender mejor los mecanismos de unión a los distintos sustratos de los CBMs, así como los

requerimientos estructurales de los mismos, que permitan el diseño de nuevas proteínas.

## **2.4. Mutagénesis como estrategia para el estudio de CBMs**

### **2.4.1. Mutagénesis sitio dirigida**

De acuerdo con lo mencionado en los apartados 2.2.3.1 y 2.2.3.2, la importancia de aminoácidos en el sitio de unión ha sido revelada, principalmente, por estudios de mutagénesis sitio dirigida. Esta estrategia permite modificar cualquier aminoácido de forma específica y conocer el impacto de dicha mutación en diversos aspectos de una proteína (Hemsley et al., 1989). En los CBMs, no sólo ha sido utilizada para conocer la importancia de aminoácidos en el sitio de unión, sino también para modificar la afinidad o la especificidad en estas proteínas (Chang et al., 1998; Linder et al., 1995; Zhang et al., 2013). Por ejemplo, la mutación Y5W en el CBM1 de la celulobiohidrolasa I de *T. reesei*, mejoró dos veces la afinidad por celulosa (Linder et al., 1995). De forma semejante, las mutaciones W68Y y W68L en el dominio CBM2 de la endoglucanasa A de *C. fimi*, cambiaron por completo la especificidad de este dominio que pasó de unir celulosa a unir fibras de polietileno tereftalato (Zhang et al., 2013).

### **2.4.2. Mutagénesis aleatoria y procesos de selección**

Aunque las bases moleculares para la predicción y el diseño de nuevas proteínas aún son limitadas, se han desarrollado diversas estrategias de mutagénesis que permiten encontrar nuevas proteínas con cambios en más de un aminoácido de

interés. Esta estrategia imita a la naturaleza al generar mutaciones aleatorias y al azar seguidas de un proceso de selección.

El despliegue en fagos se ha utilizado en la búsqueda de nuevas proteínas a partir de CBMs de la familia 1 (Lehtiö et al., 2000; Smith et al., 1998) y 4 (Gunnarsson et al., 2007, 2006a, 2004). Respecto al dominio CBM1 de la celobiohidrolasa I de *T. reesei*, mediante la mutagénesis semi-aleatoria con oligonucleótidos degenerados, fueron mutados siete residuos. La biblioteca generada fue desplegada en el fago M13 y seleccionada por su afinidad a celulosa, así como a tres diferentes proteínas ( $\beta$ -glucoronidasa, fosfatasa alcalina bovina y  $\alpha$ -amilasa). De la selección se reportaron dos variantes del dominio CBM1 con capacidad de interactuar con  $\beta$ -glucoronidasa o fosfatasa alcalina bovina, perdiendo su capacidad de unir celulosa (Smith et al., 1998), dos variantes más que interactúan con la  $\alpha$ -amilasa porcina (Lehtiö et al., 2000) y dos más presentaron la capacidad de unir  $\text{Ni}^{2+}$  (Wernéus et al., 2001).

Por otro lado, el dominio CBM4 de una xilanasa de *Rhodotermus marinus* fue mutado semi-aleatoriamente al cambiar doce aminoácidos, mediante oligonucleótidos degenerados. La biblioteca generada también fue desplegada en el fago M13 y fue seleccionada con diferentes carbohidratos como xilano, xiloglucano, y la glucoproteína humana IgG4 (Gunnarsson et al., 2004). La estrategia permitió encontrar proteínas con mejor especificidad hacia xilano (Gunnarsson et al., 2007), mejor afinidad por xiloglucano (Gunnarsson et al., 2006b) y específicas a IgG4 (Gunnarsson et al., 2006a). Variantes como la X-2 y XG-34 obtenidas a partir del dominio CBM4, han dado a conocer algunas de las



características moleculares y estructurales que determinan la especificidad y afinidad por xilano y xiloglucano, respectivamente (Gullfot et al., 2009; Schantz et al., 2009).

Por lo tanto, la mutagénesis aleatoria en combinación con tecnologías, como el despliegue en fagos, son herramientas poderosas que pueden contribuir a entender los determinantes estructurales durante la interacción CBM-carbohidratos, así como ser una estrategia para la búsqueda de nuevas proteínas con características estructurales interesantes que pueden ser utilizadas en diferentes aplicaciones biotecnológicas.

### 3. Antecedentes

Dentro de la clasificación de CBMs, un subgrupo se conoce como dominios de unión al almidón (SBDs, por sus siglas en inglés), ya que tiene la capacidad de unir almidón y otros  $\alpha$ -glucanos. Las familias que pertenecen a este subgrupo son: CBM20, CBM21, CBM25, CBM26, CBM34, CBM41, CBM45, CBM48, CBM53, CBM58, CBM68, CBM69, CBM74, CBM82 y CBM83 (CaZy [www.cazy.org](http://www.cazy.org), revisado octubre, 2018).

Curiosamente, algunos SBDs forman tándems de hasta cinco dominios, tal es el caso de algunos lactobacilos (*L. plantarum*, *L. manihotivorans* y *L. amylovorus*) que presentan cuatro o cinco dominios de la familia 26 (CBM26) (Giraud and Cuny, 1997; Morlon-Guyot et al., 2001), pero regularmente se encuentran tándems con dos dominios, como en la amilasa de *B. halodurans* C125 con una serie CBM25-CBM26 (Boraston et al., 2006). El dominio de unión al almidón (SBD) de la  $\alpha$ -amilasa de *L. amylovorus* está compuesto por cinco módulos idénticos en tándem, pertenecientes a la familia 26 de módulos de unión a carbohidratos (CBMs) (Giraud and Cuny, 1997; Rodríguez-Sanoja et al., 2005).

Particularmente, de la familia CBM26 se conocen la estructura de tres dominios: el dominio CBM26 de la amilasa de *B. halodurans* C125 (PDB ID 2C3G); el dominio CBMb (CBM26-1) y CBMc (CBM26-2) de la  $\alpha$ -amilasa de *Eubacterium rectale* (cristalizado en tándem, PDB ID 6B15), los cuales adoptan el plegamiento  $\beta$ -sandwich característicos en los CBMs. De acuerdo con la información de las estructuras obtenidas con maltosa y maltoheptosa, respectivamente, hay tres aminoácidos aromáticos importantes para la unión al almidón: W36, Y23 y Y25 en

el CBM26 de *B. halodurans* C125; W208, Y196 y Y198 en el CBM26-1 y W314, Y297 y W299 en el CBM26-2 de *E. rectale* (Boraston et al., 2006, 2004; Cockburn et al., 2018). La información estructural de los dominios cristalizados muestra que, además de los aminoácidos aromáticos mencionados, hay diferentes aminoácidos polares que establecen enlaces de puentes de hidrógeno, pero que no están conservados en todos los dominios.

Por otro lado, un dominio CBM26 de la  $\alpha$ -amilasa de *L. amylovorus* ha sido estudiado mediante mutagénesis sitio dirigida, para conocer el papel de distintos aminoácidos aromáticos en la unión al almidón (Rodríguez-Sanoja et al., 2009). Los autores mencionan que sólo tres aminoácidos aromáticos son importantes para la unión de almidón: W32, Y18 y Y20, siendo el W32 indispensable para la función, ya que su mutación provoca la pérdida total del reconocimiento hacia almidón. Estos datos concuerdan con la importancia de los aminoácidos aromáticos como mediadores de la interacción proteína-carbohidrato (Boraston et al., 2004; Hashimoto, 2006).

Sin embargo, en el dominio CBM26 de la  $\alpha$ -amilasa de *L. amylovorus* no se conocen todos los aminoácidos que interaccionan con almidón u otros maltooligosacáridos, ya que se desconoce la estructura. Aunque, información estructural del dominio en complejo con el sustrato es deseable, la naturaleza autoproteolítica de este dominio ha dificultado el obtener cristales para conocer su estructura. Por lo que, estrategias de mutagénesis son una alternativa para elucidar la participación de distintos aminoácidos conservados en el sitio de unión.

#### **4. Justificación**

La importancia de los CBMs en la unión de carbohidratos demanda más estudios que aporten información acerca de la relevancia de los aminoácidos que participan en la interacción con el sustrato, ya que poco se conoce de las condiciones estructurales que determinan su especificidad y afinidad.

Actualmente, se desconoce la estructura terciaria del SBD de la  $\alpha$ -amilasa de *L. amylovorus*. Por lo que son necesarias otras estrategias que permitan conocer el papel de los distintos aminoácidos conservados durante la unión con almidón.

Entender este proceso es de vital importancia para el diseño de nuevos CBMs y el desarrollo o mejoramiento de aplicaciones biotecnológicas.

#### **5. Hipótesis**

La mutación puntual o simultánea de los aminoácidos no aromáticos del sitio de unión de un dominio CBM26 de la  $\alpha$ -amilasa de *L. amylovorus* evidenciará su importancia estructural y funcional en el proceso de unión.

## **6. Objetivos**

### **6.1. Objetivo general**

Conocer la importancia funcional de distintos aminoácidos que se encuentran en el sitio de unión de un CBM26 (LaCBM26) de la  $\alpha$ -amilasa de *L. amylovorus*.

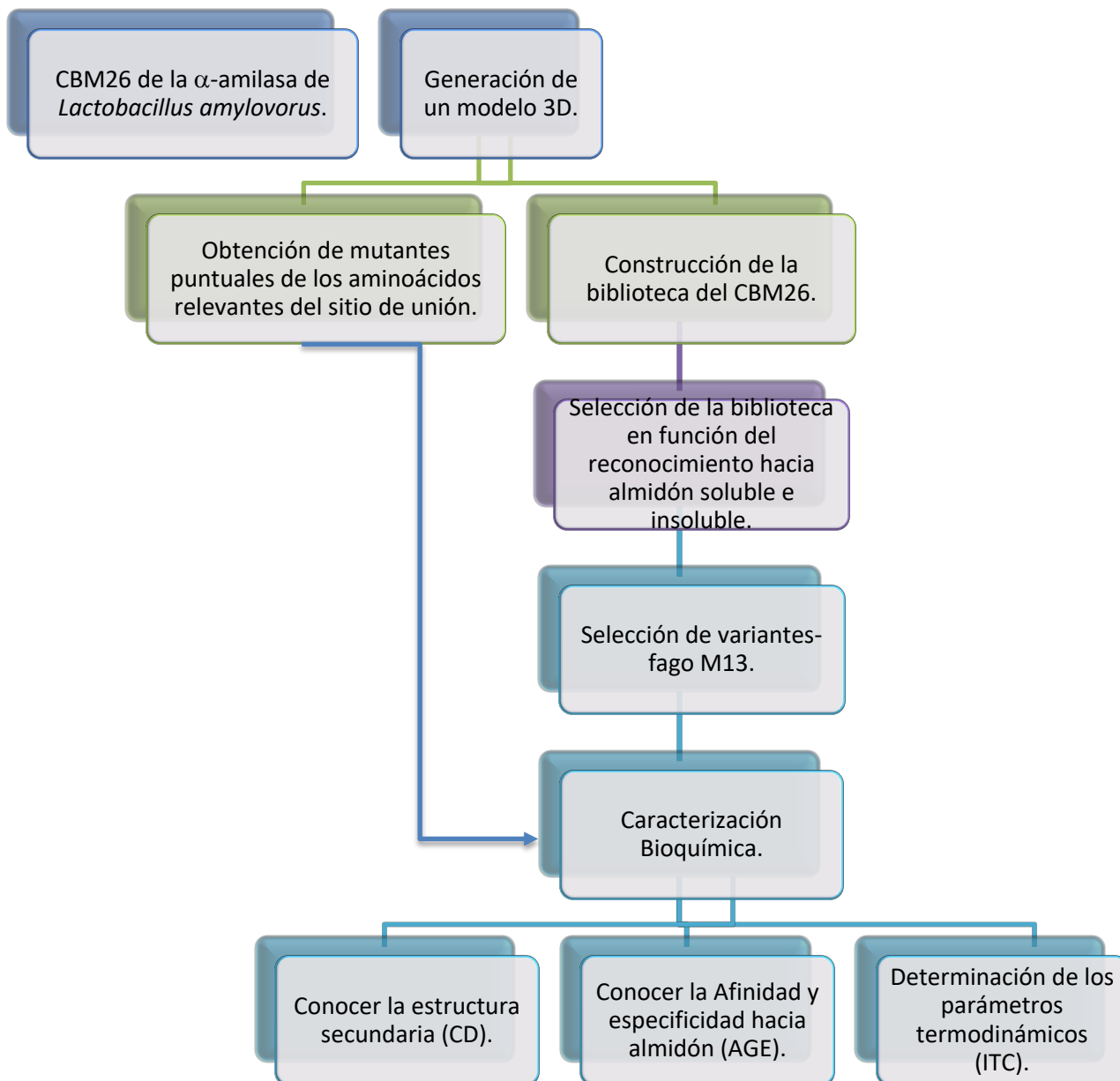
### **6.2. Objetivos específicos**

1. Identificar los aminoácidos del sitio de unión en el LaCBM26 que podrían estar involucrados en la unión de almidón.
2. Caracterizar el efecto de mutaciones puntuales de los aminoácidos seleccionados en el LaCBM26 sobre la capacidad de unión con almidón.
3. Establecer el efecto de las mutaciones combinatorias de los aminoácidos seleccionados en el LaCBM26 sobre la capacidad de unión con almidón.

## 7. Metodología.

### 7.1. Estrategia experimental

El siguiente esquema indica los diferentes pasos y metodologías llevadas a cabo para la realización de este proyecto.



## **7.2. Predicción de un modelo estructural del dominio CBM26 de la $\alpha$ -amilasa de *L. amylovorus***

La predicción del modelo tridimensional del dominio CBM26 de la  $\alpha$ -amilasa de *L. amylovorus* se realizó con el servidor I-TASSER (Roy et al., 2010; Yang et al., 2015; Zhang, 2008). El modelo obtenido tiene como molde el dominio CBM26 de la amilasa de *B. halodurans* C125 (PDB: 2c3g), que es el dominio con mayor identidad (28%) de la familia CBM26 del que se conoce su estructura, ya que con los dominios CBM26 de *E. rectale* el porcentaje de identidad es todavía menor (con el dominio CBM26-1 tiene 26% de identidad y con el CBM26-2 el 15%).

El modelo obtenido fue sometido a minimizaciones de energía de gradiente conjugado utilizando el paquete de Amber 16 con Amber ff99SB-ILDN campo de fuerza (protein force field; Case et al., 2016). La validación estructural del modelo tridimensional fue realizada mediante diferentes herramientas como los servidores: PROCHECK (Laskowski et al., 1993) y WHAT\_CHECK (Hoofst et al., 1996). La calidad estereoquímica de la estructura proteica fue revisada residuo por residuo indicando que > 85% de los residuos estaban en regiones correctas y en regiones preferidas / permitidas en la gráfica Ramachandran, mostrando una alta calidad. Finalmente, el modelo fue depositado en la base de datos de modelos de proteínas (PMDB) con el número PM0081297.

## **7.3. Mutagénesis sitio dirigida**

Se introdujeron mutaciones puntuales para cambiar los aminoácidos seleccionados a leucina. Las mutaciones se generaron utilizando el kit QuikChange Multi Site-

Directed Mutagenesis (Agilent Technologies), de acuerdo con el manual del proveedor, con dos oligonucleótidos antiparalelos (Tabla 1). Como molde se utilizó el gen *cbm26* en el vector pQ-1 (Guillén et al., 2007). Las PCR se realizaron utilizando 50 ng de templado, 10 pmol de oligonucleótidos (para insertar la mutación deseada), 0.5 mM de NTPs y 0.4  $\mu$ L de ADN polimerasa Phire Hot Start II (F122S Thermo Scientific) para un volumen final de 20  $\mu$ L, bajo las siguientes condiciones: 98 °C 30 s por 1 ciclo, 98°C 10 s, 55°C 10 s, 72°C 2 min por 30 ciclos, 72 °C 2 min 1 ciclo. Después de la PCR, los productos se trataron con 1  $\mu$ L de Dpn I a 37 °C por 1 h para digerir el ADN parental y luego se transformaron en células de *E. coli* XL10-Gold plaqueando cajas con medio LB-agar-Amp (100  $\mu$ g/mL). Todas las construcciones se validaron mediante secuenciación de ADN (Laboratorio de Secuenciación Genómica de la Biodiversidad y de la Salud, UNAM).

#### **7.4. Construcción de una biblioteca combinatoria de un dominio CBM26 del dominio LaCBM26 de la $\alpha$ -amilasa de *L. amylovorus***

La biblioteca se construyó al ensamblar siete oligonucleótidos, dos de los cuales tienen sitios con el codón degenerado “NNK”, N se refiere a las bases A, T, G y C; y K solo a G y C, de tal forma que las combinaciones de las bases nucleotídicas permiten codificar los 20 aminoácidos del código genético y el codón de paro TAG; así el tamaño teórico de la biblioteca fue de  $3.3 \times 10^7$  variantes y el producto del ensamblaje es el gen que codifica para el dominio CBM26 con cinco tripletes mutados al azar.



Tabla 1. Oligonucleótidos usados para mutagénesis sitio dirigida y en la construcción de la biblioteca del LaCBM26.

Name	Primers
<b>pQ-Q68L</b>	P1f: 5'-CCGATGGGACAAAGCTTACACCAGCAGCTAATG P2r: 5'-CATTAGCTGCTGGTGTAAGCTTTGTCCCATCGG
<b>pQ-E74L</b>	P1f: 5'-GCAAACACCAGCTGCTAATCTGGCTGGTTTTACC P2r: 5'-GGTAAAACCAGCCAGATTAGCAGCTGGTGTGGTGC
<b>pQ-F77L</b>	P1f: 5'-GAGGCTGGTTTAACCTTTACTGCAGATGCTACTTATG P2r: 5'-CATAAGTAGCATCTGCAGTAAAGGTTAAACCAGCCTC
<b>1/Rv1</b>	5'-TGCTACCCCAGCTTGACGGCTTTTCGAAGTAAACCTTTTAGTTTCTGTTGTTCCATGGCTGCTG
<b>2/Fw2a</b>	5'-CAAGCTGGGGTAGCAGAGTGTATGCCNNKGTANNKAATAAGAACACGAACAAGG <sup>a</sup>
<b>3/Rv2a</b>	5'-CTAAAGCGGTCATTTTTTTGCCAGGCCAAGCTGAAGTTATAGCCTTGTTTCGTGTTTC
<b>4/Fw3</b>	5'-AAATGACCGCTTTAGGTAACGACAAGTATGAATTGGATCTCGACACTGATGAAG
<b>5/Rv3a</b>	5'-GTCCCATCGGTGAATATAACAGCTAAATCAGAGTCATCTTCATCAGTGTCGA
<b>6/Fw4b</b>	5'-ATTCACCGATGGGACTAAGNNKACACCAGCAGCTAATNNKGCAGGTNNKACATTACAGCAGA <sup>a</sup>
<b>7/Rv4b</b>	5'-TTAAGTTAGGATCCTACGTATTTTACGACACCATTTTGGTCATAAGTGGCATCTGCTGTGAATGT

<sup>a</sup> Codón degenerado: K = T or G; N = A, C, G or T

Para llevar a cabo el ensamblaje se agregaron 20 pmol de cada oligonucleótido (el ensamblaje es para crear el templado para la subsecuente amplificación de DNA) (oligos 1-7; Tabla 1), 1.5 mM de MgSO<sub>4</sub> y 0.5 mM de dNTPs. Se obtuvieron fragmentos ensamblados después de 8 ciclos de PCR (95 °C por 1 min, 56 °C por 2 min, 68 °C por 2 min, con 1 U de la enzima Pfu polymerase platinum® de invitrogen). A partir de la reacción anterior, se utilizó una alícuota de 0.5, 1, 3 y 5 µL, para re-amplificar los fragmentos con 25 pmol de los oligonucleótidos 5'-CAGCAGCCATGGAACAACAGAAAC-3' y 5'-TTAAGTTAGGATCCTACGTATTTTAC-3' mediante 30 ciclos de PCR (98 °C por 30 s, 59 °C por 30 s, 72 °C por 30 s, con 1 µL de la enzima Phire polymerase Thermo scientific), e introducir los sitios que reconocen las enzimas de restricción *NcoI* (Thermo scientific) y *BamHI* (Thermo scientific), para clonar la biblioteca obtenida en el fagémido PG8SAET (Amablemente donado por el Dr. Karen Manoutcharian). El fagémido pG8SAET contiene el caset para expresar la proteína de interés

fusionada a la proteína p8 del fago M13 (Pedroza-Roldan et al., 2009). De la misma forma se clonó en el fagémido el gen del dominio silvestre (control positivo).

La generación y propagación de la biblioteca se realizó con células electrocompetentes de *E. coli* TG1 en medio LB-Amp (100 µg/mL). Para la transformación de la biblioteca se realizaron 16 electrotransformaciones dando un pulso de 1250 V con el electroporador 2510 marca Eppendorf. El tamaño y variabilidad de la biblioteca se verificó por secuenciación. Se realizó la extracción del DNA de la biblioteca con el kit Qiagen (No. de cat. 27104) de acuerdo con el protocolo del proveedor. El DNA de la biblioteca generada se almacenó a -20 °C. El despliegue de las variantes en el bacteriófago se realizó con el fago helper M13K07 (Invitrogen) y la cepa TG1 de *E. coli*, transformada con la biblioteca (ver Anexo 2; Pedroza-Roldan et al., 2009).

#### **7.5. Producción de partículas de fagos con variantes del LaCBM26 desplegadas en su superficie**

Se inocularon 5 mL de medio 2-YT/Amp (100 µg/mL) con una colonia o 10 µL de un stock de glicerol durante toda la noche (OVN) a 37 °C. Del cultivo anterior, se tomaron 50 µL y se inocularon de nuevo 5 mL de medio 2-YT/Amp (100 µg/mL) fresco (Dilución 1:100), se incubó a 37 °C hasta alcanzar una D.O. a 600nm de aproximadamente 0.7. Enseguida, se realizó la infección con fago helper (M13K07) a 10 MOI (multiplicidad de infección) diluido en 100 µL de medio 2-YT/Amp (100 µg/mL). El cultivo se incubó por 30 min a 37 °C sin agitación y 30 min más con agitación a 250 rpm. Del cultivo anterior se tomaron 2 mL y se inocularon 48 mL de

medio 2-YT/Amp (100 µg/mL) / kanamicina (50 µg/mL) fresco, mismo que se incubó a 37 °C durante 16-20 h con agitación de 250 rpm. El cultivo obtenido contenía las partículas de fagos, a partir del cual se realizó el rescate de estos.

El volumen de cada matraz se centrifugó a 8000 x g a 4 °C por 10 min. El sobrenadante se trasvasó en un tubo limpio y se adicionó 1/6 del volumen de PEG-8000 20% / NaCl 2.5 M para concentrar las partículas de fagos, el tubo se incubó a 4 °C durante toda la noche. Posteriormente, se centrifugó a 14 810 x g por 10 min a 4 °C, y el líquido se decantó. Las partículas fueron re-suspendidas en 1 mL de amortiguador TBS. Se realizó una segunda precipitación con 0.3 mL de PEG-8000 20% / NaCl 2.5 M, pero los tubos sólo se incubaron a 4 °C por 1 h. Se centrifugó de la misma forma, y se resuspendieron las partículas de fago en 0.6 mL de TBS totales. Las partículas de fagos se almacenaron a 4 °C hasta su uso (Cerillo, 2012).

#### **7.6. Inmovilización de almidón en placas de poliestireno**

Se agregaron 100 µL de una solución de almidón al 10% (Prolabo Fontenay-sous-Bois, France) en cada pozo de placas de poliestireno (Placas Costar de 96 pozos, No. Cat. 3370). Las placas con almidón se incubaron a temperatura ambiente (TA) por 22 h. Se lavó la placa con amortiguador PBS 1X y el exceso de líquido se eliminó colocando la placa sobre papel absorbente por 10 min. Para asegurar que el almidón se encontraba inmovilizado en la placa, se determinó el contenido de carbohidratos totales por el método de fenol-sulfúrico (Masuko et al., 2005). Se adicionaron 50 µL de amortiguador PBS 1X y se agitó la placa durante 30 s en el Multiskan FC (Thermo Scientific), la determinación de carbohidratos totales se

realizó al adicionar 150  $\mu\text{L}$  de  $\text{H}_2\text{SO}_4$  concentrado y 30  $\mu\text{L}$  de fenol al 5% (Masuko et al., 2005).

Para comprobar que el almidón inmovilizado se encontraba disponible para ser reconocido por los dominios de unión al almidón, diferentes concentraciones de proteína (1  $\mu\text{g}$  /mL, 10  $\mu\text{g}$  /mL y 25  $\mu\text{g}$  /mL) del LaCBM26, W32L, y LaSBD fueron adicionadas a placas con almidón inmovilizado, pre-incubadas en PBS-BSA (PBS con 0.2% de BSA) durante 1 h a TA, como previamente se describió. Las placas se incubaron por 3 h a TA. Después las placas se lavaron con PBS-5T (PBS con tween-20 0.05%) para eliminar la proteína no unida. Se adicionaron 100  $\mu\text{L}$  de anticuerpo primario anti-His (GeneTex; dilución 1:1000) en PBS-BSA porque las proteínas tienen un tallo histidinas en el extremo C-terminal, se incubó 1 h a TA. Como anticuerpo secundario se utilizó anti-mouse IgG acoplado a peroxidasa alcalina (Invitrogen; dilución 1:5000) en PBS-BSA, las placas se incubaron 1.5 a TA. Se cuantificó usando el sustrato FAST OPD (SIGMA), de acuerdo con las instrucciones del fabricante. La absorbancia se determinó a 450 nm en el Multiskan FC (Thermo Scientific).

### **7.7. Selección de variantes del dominio LaCBM26 con almidón**

La selección de la biblioteca en el fago M13 se realizó con almidón inmovilizado en placas de poliestireno y con almidón insoluble en suspensión. En el primer caso, se agregaron  $10^{11}$  ufp (unidades formadoras de partículas) diluidas en solución de bloqueo PBS-5T-BSA (pre-incubación por al menos 40 min en PBS con tween-20 al 0.05% y BSA al 0.2%) a cada pozo con almidón inmovilizado y se incubó toda la noche a 4°C. Después se realizaron cuatro lavados con PBS-5T y una vez más con

PBS. Los fagos unidos fueron eluidos por cambio de pH adicionando 200  $\mu$ L de Trietilamina (TEA) e incubando por 30 min a TA, después se neutralizó con la adición de 200  $\mu$ L de amortiguador Tris-HCl 1 M (pH 7.4). Se guardaron 50  $\mu$ L de la solución anterior para la cuantificación y búsqueda de variantes, mientras que el resto se utilizó para re-amplificar los fagos y realizar la siguiente ronda de selección. Se realizaron tres rondas de selección.

Para la selección con almidón insoluble, se utilizaron  $10^{11}$  ufp en PBS-T-BSA (PBS-BSA y 0.1% de tween-20) y se adicionaron a 500  $\mu$ L de una suspensión de almidón al 5% en PBS-T-BSA. La mezcla se incubó toda la noche a 4°C en un agitador rotativo. Después, la muestra se centrifugó a 8330 x g por 10 min. La muestra se lavó cuatro veces con PBS-T y una vez más con PBS. Los fagos unidos fueron eluidos con 300  $\mu$ L de una solución de  $\beta$ -ciclodextrina (9 mg/mL), incubando por 40 min a 4°C o por cambio de pH adicionando 200  $\mu$ L de TEA e incubando por 30 min a TA y neutralizando con 200  $\mu$ L de amortiguador Tris-HCl 1M (pH 7.4; Modificado de Gunnarsson et al., 2004). Se guardaron 50  $\mu$ L para la cuantificación y búsqueda de variantes, mientras que el resto se utilizó para re-amplificar los fagos y realizar la siguiente ronda de selección. Se realizaron dos rondas de selección.

Las variantes M13-Biblioteca LaCBM26 seleccionadas fueron analizadas por ensayos de ELISA. El ensayo de ELISA consistió en dejar interactuar a las proteínas desplegadas en el fago M13, con almidón inmovilizado en placas (Corning MaxiSorp), la amplificación y detección de la señal se obtuvo por la interacción con el anticuerpo anti-M13/HRP (dilución 1:5000) en PBS-5T-BSA. Se cuantificó usando el sustrato FAST OPD (SIGMA), de acuerdo con las instrucciones del fabricante. La

absorbancia se determinó a 450 nm en el Multiskan FC (Thermo Scientific). Como controles se utilizaron el dominio silvestre LaCBM26 desplegado en el fago M13 (control positivo) y un péptido no relacionado B22 (seq: QAALQRLFETCNE; antígeno con mutaciones a partir de la proteína p24 GAG de VIH; control negativo).

### **7.8. Clonación, expresión y purificación de las proteínas generadas de forma soluble**

Se realizó la clonación de cada gen que codifica para las diferentes variantes en el vector pET22b(+) (Novagen). La amplificación se realizó con los oligos **5'-TCGTTTCATATGTCCAAAAGGTTTACTTCG-3'** y **5'-TTATCCTGCGGCCGCTACGACACCATTTTG-3'**, los cuales, introdujeron los sitios que reconocen las enzimas de restricción *NdeI* y *NotI* (Thermo Scientific). Las condiciones de la PCR fueron (98 °C 5 s, 55 o 60 °C 5s, 72 °C 5 s, 25 ciclos), se utilizó 1.25 U de la enzima Pfu polymerase (Thermo scientific), 1 mM de MgSO<sub>4</sub> y 0.5 mM de dNTPs. Se clonaron 14 variantes diferentes. Se verificó cada clonación mediante secuenciación.

Se transformaron células termocompetentes de la cepa de *E. coli* Rosetta2 (DE3). Las células se incubaron en medio LB /Amp (100 µg/ mL) hasta alcanzar una DO aprox. de 0.8. La expresión fue inducida por la adición de IPTG (isopropil-β-D-tiogalactopiranosido) a 0.4 mM por 2 h a 37 °C y agitación a 250 rpm. El cultivo se centrifugó (8000 x g, por 10 min) y el pellet se resuspendió en 1/100 del cultivo original en solución de resuspensión (NaCl 0.5 M, imidazol 25 mM y fosfatos 20 mM pH 7.4 con inhibidor de proteasas para purificación con histidinas (P8849, SIGMA).

Las células fueron sonicadas por 30 s a 60 Hz con descansos de 30 s durante 30 min con el sonicador (Vibra-Cell, Inc). Después se centrifugó a 8000 x g por 10 min, el extracto obtenido fue filtrado por 0.8 y 0.22  $\mu\text{m}$ , posteriormente se purificó por cromatografía de afinidad con la resina ProBond (46-0019, Invitrogen) de acuerdo con las especificaciones del fabricante. Algunas ocasiones la purificación también se realizó mediante una cromatografía de intercambio iónico (HiTrap™ DEAE FF 17-5055-01, GE Healthcare) utilizando fosfatos 20 mM pH 7.4 y EDTA 1 mM como solución de unión y fosfatos 20 mM pH 7.4, EDTA 1 mM y NaCl 1M como solución de elución. Las fracciones obtenidas siempre fueron revisadas por SDS-PAGE (Laemmli, 1970).

### **7.9. Electroforesis de afinidad (AGE)**

La afinidad a almidón soluble de cada mutante, así como del dominio LaCBM26 silvestre se determinó por AGE. Se prepararon geles de poliacrilamida al 12% sin y con 1% de almidón soluble (Prolabo, Fontenay-sous-Bois, France). La migración de las proteínas se realizó en frío a 20 mA (Bio-Rad Mini-Protean III system). Como control negativo se utilizó BSA (1 mg/mL). La tinción de los geles se realizó con azul de Coomassie R-250 (22) (Tomme et al., 2000).

### **7.10. Dicroísmo Circular**

Se obtuvieron los espectros en el UV lejano (190-260 nm) de las distintas proteínas. La muestra se dializó con un amortiguador de fosfatos 10 mM pH 7.4, se filtró por 0.22  $\mu\text{m}$  y se determinó su concentración por absorbancia a 280 nm (0.20 mg/mL). Los espectros se obtuvieron por triplicado en un espectropolarímetro J-710 (Jasco

Inc., Easton, MD, USA) a 20 °C, a una velocidad de 10 nm/min, con un tiempo constante de 10 s, resolución 1 nm y sensibilidad de 20 mdeg. Los datos crudos fueron convertidos a elipticidad molar media por residuo  $[\theta]_{mrw,\lambda}$  mediante la fórmula:  $[\theta]_{mrw,\lambda} = MRW \times \theta_{\lambda} / (10 \times l \times C)$ , donde MRW es la masa media del residuo ( $MRW = M / (N_r - 1) = 110$  Da, M es la masa molecular de en la cadena polipeptídica (Da) y  $N_r$  es el número de residuos; para la mayoría de los casos se asigna el valor de 110 Da),  $\theta_{\lambda}$  es el valor de elipticidad determinado a cada longitud de onda, l es la longitud de la celda (cm), C es la concentración de la proteína (g/mL; Kelly et al., 2005).

### 7.11. Calorimetría de Titulación Isotérmica (ITC)

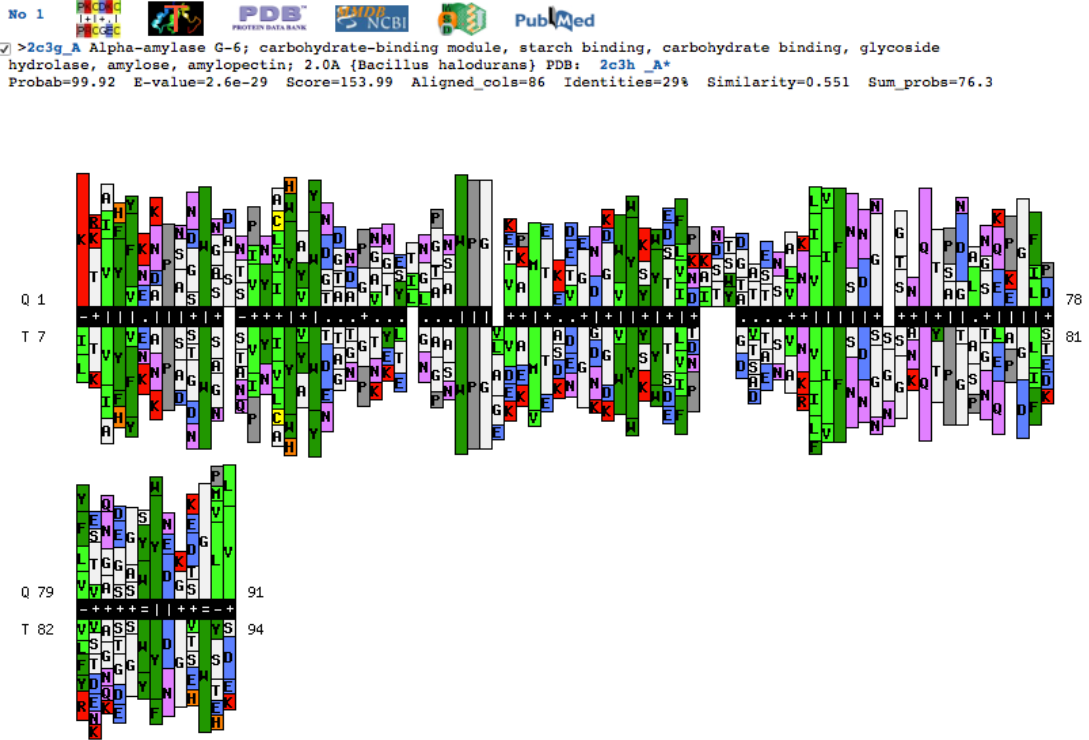
Se realizó el ITC en el ITC<sub>200</sub> (Malvern) a 25 °C. Las proteínas fueron dializadas exhaustivamente en amortiguador citrato-fosfato pH 5.5. La concentración de las proteínas se determinó por absorbancia a 280 nm. Las concentraciones que se utilizaron variaban entre 5 y 100  $\mu$ M (celda de reacción), y de ligando entre 1-10 mM de  $\beta$ -ciclodextrina (jeringa). El ligando se preparó en el mismo amortiguador en el que se dializaron las proteínas. Antes de la determinación las muestras fueron centrifugadas (16110 x g). La constante de unión ( $K_a$ ), y el cambio de entalpía ( $\Delta H$ ) se determinaron por ajuste no lineal usando un modelo de unión de un sitio de acuerdo con el software MicroCal Origin 7.0 (Wiseman et al., 1989). Los datos también fueron analizados con el software AFFINImeters (<https://www.affinimeter.com>). El cambio de energía libre de Gibbs ( $\Delta G^\circ$ ) y de entropía ( $\Delta S^\circ$ ) fue calculado de la fórmula  $\Delta G^\circ = -RT \ln K_a$  y  $\Delta G^\circ = \Delta H^\circ - T \Delta S^\circ$ , donde R es la constante de los gases ideales (1.987207 cal/mol K) y T la temperatura expresada en grados kelvin (K).



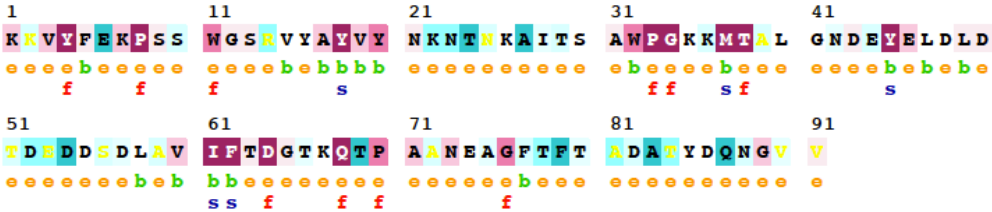
## 8. Resultados

### 8.1 Análisis *in silico* de un módulo CBM26 de la $\alpha$ -amilasa de *L. amylovorus*

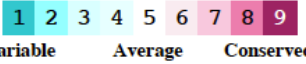
El SBD de la  $\alpha$ -amilasa de *L. amylovorus* tienen cinco dominios de 91 aminoácidos, organizados en tándem y con secuencia primaria idéntica. Con la secuencia de un solo módulo LaCBM26 se realizó una comparación con diferentes CBMs de secuencia homólogas generando un perfil estadístico de los residuos en cada posición (servidores HHpred-Söding et al., 2005 y ConSurf-Ashkenazy et al., 2016). Se observó que once posiciones dentro de la secuencia del dominio LaCBM26 están conservadas (W11, Y18, Y20, W32, P33, G34, Y45, Q68, G76, F77 y Y85) (Figura 9). Sin embargo, se sugiere que no todos estos residuos participan de igual manera en la unión. En un trabajo previo se revisó el papel de algunos de estos residuos, por ejemplo, la sustitución de W11, P33, Y45, y Y85 por leucina no modificó la capacidad del dominio de unirse al almidón, a pesar del carácter aromático e hidrofóbico de los residuos (Rodríguez-Sanoja et al., 2009). Por el contrario, la sustitución de Y18, Y20 y W32 si disminuyó la unión al almidón, y en el caso particular del W32 la eliminó (Rodríguez-Sanoja et al., 2009). En el caso de los aminoácidos G34, Q68, G76 y F77 no se tiene información respecto a alguna interacción directa con el sustrato durante la unión.



### ConSurf Results



The conservation scale:

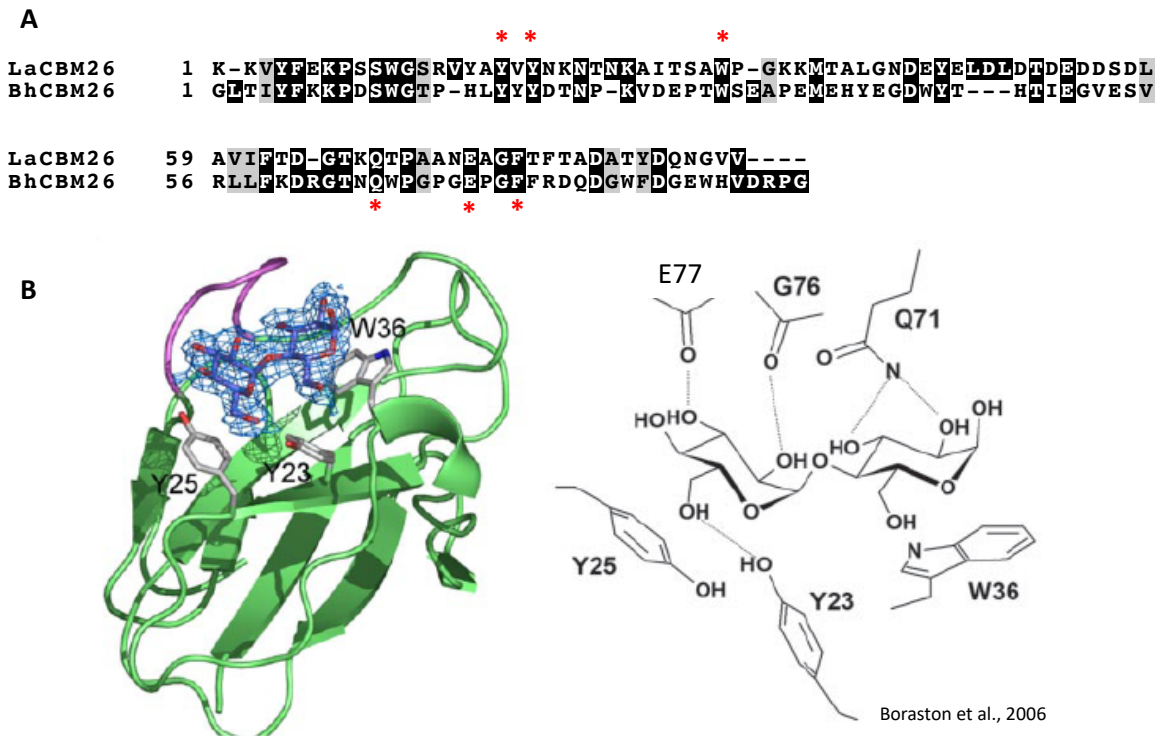


- o - An exposed residue according to the neural-network algorithm.
- b - A buried residue according to the neural-network algorithm.
- f - A predicted functional residue (highly conserved and exposed).
- s - A predicted structural residue (highly conserved and buried).
- x - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

**Figura 9.** Aminoácidos conservados en los dominios de unión al almidón de la familia CBM26. Resultados obtenidos con los servidores HHpred (arriba) y ConSurf (abajo) ([http://toolkit.tuebingen.mpg.de/hhpred/histograms\\_makemodel](http://toolkit.tuebingen.mpg.de/hhpred/histograms_makemodel)).

Se revisó la secuencia del LaCBM26 con la herramienta Basic Local Alignment Search Tool (BLAST), acotando la búsqueda a la base de datos del Protein Data Bank (PDB), la secuencia con menor valor de expectancia E (0.001) y mayor cobertura (93%) corresponde al dominio CBM26 (BhCBM26) de la amilasa de *B. halodurans*, con el cual el dominio LaCBM26 tiene 28% de identidad.

En el dominio BhCBM26 (PDB ID 2C3H) seis aminoácidos son importantes en la unión de maltosa, el W36 y Y25 participan mediante interacciones hidrofóbicas y CH- $\pi$ , mientras que la Y23, E77, G76 y Q71 lo hace a través de enlaces de puentes de hidrógeno (Figura 10; Boraston et al., 2006). A pesar del bajo porcentaje de identidad con el dominio LaCBM26, el alineamiento de la Figura 10 muestra que los aminoácidos Y18, Y20, W32, Q68, E74 y F77 en el LaCBM26 están conservados entre ambos dominios y que de acuerdo al dominio BhCBM26 podrían participar entonces en la unión del sustrato, mientras que la G34 y G76 no están conservadas entre los dominios (Boraston et al., 2006). Por lo que es necesario conocer más del papel de estos aminoácidos en el dominio LaCBM26.

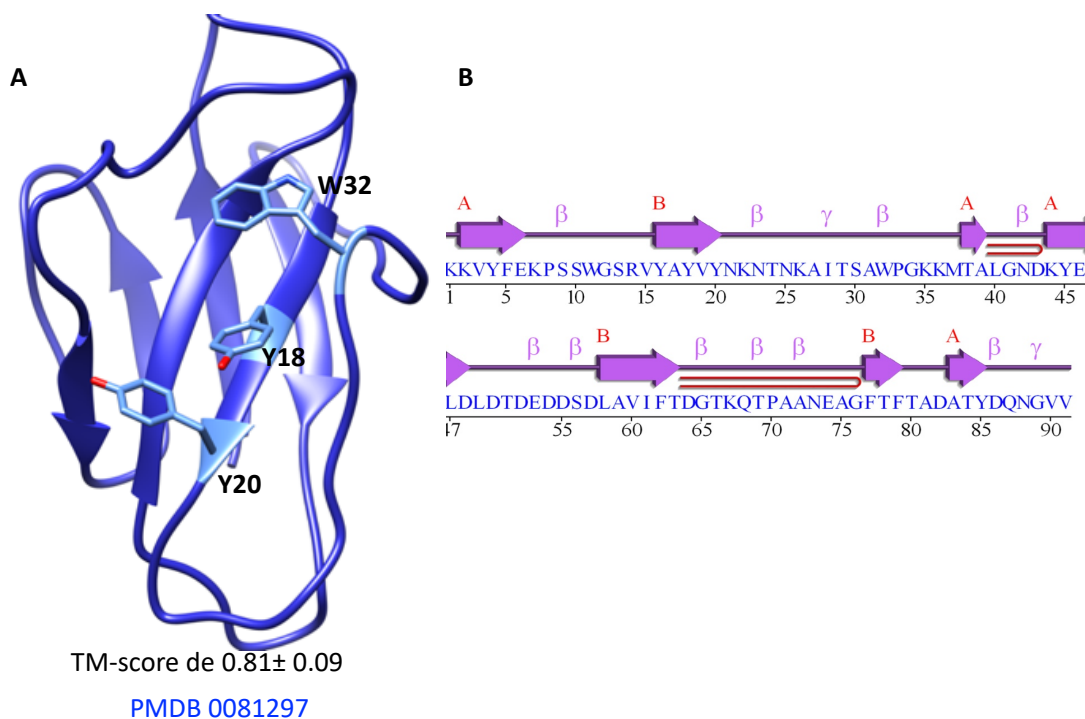


**Figura 10.** Aminoácidos conservados entre dos dominios de unión a almidón. A) Secuencia de los dominios LaCBM26 y *BhCBM26*. Los aminoácidos que participan en la unión de maltosa y que están conservados con el dominio LaCBM26 se indican con un asterisco rojo. B) Estructura del dominio *BhCBM26* con maltosa y representación esquemática de las interacciones que se establecen durante la unión del sustrato (Boraston et al., 2006).

Debido a que se desconoce la estructura tridimensional del dominio LaCBM26, se generó un modelo tridimensional por homología con el programa I-TASSER (Figura 11) (Roy et al., 2010; Yang et al., 2015; Zhang, 2008). El valor de TM-score fue  $0.81 \pm 0.09$ , lo cual sugiere que el modelo obtenido es aceptable, ya que valores mayores a 0.5 son significativos e indican una buena predicción (Zhang, 2008).

Se realizó la minimización de energía y validación del modelo (ver sección de Metodología). El modelo tridimensional obtenido (Figura 11) sugiere que el módulo

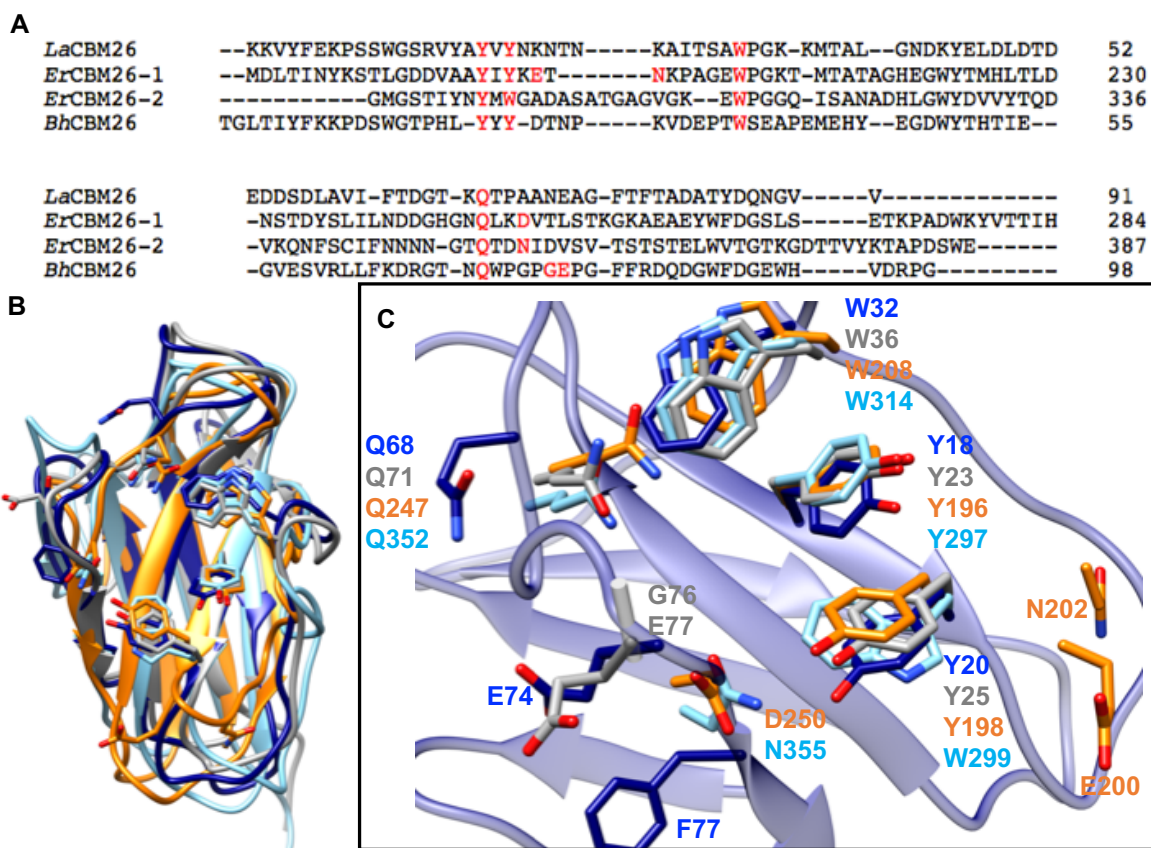
LaCBM26 presenta un plegamiento tipo  $\beta$ -sandwich con dos hojas  $\beta$ , cada una con tres (1:  $\beta$ 3,  $\beta$ 2,  $\beta$ 6) o cuatro (2:  $\beta$ 7,  $\beta$ 1,  $\beta$ 5,  $\beta$ 4) hebras beta antiparalelas, siendo éste el principal plegamiento reportado en la mayoría de los CBMs (Boraston 2004; Hashimoto 2006). En el modelo se puede ver que el sitio de unión está localizado en el lado cóncavo del  $\beta$ -sandwich, que es donde están ubicados los aminoácidos W32, Y18 y Y20 (Figura 11).



**Figura 11.** Modelo del dominio LaCBM26. A) Localización de los aminoácidos aromáticos importantes en la unión. B) Elementos de estructura secundaria del dominio.

Recientemente se reportó la estructura cristalográfica de dos dominios más de la familia 26, el dominio CBMb (*ErCBM26-1*) y CBMc (*ErCBM26-2*) (PDB: 6B15) de la  $\alpha$ -amilasa de *E. rectale* (Cockburn 2018). Nuevamente, la identidad que comparten cada uno de estos dominios con respecto al dominio LaCBM26 es baja, entre 15-

26%. Se realizó un alineamiento de estructura terciaria entre el modelo 3D del LaCBM26 y las tres estructuras de la familia CBM26 que se han reportado hasta el momento (Figura 12).



**Figura 12.** Comparación estructural entre cuatro dominios de unión a carbohidratos de la familia 26. A) Alineamiento de la estructura primaria obtenido por Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>); en rojo se indican los residuos conservados que interaccionan con el ligando de acuerdo con los estudios de estructura o por mutagénesis sitio dirigida. B) Alineamiento de la estructura terciaria del dominio *BhCBM26* (PDB: 2c3g; gris), dominio *ErCBM26-1* (anaranjado) y *ErCBM26-2* (azul claro) (PDB: 6B15) y el modelo de dominio *LaCBM26* (PM0081297; azul). C) Acercamiento del sitio de unión de los cuatro dominios CBM26, que muestran los residuos de unión importantes.

Las estructuras son similares entre sí, con una raíz de desviación cuadrática media RMSD (por sus siglas en inglés) de 1.3 Å, 2.5 Å y 2.1 Å al comparar los dominios *BhCBM26*, *ErCBM26-1* y *ErCBM26-2* con respecto al *LaCBM26*, donde las mayores diferencias se observan en las asas.

En el alineamiento de la Figura 12, se muestra la cadena lateral de los aminoácidos que participan en la unión y aquellos obtenidos del análisis de estructura primaria del dominio *LaCBM26*. Se observó que, mientras la orientación es semejante para los residuos aromáticos que intervienen en la unión, es diferente entre los aminoácidos polares, sobre todo en el caso de la glutamina, la cual, está conservada en la mayoría de los dominios de unión al almidón de la familia 26 (Majzlová and Janeček, 2014).

Se midieron las distancias entre el  $C\alpha$  y un átomo específico del grupo funcional de cada residuo. La distancia entre el  $C\alpha$  de la Q68 (*LaCBM26*), con el  $C\alpha$  de la Q71 (*BhCBM26*), Q247 (*ErCBM26-1*) y Q352 (*ErCBM26-2*) fue de 1.96 Å, 2.78 Å y 2.04 Å, respectivamente. Comparando el  $C\beta$  de la cadena lateral de este mismo aminoácido, se obtuvieron diferencias de 4.43 Å, 7.36 Å y 4.18 Å, respectivamente. Como se puede ver este aminoácido se ubica en una región de asas (Figura 12).

Para el caso del W32 (*LaCBM26*), un aminoácido indispensable para la unión al almidón, se encontró que entre el  $C\alpha$  de este residuo comparado con el triptófano en cada dominio, W36 (*BhCBM26*), W208 (*ErCBM26-1*) y W314 (*ErCBM26-2*), existe una distancia de 1.51 Å, 2.87 Å y 1.22 Å, respectivamente, mientras que al medir la distancia entre el átomo  $N\epsilon 1$  entre cada residuo fue de 2.46 Å, 3.42 Å y 1.88 Å, respectivamente.



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

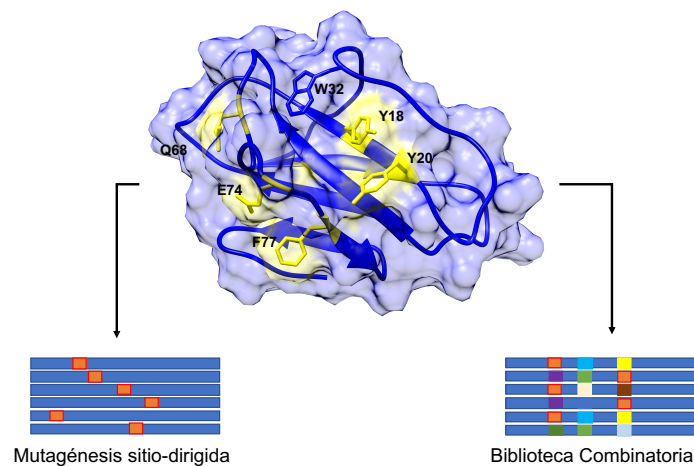
Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



Por el contrario, los residuos localizados en las hebras beta presentan menores desviaciones entre sí. Las dos tirosinas importantes para la unión en los cuatro dominios están localizadas dentro de una hebra beta. En consecuencia, se observaron menores diferencias entre estos aminoácidos, para cada uno de los dominios. Por ejemplo, la distancia entre el C $\alpha$  de la Y18 con respecto a la Y23, Y196, y Y297 es de 0.47 Å, 0.78 Å y 1.02 Å, respectivamente. De similar magnitud fueron las distancias entre el C $\alpha$  para las tirosinas correspondientes a la posición 20 del LaCBM26.

Por lo tanto, el modelo obtenido permitió localizar y revisar los aminoácidos conservados en el sitio de unión del LaCBM26 que podrían tener un papel importante para la unión de almidón. Por lo que, para conocer la importancia de estos aminoácidos se eligieron seis aminoácidos a mutagenizar (Y18, Y20, W32, Q68, E74 y F77; Figura 13) mediante dos estrategias, por mutagénesis sitio-dirigida y la construcción de una biblioteca combinatoria.



**Figura 13.** Aminoácidos del LaCBM26 con posible importancia en la unión de almidón (se muestran en sticks y de color amarillo), seleccionados para mutar de forma puntual o aleatoria.

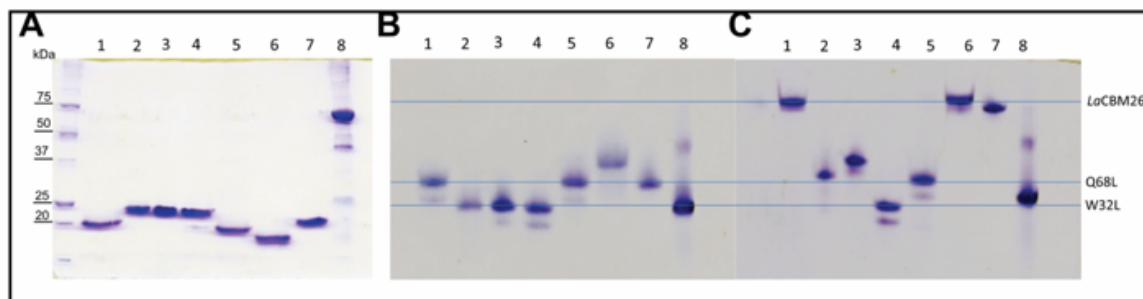
## **8.2. Caracterización bioquímica de mutantes generadas mediante mutagénesis sitio-dirigida**

### 8.2.1. Determinación de la capacidad de unión al almidón del dominio LaCBM26 y seis mutantes puntuales

Se caracterizaron seis mutantes puntuales (Y18L, Y20L, W32L, Q68L, E74L y F77L) del dominio LaCBM26. Primero, las proteínas fueron expresadas y purificadas por afinidad a Níquel (Figura 14A). Algunas de las proteínas mutadas migraron diferencialmente tanto en geles SDS-PAGE como PAGE, lo cual se atribuyó a que al cambiar un aminoácido polar que puede tener carga se puede modificar la migración de la proteína, aunque no se descarta algún tipo de migración aberrante, sobre todo en el caso de SDS-PAGE que en ocasiones se observa con este tipo de proteínas.

Se revisó el efecto de las mutaciones sobre la unión al almidón en geles de retardo, también se determinaron las constantes de afinidad hacia  $\beta$ -ciclodextrina por ITC (Wiseman et al., 1989). La Figura 14B y C muestra que dos proteínas mutadas, Q68L y W32L no fueron capaces de unir almidón soluble, ya que la migración en los geles de retardo no se modificó debido a la presencia del sustrato. Estos resultados, por un lado, corroboraron lo obtenido por Rodríguez-Sanoja et al. (2009) con la mutante W32L. Por el otro lado, el resultado de la mutante Q68L que no fue capaz de unir almidón, aún en presencia del W32, fue totalmente inesperado, y sugieren que la Q68 también es indispensable en la unión del dominio con almidón y no sólo el W32 como previamente se había establecido (Rodríguez-Sanoja et al., 2009). Por otro lado, la Y18 y Y20 retardan su migración en presencia del sustrato, pero ésta

es menor que el dominio silvestre (Figura 14C). Otras de las proteínas analizadas fueron las mutantes E74L y F77L las cuales no tienen afectada su capacidad de unir almidón soluble, ya que muestran un retardo similar al dominio LaCBM26.



**Figura 14.** Efecto en la unión a almidón de seis mutantes puntuales del dominio LaCBM26. (A) SDS-PAGE con las distintas mutantes puntuales del dominio LaCBM26 (B) PAGE en condiciones nativas de las mutantes del dominio LaCBM26 (C) AGE en presencia de almidón soluble. 1: LaCBM26, 2: Y18L, 3: Y20L, 4: W32L, 5: Q68L, 6: E74L, 7: F77L, 8: BSA (control negativo). Las líneas horizontales indican el retardo en la migración de las proteínas LaCBM26, Q68L y W32L.

Por otra parte, se determinaron los parámetros termodinámicos utilizando  $\beta$ -ciclodextrina como ligando; particularmente con el dominio LaCBM26 también se realizó la calorimetría con maltoheptosa (G7). El dominio LaCBM26 presenta una constante de unión del orden de  $10^4 \text{ M}^{-1}$ , tanto para  $\beta$ -ciclodextrina como para maltoheptosa (Tabla 2), la cual es semejante a otros dominios de unión a almidón de la misma familia o otras relacionadas (*BhCBM25* y *BhCBM26* (Boraston et al., 2006); CBM26-1/CBM26-2 (Cockburn et al., 2018); CBM41 (Cockburn et al., 2018; Lammerts Van Bueren et al., 2004); CBM45 (Glaring et al., 2011)).

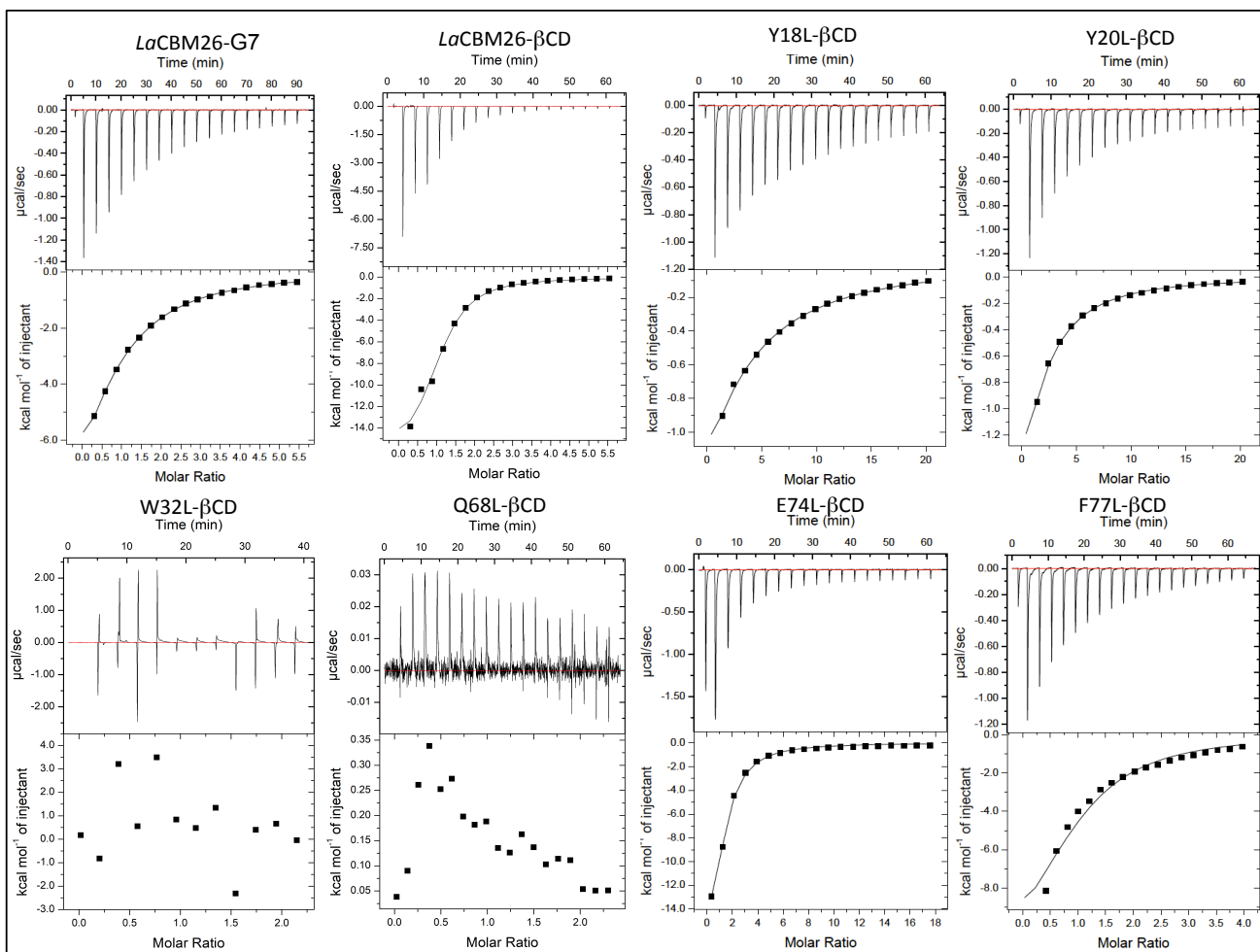
Del mismo orden de magnitud ( $10^4 \text{ M}^{-1}$ ) fueron las constantes obtenidas por las mutantes E74L y F77L, lo cual corrobora que la afinidad por el ligando de estas mutantes no está afectada como se observó en la electroforesis de afinidad. En el caso de las mutantes Y18L y Y20L, las constantes obtenidas fueron 50 y 20 veces menores a la constante de unión del dominio silvestre, respectivamente. También la incapacidad de las mutantes W32L y Q68L de unir carbohidratos fue corroborada por ITC (Tabla 2, Figura 15).

Por otro lado, los parámetros termodinámicos mostraron que la interacción del dominio LaCBM26, así como de las mutantes analizadas, hacia el sustrato son entálpicamente favorables ( $\Delta H$ ) y entrópicamente desfavorables ( $\Delta S$ ) en todos los casos, lo cual es característico durante la interacción de los CBMs con carbohidratos (Boraston et al., 2006; Glaring et al., 2011).

**Tabla 2.** Parámetros termodinámicos del dominio LaCBM26 y sus diferentes mutantes durante la unión a  $\beta$ -ciclodextrina determinados por ITC.

	$K_a$ $\times 10^{-4} \text{ M}^{-1}$	$\Delta G$ kcal/mol	$\Delta H$ kcal/mol	$T\Delta S$ kcal/mol
<b>Maltoheptosa</b>				
<b>CBM26</b>	2.41±0.4	-5.98	-11.4±1.0	-5.46
<b><math>\beta</math>-ciclodextrina</b>				
<b>CBM26</b>	5.69±0.9	-6.49	-16.8±0.9	-10.31
<b>Y18L</b>	0.11±0.01	-4.15	-7.8±0.01	-3.65
<b>Y20L</b>	0.25±0.04	-4.64	-8.01±0.03	-3.37
<b>W32L</b>			NB	
<b>E74L</b>	2.71±0.3	-6.04	-11.32±0.16	-5.27
<b>F77L</b>	3.66±0.3	-6.22	-14.4±1.1	-8.17
<b>Q68L</b>			NB	

NB: No Unión.



**Figura 15.** Curvas de calorimetría de titulación isotérmica (ITC) de LaCBM26 y diferentes mutantes puntuales con β-ciclodextrina.

### 8.2.2. Determinación de la estructura secundaria del dominio LaCBM26 y seis mutantes puntuales

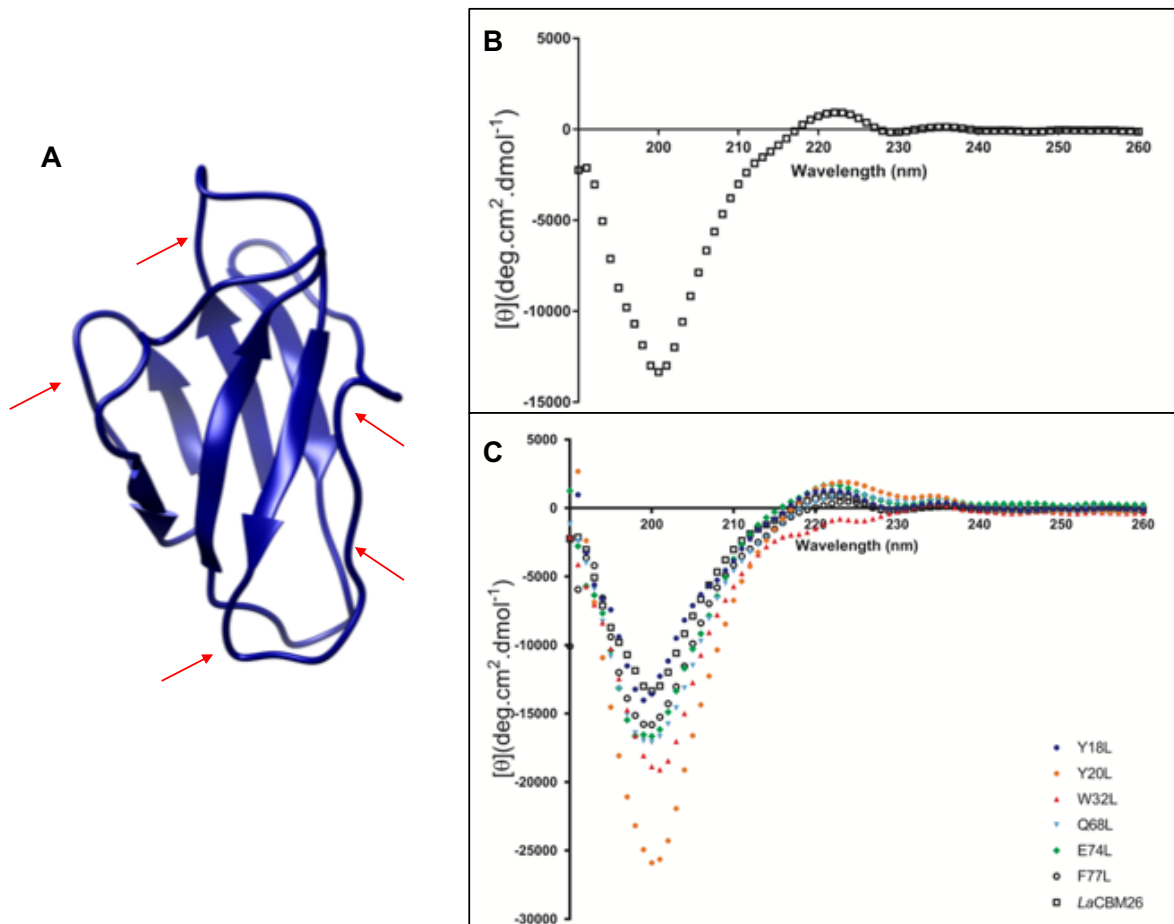
Se determinó el espectro de dicroísmo circular en el UV lejano (190 -260 nm) del dominio LaCBM26, así como de las seis mutantes puntuales obtenidas anteriormente para conocer si las mutaciones afectaron la estructura secundaria de las proteínas (Figura 16). Los espectros muestran un mínimo entre 198-202 nm de longitud de onda, así como dos máximos entre 218-227 nm y 231-239 nm de

longitud de onda. Con base en lo anterior, se sugiere que, si bien el dominio puede adoptar un plegamiento de  $\beta$ -sandwich, tiene un alto contenido de zonas sin plegamiento, lo cual se ve reflejado en los espectros atípicos obtenidos. Por lo que se realizó la estimación de elementos de estructura secundaria a partir de los espectros de la Figura 16 con el servidor BeStSel (<http://bestsel.elte.hu/index.php>) (Tabla 3). Los datos obtenidos concuerdan en que hay un alto contenido de giros, ya que de hebras  $\beta$  se obtuvo aproximadamente 38.7%, mientras que de giros y otras estructuras fue de 61.3%. No se obtuvieron cambios importantes en la estimación de elementos de estructura secundaria entre las distintas mutantes.

**Tabla 3.** Estimación de elementos de estructura secundaria de LaCBM26 y mutantes puntuales mediante el servidor BeStSel (<http://bestsel.elte.hu/index.php>).

Proteína	$\beta$ -antiparalela				Giros	Otros
	Suma	Left-twisted	relaxed	Right-twisted		
<b>LaCBM26</b>	<b>38.7</b>	2.6	18.1	18.0	<b>13.3</b>	<b>48.0</b>
<b>W32L</b>	<b>41.0</b>	2.1	18.0	21.0	<b>13.5</b>	<b>45.5</b>
<b>Q68L</b>	<b>42.5</b>	2.4	19.5	20.6	<b>13.7</b>	<b>43.8</b>
<b>Y18L</b>	<b>44.6</b>	2.7	20.8	21.1	<b>14.0</b>	<b>41.4</b>
<b>Y20L</b>	<b>42.8</b>	2.1	19.4	21.4	<b>13.8</b>	<b>43.4</b>
<b>E74L</b>	<b>42.6</b>	2.1	19.3	21.1	<b>13.9</b>	<b>43.5</b>
<b>F77L</b>	<b>41.0</b>	2.5	18.7	19.7	<b>13.7</b>	<b>45.3</b>

Suma: Contenido total de hebras  $\beta$  en láminas antiparalelas, dividido en left-twisted, relaxed o right-twisted.



**Figura 16.** Espectros de Dicroísmo Circular en el UV lejano del dominio LaCBM26 y mutantes puntuales. (A) Modelo tridimensional del dominio LaCBM26, las flechas rojas indican las zonas de asas. Espectros de Dicroísmo Circular en el UV lejano del dominio LaCBM26 (B) y diferentes mutantes puntuales (C). Los espectros fueron obtenidos en fosfato de sodio 10 mM pH 7.4, a 20 °C.

Sin embargo, todos los espectros obtenidos fueron semejantes entre sí, lo que indica que las mutaciones no provocaron cambios a nivel de estructura secundaria. Aunque el mayor cambio se observó en el espectro de la mutante W32L, el cual muestra una disminución del máximo entre 218-227 nm, que podría estar relacionado con la ausencia del anillo aromático. Hay reportes que indican que es

posible observar cierta contribución en el UV lejano entre 220 y 235 nm de longitud de onda (Woody, 1994), lo que no representaría un cambio a nivel de estructura secundaria.

### **8.3. Caracterización de mutantes del dominio LaCBM26 generadas mediante la construcción de una biblioteca combinatoria**

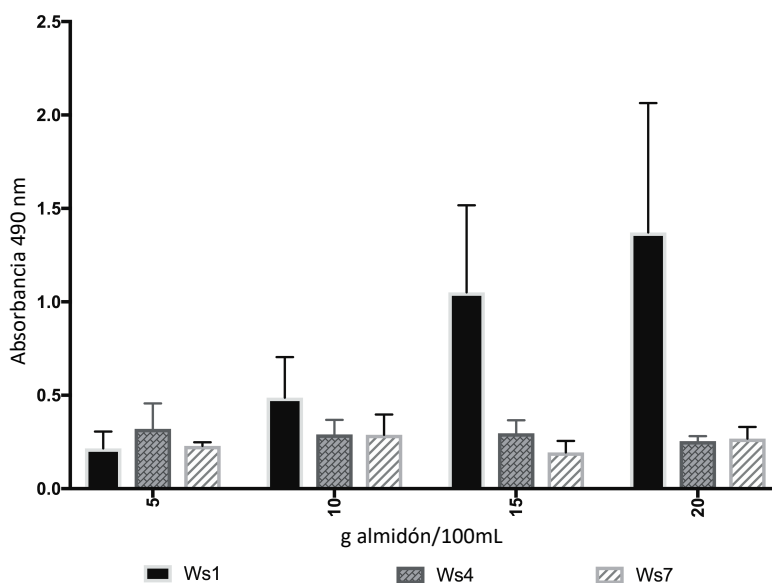
Se generó una biblioteca combinatoria del dominio LaCBM26, mutando los aminoácidos de cinco posiciones diferentes. Se mutagenizaron combinatoriamente la Y18, Y20, Q68, E74 y F77 (Figura 13) por cualquiera de los veinte aminoácidos existentes mediante el uso de oligonucleótidos degenerados. El W32 se mantuvo sin cambio para el diseño de la biblioteca para asegurar una plataforma hidrofóbica, mientras que la Q68 fue sometida a cambios, para evaluar si era posible que otro aminoácido, probablemente polar, supliera su papel durante la unión de almidón.

El tamaño de la biblioteca obtenida después de las transformaciones en *E. coli* TGI fue de  $1.1 \times 10^6$  variantes. De las colonias analizadas el 85% presentó amplificadas esperados en correcto marco de lectura, con presencia de 17 de los veinte aminoácidos, lo que sugirió una adecuada variabilidad.

Debido a que la biblioteca fue seleccionada bajo dos esquemas de presentación de sustrato, primeramente, se estandarizó y determinó la presencia de almidón soluble inmovilizado sobre la superficie de la placa (Masuko et al., 2005). El contenido de almidón inmovilizado se determinó después de 1, 4 y 7 lavados, a manera de simular el procedimiento característico de un ensayo de ELISA (Figura 17). Los datos



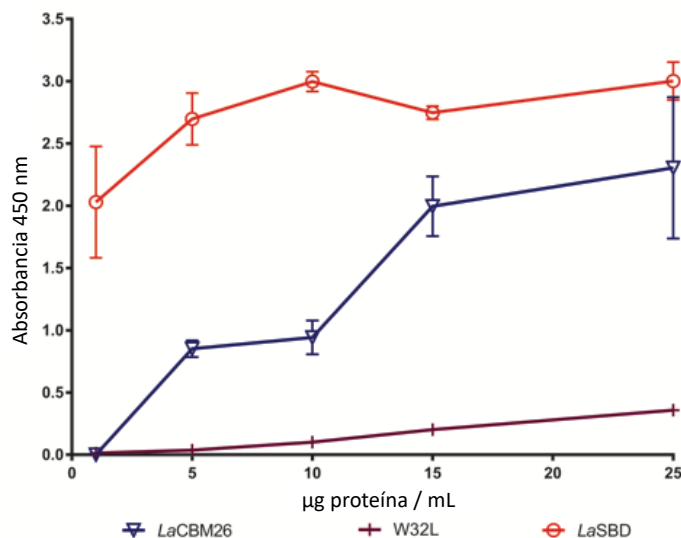
mostraron que el almidón se mantiene en la placa al menos después de siete lavados.



**Figura 17.** Cuantificación de almidón inmovilizado en placas de poliestireno. Determinación mediante el método de fenol-sulfúrico (Masuko et al., 2005). Ws1: Un lavado, Ws4: Cuatro lavados, Ws7: Siete lavados.

Después, se determinó si el almidón inmovilizado en la placa era reconocido por tres proteínas de unión a almidón (LaCBM26, LaSBD -dominio de unión al almidón en tándem de la  $\alpha$ -amilasa de *L. amylovorus* y W32L-control negativo). La Figura 18 muestra los valores de absorbancia conforme se incrementa la concentración de proteína. En el caso del LaCBM26, se observó un incremento de 3 veces al incrementar la concentración de proteína de 5  $\mu$ g/mL a 25  $\mu$ g/mL. Los datos obtenidos para el LaSBD son semejantes desde 1  $\mu$ g/mL a 25  $\mu$ g/mL de proteína, lo que indica que el sistema se saturó de inmediato, lo que resulta normal puesto que este dominio en tándem tiene 100 veces mayor afinidad hacia almidón que un

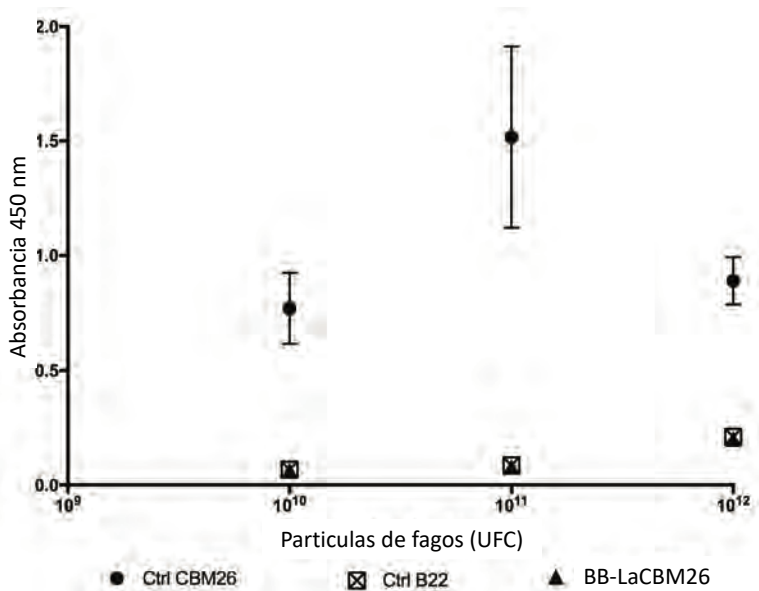
sólo módulo CBM26 (Guillén et al., 2007, García 2015). De acuerdo con lo esperado, la mutante W32L no se une al almidón inmovilizado en la placa. Por lo tanto, el sistema permitió la unión de proteínas en función de su afinidad al sustrato demostrando que el almidón inmovilizado en la placa está disponible para que sea reconocido.



**Figura 18.** Determinación de la capacidad de unión de almidón inmovilizado en placas de poliestireno de tres diferentes proteínas. LaCBM26: dominio de unión al almidón. W32L: mutante del dominio que no puede unir almidón. LaSBD: Dominio de unión al almidón con cinco módulos en tándem.

Posteriormente, se realizó la selección de la biblioteca combinatoria del dominio LaCBM26, desplegada en la superficie del bacteriófago M13. De la segunda y tercera ronda de selección con almidón soluble se revisaron cerca de 200 colonias y de la segunda ronda de selección con almidón insoluble aproximadamente 100 colonias, obteniéndose secuencias con mutaciones y deleciones espontáneas, lo que sugería falsos positivos.

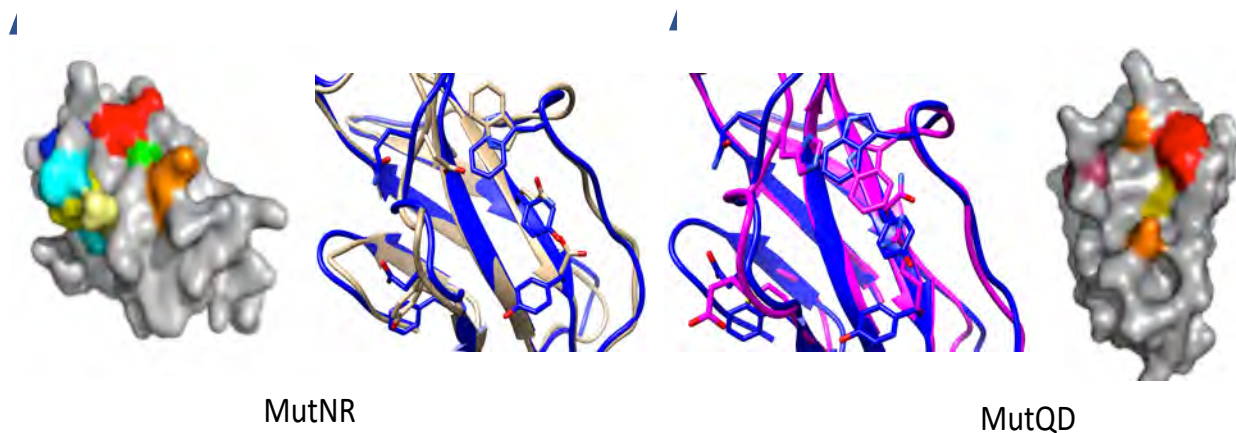
Se revisó si del total de la biblioteca existían variantes con la capacidad de unir almidón, mediante ELISA. Contrario a lo esperado, los valores de absorbancia obtenidos de la biblioteca, desplegada en el fago M13, fueron semejantes al control negativo (péptido derivado de HIV-1Gap). Lo cual sugiere que no hay suficientes partículas de fagos desplegando mutantes afines al almidón para poder ser detectadas por este método (Figura 19), o bien, es posible que las mutantes desplegadas en el fago M13 tienen un plegamiento distinto al del dominio LaCBM26, producto de las mutaciones introducidas durante el diseño de la biblioteca combinatoria.



**Figura 19.** Interacción del dominio CBM26, desplegado en el fago M13, con almidón soluble inmovilizado en placas de polipropileno. Control negativo, péptido no relacionado desplegado en el fago M13 (CtrlB22). Control positivo, dominio CBM26 desplegado en el fago M13 (CtrlCBM26). Biblioteca combinatoria del dominio CBM26 desplegada en el fago M13 (BB-LaCBM26).

A partir de la selección con almidón, se analizaron dos proteínas (MutQD y MutNR) que tenían una secuencia de aminoácidos relacionada con el dominio LaCBM26, es decir, la proteína MutQD sólo tenía mutaciones en los sitios asignados durante la construcción de la biblioteca, y la proteína MutNR además de tener mutaciones en los sitios asignados, tenía deletados seis aminoácidos en el extremo carboxilo terminal. En la Figura 20 se muestra la secuencia y representación tridimensional de las proteínas MutQD y MutNR donde se muestran las mutaciones en los sitios asignados.

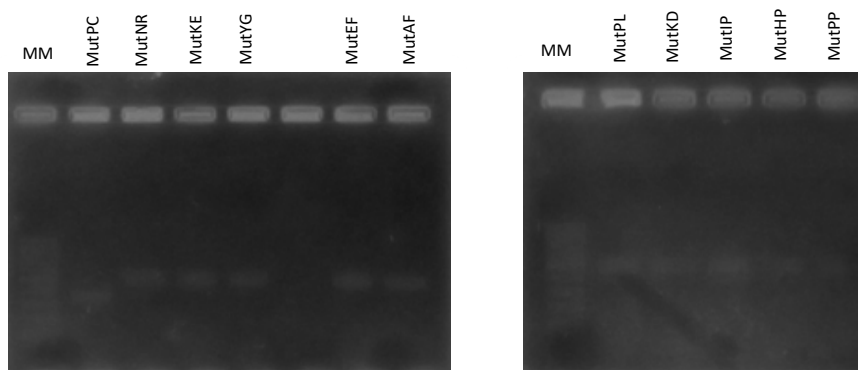
LaCBM26	1	KKVYFEKPSWGSRVYAYVYNKNTNKAITSAWPGKKMTALGNDKYELDLDTDEDDSDLAVIFTDGTKQTPAANEAGFTFTADATYDQNGVV	91
MutQD	1	.....Q.D.....D.....Y..L.....	91
MutNR	1	.....N.R.....H...M.QQMPLMTKMVS-----	85



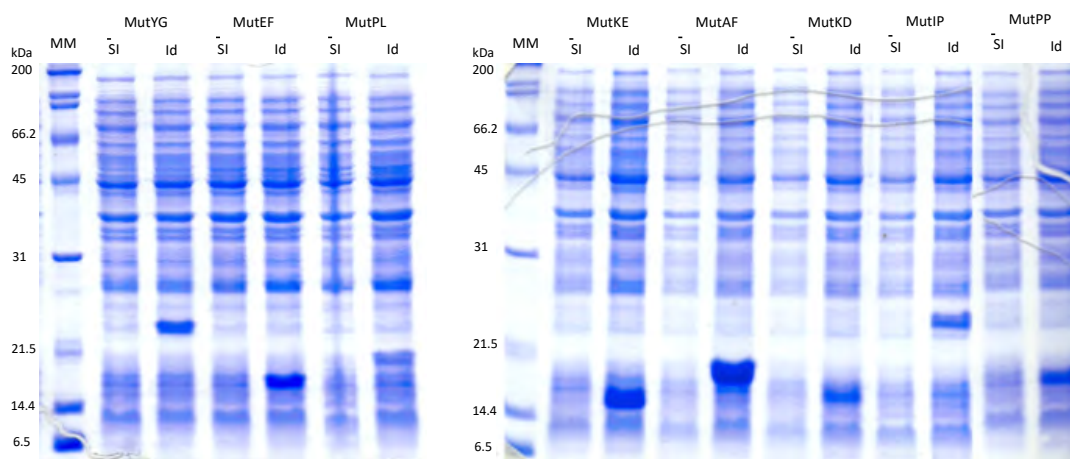
**Figura 20.** Mutantes aleatorias obtenidas durante la selección con almidón. Modelo de la estructura tridimensional obtenido por I-TASSER, marcando los sitios con mutaciones (izquierda), y el alineamiento de estos residuos con respecto al dominio LaCBM26 (representada en color azul) (derecha), para las mutantes aleatorias MutNR (representada en color gris) y MutQD (representada en color magenta), respectivamente.

Debido a lo anterior, diez colonias seleccionadas al azar (Figura 21), además de las dos variantes seleccionadas durante el proceso de selección con almidón, fueron expresadas en la cepa Rosetta 2 (DE3) y analizadas como proteínas solubles (Figura 22).

LaCBM26	1	KKVYFEKPSWGSRVYAYVYNKNTNKAITSAWPGKKMTALGNDEVELDLDTDEDDSDLAVIFTDGKQTPAANEAGFTFTADATYDQNGVV	91
MutKE	1	.....K.E.....F.....F.....	91
MutEF	1	.....E.F.....F.....F.....	91
MutYG	1	.....G.....P.....L.R.....	91
MutAF	1	.....A.F.....T.....R.G.....	91
MutPL	1	.....P.L.....Y.....F.....	91
MutHP	1	.....H.P.....G.....F.V.....	91
MutKD	1	.....K.D.....L.....F.I.....	91
MutPP	1	.....P.P.....T.....F.....	91
MutIP	1	.....I.P.....E.....Y.S.....	91
MutQL	1	.....Q.L.....D.....R.I.....	91



**Figura 21.** Mutantes aleatorias del dominio LaCBM26 seleccionadas al azar. Secuencia primaria de diez clonas seleccionadas al azar a partir de la construcción de la biblioteca del dominio LaCBM26, en cada una se muestran las diferencias de aminoácidos con respecto al dominio silvestre (arriba). Geles de agarosa al 1% que muestran el gen amplificado a partir de la construcción en el vector pET22b(+) de cada una de las mutantes aleatorias (abajo).

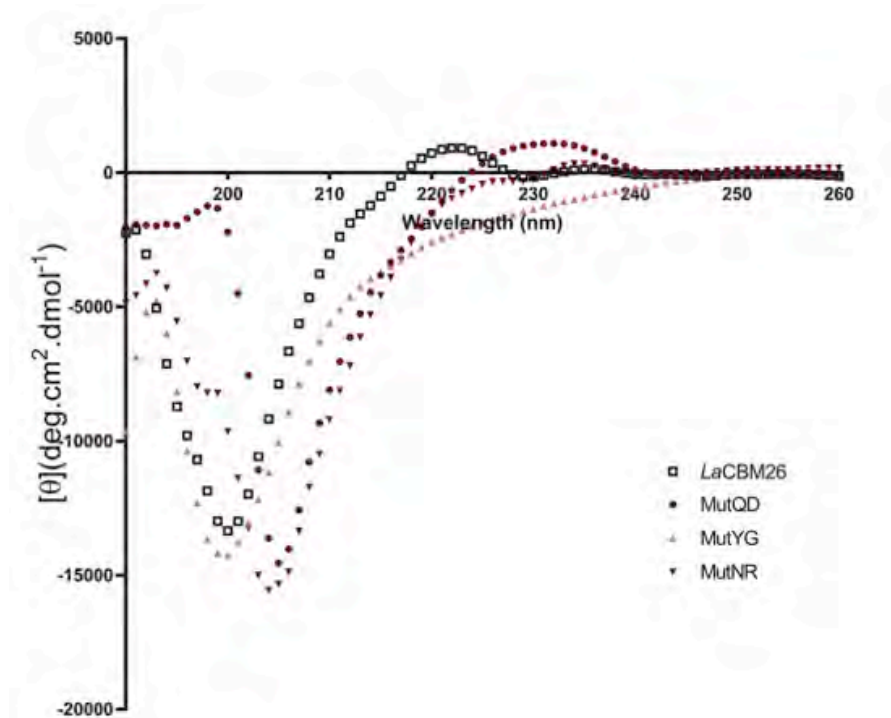


**Figura 22.** SDS-PAGE que muestra la expresión en Rosetta 2 (DE3) de variantes aleatorias del dominio CBM26. SI: Sin inducir. Id: Inducidas con 1mM de IPTG. Geles de poliacrilamida al 12%

La capacidad de unir almidón de cada una de las proteínas anteriores fue revisada mediante electroforesis de afinidad. Contrario a lo esperado, ninguna de las variantes revisadas conservó la capacidad de unir almidón, aun cuando todas contienen dentro su secuencia el W32 (Figura 23, Anexo 3). Por lo tanto, mantener el W32 no fue suficiente para que se lleve a cabo la unión al almidón. De la misma forma, ninguna de las mutantes fue capaz de unir  $\beta$ -ciclodextrina, determinado por ITC. En la figura 23 se muestra a manera de ejemplo el análisis de dos de las mutantes obtenidas, en las que se revisó por AGE no solo su capacidad de unirse a almidón sino también a amilosa, amilopectina y glucógeno.



nivel de estructura secundaria entre las variantes y el dominio LaCBM26, ya que se ven desplazados los máximos y mínimos entre las proteínas, observándose en diferentes longitudes de onda (Figura 24). Aunque estas mutantes tienen en su secuencia al W32, ninguno de los espectros obtenidos tiene el máximo observado entre 219-227 nm, lo que sugiere cambios conformacionales que afectan algunos elementos de estructura secundaria que podrían verse reflejados en la orientación del W32, y probablemente esto conduzca a la pérdida de la afinidad hacia almidón.



**Figura 24.** Espectros de Dicroísmo circular en el UV lejano de variantes aleatorias del dominio LaCBM26. Los espectros fueron obtenidos en fosfato de sodio 10 mM pH 7.4, a 20 °C.

Cabe señalar que no se descarta la opción de que las mutaciones encontradas en las variantes combinatorias conlleven a una afectación en la estabilidad dinámica y/o movilidad intrínseca de la proteína que pueda repercutir de forma importante en



la función de dichas proteínas. Además, es importante destacar que las variantes revisadas tienen diferentes aminoácidos en la posición 68 y en ningún caso es glutamina, el cual es indispensable para la función del dominio, por lo que son necesarios más experimentos para determinar a fondo su papel en la unión de almidón.

## 9. Discusión

Los CBMs unen diversos polisacáridos, entre ellos almidón, sin embargo, los determinantes estructurales de su especificidad y afinidad aún se desconocen, sobre todo en dominios que carecen de información estructural. Se han sugerido modelos generales para predecir la importancia de algunos aminoácidos en la unión de carbohidratos. Se ha propuesto que los CBMs que unen  $\alpha$ -glucanos tienen en el sitio de unión tres regiones importantes, A, B y C, en la región A siempre hay un triptófano que establece interacciones hidrofóbicas y CH- $\pi$ ; en la región B hay un aminoácido aromático que puede ser un triptófano o tirosina que también establece interacciones hidrofóbicas y CH- $\pi$  contribuyendo así a la plataforma hidrofóbica para la unión de carbohidratos, y en la región C un aminoácido polar (Abbott and van Bueren, 2014).

Se conocen únicamente tres estructuras de las 533 secuencias depositadas de la familia 26 de módulos de unión a carbohidratos (CaZy, [www.cazy.org](http://www.cazy.org), octubre 2018). Como se mencionó, el dominio de unión al almidón (*LaSBD*) de la  $\alpha$ -amilasa de *L. amylovorus* tiene una estructura particular, ya que está formado de cinco módulos de unión a carbohidratos de la familia 26 idénticos y en tándem (Giraud & Cunny 1997; Rodríguez Sanoja et al., 2000), se desconoce la estructura terciaria de este dominio y dada la naturaleza autoproteolítica de la proteína (Sánchez-Cuapio, en preparación), los estudios estructurales son sumamente complicados.

Por lo tanto, se eligieron distintos residuos con base en el análisis de estructura primaria (Figura 9 y 10) y localización en el modelo tridimensional obtenido por homología del LaCBM26 (Figura 11) para determinar su papel en el sitio de unión

del dominio LaCBM26 y revisar si podrían corresponder a las regiones A, B y C mencionadas anteriormente. El modelo tridimensional obtenido sugiere un plegamiento tipo  $\beta$ -sandwich, el cual es el tipo de plegamiento predominante entre los dominios de unión a carbohidratos (Boraston et al., 2006, 2004; Cockburn et al., 2018; Hashimoto, 2006). El alineamiento tridimensional de los dominios BhCBM26, ErCBM26-1 y ErCBM26-2, con respecto al modelo (Figura 12) indicó que tienen estructuras similares (RMSD 1.3 Å, 2.5 Å y 2.1 Å, respectivamente), donde las principales diferencias se observan en las asas. Es posible que el alto contenido de regiones sin plegamiento que tiene el dominio LaCBM26 justifique el tipo de espectro atípico de dicroísmo circular obtenido en el UV-lejano (Figura 16), en congruencia con esta observación, los datos obtenidos mediante la estimación de estructura secundaria por el servidor BeStSel indicaron pocos elementos de estructura secundaria y un número significativo de asas (Tabla 3). De forma similar, espectros atípicos de proteínas con un alto contenido de asas se ha reportado en otros dominios de unión al almidón, por ejemplo, en un dominio de unión al almidón (SBD1) de la  $\alpha$ -amilasa de *Kocuria varians* (Yamaguchi et al., 2012), así como en otras proteínas que unen carbohidratos, por ejemplo, la heveina (Rodríguez-Romero et al., 1989).

En cuanto a los estudios de unión, el dominio LaCBM26 tiene una afinidad de  $\sim 10^4$  M<sup>-1</sup> tanto por la  $\beta$ -ciclodextrina como por la maltoheptosa (Tabla 2), aunque la afinidad es dos veces mayor por  $\beta$ -ciclodextrina, probablemente por la semejanza estructural de la  $\beta$ -ciclodextrina con la amilosa, y debido a la penalización entrópica adicional asociada con la estabilización de la conformación de la maltoheptosa al

ocurrir la unión (Williamson and Belshaw, 1993). La afinidad por almidón que se reporta para el dominio LaCBM26 es muy semejante a la reportada en otros dominios de unión al almidón, por ejemplo, los dominios CBM26-1 y CBM26-2 de la  $\alpha$ -amilasa de *E. rectale* (Cockburn et al., 2018), los dominios CBM25 y CBM26 de la amilasa de *B. halodurans* (Boraston et al., 2006), el dominio CBM45 de *Arabidopsis thaliana* (Glaring et al., 2011), entre otros, además en todos los casos la estequiometría del sistema es 1:1. Con base en los parámetros termodinámicos, la interacción del LaCBM26 con carbohidratos es estable ( $\Delta G = -6.49$  kcal/mol), exotérmica y manejada entálpicamente, lo que es compensado por la pérdida de la entropía durante la unión del ligando.

Por otro lado, la relación estructura-función de los aminoácidos conservados en el sitio de unión, revisada por mutagénesis sitio-dirigida, demostró la importancia de aminoácidos polares (Q68) en el sitio de unión, a través de la pérdida de la unión al almidón aún y cuando está intacta la plataforma hidrofóbica (W32, Y18 y Y20). Posiblemente, la Q68 está formando puentes de hidrógeno con el sustrato de forma semejante al dominio BhCBM26, y el hecho que no esté presente elimina estas interacciones, lo cual no permite la unión, aunque no se descarta la posibilidad de cambios conformacionales alrededor del W32 que impidan su interacción y en consecuencia se pierda la función. La presencia del W32 es importante debido a que establece interacciones hidrofóbicas y CH- $\pi$ , que son dependientes de la distancia y orientación del aminoácido y de las moléculas de glucosa (Figura 5), por lo que cualquier cambio podría ser relevante para la interacción con el polisacárido.

Por el contrario, las mutaciones en Y18 y Y20, que teóricamente también forman un enlace de hidrógeno y una interacción hidrofóbica, respectivamente, todavía permiten la unión del sustrato (Figura 14-15, Tabla 2). En este caso, es posible que la pérdida de enlaces de hidrógeno conduzca a una redistribución de las interacciones que compense dichos enlaces, aunque los parámetros termodinámicos sugirieron que las mutaciones realizadas disminuyeron la estabilidad de la interacción (Boraston et al., 2001). Sin embargo, es difícil conocer con precisión el efecto de las mutaciones en el sitio de unión que estén implicados en cambios conformacionales locales sin tener información estructural tridimensional, por lo que son necesarios más estudios al respecto. Por otro lado, la mutagénesis de los aminoácidos E74 y F77 no modifica la afinidad por almidón (Figura 14-15, Tabla 2).

Los espectros de CD en el UV-lejano de los mutantes analizadas mostraron que la estructura secundaria del dominio LaCBM26 no está afectada por mutaciones puntuales, aunque las mínimas diferencias observadas sugieren cambios conformacionales locales (Figura 16). El mayor efecto se observó para el caso de la mutante W32L, en la cual ya no se observa el máximo observado entre 218 y 227 nm; posiblemente la mutación W32L provoca cambios conformacionales en la proteína, modificando así el espectro de CD, aunque esto no representa cambios en el contenido de los elementos de estructura secundaria, ya que los valores estimados con el servidor BeStSel son muy semejantes con respecto al dominio silvestre (Tabla 3). Al respecto, Woody, (1994) menciona que los valores de elipticidad molar entre 220-235nm puede explicarse por una contribución de la presencia de aromáticos en la estructura expuestos a la superficie, relacionados con

la interacción con carbohidratos, lo cual se ha observado de forma similar en otras proteínas como la proteína de unión a glucano A de *Streptococcus mutants* (Haas et al., 1998) y el dominio de unión a almidón de *Aspergillus niger* (Ota et al., 2016) y *R. oryzae* (Jiang et al., 2012).

Por lo tanto, los resultados obtenidos sugieren que en el dominio LaCBM26 al menos son indispensables dos aminoácidos, el W32 y la Q68, que podrían corresponder a la región A y C en el análisis del sitio de unión a  $\alpha$ -glucanos propuesto por Abbott et al. (2014), mientras que la Y20 podría corresponder a la región B, ya que con base al alineamiento de estructura terciaria (Figura 12), empalma con la Y25, Y198 y W299 que interaccionan con el ligando a través de interacciones CH- $\pi$ .

Por otra parte, en este trabajo también se construyó una biblioteca combinatoria del dominio LaCBM26, para modificar de manera aleatoria a cinco residuos del sitio de unión que permitiera obtener variantes que aportaran información referente a los determinantes estructurales en el sitio de unión de este dominio.

El uso de los CBMs como plataformas para modificar aminoácidos importantes en la unión ha sido poco explorado. Esta estrategia ha sido utilizada para mutagenizar de forma combinatoria al dominio CBM4-2 de la xilanasas 10A de la bacteria termófila *R. marinus*, el cual une xilano, xiloglucano, manano y celulosa, obteniendo variantes específicas a algunos de estos sustratos, por ejemplo la variante X-2 que une específicamente xilano o la variante XG-34 que une específicamente xiloglucano (Gunnarsson et al., 2004). A pesar de las aplicaciones potenciales que se derivan de esta tecnología, pocos han sido los dominios analizados. Recientemente,

también fue mutagenizado, de forma totalmente aleatoria, el dominio CBM11 de CelH (celulosoma asociado a proteínas multidominio:  $\beta$ -1,3-1,4 endoglucanasa Lic26A y la enzima bifuncional celulasa/xilanasa Cel5E), a partir del cual se encontraron variantes de dominio con una o hasta cuatro mutaciones que incrementaba su afinidad hasta 15 veces en comparación con el dominio silvestre (Furtado et al., 2018).

Sin embargo, los resultados obtenidos a partir de la construcción de la biblioteca combinatoria sugieren que el dominio LaCBM26 tiene poca tolerancia al cambio y a la inserción de mutaciones al azar en los sitios seleccionados, ya que si bien se logró construir una biblioteca de hasta 1.5 millones de variantes, fue difícil seleccionar variantes que mantuvieran o mejoraran la unión hacia almidón.

Al analizar los espectros obtenidos de las mutantes con modificaciones aleatorias se observó que en cada caso las mutaciones provocan cambios a nivel de estructura secundaria, así como la pérdida de la función, por lo que es posible que solo algunas arquitecturas de proteínas compensen el impacto de mutar residuos importantes relacionados con la función y soporten la innovación funcional.

Como se mencionó anteriormente, el dominio LaCBM26 tiene pocos elementos de estructura secundaria y una cantidad importante de giros y regiones poco estructuradas, esto puede provocar que no exista en el dominio estabilidad estructural que permita mutagenizar de manera aleatoria los aminoácidos seleccionados, como regularmente es requerido (Skerra, 2000).

En el caso de los dominios CBM4-2 y CBM11, utilizados como plataformas para la búsqueda de variantes con mejor afinidad a diferentes sustratos, tiene una

estructura con mayor cantidad de hebras  $\beta$  CBM4-2, 50% (PDB ID 1k42) y CBM11, 48% (PDB ID 1V0A); que el dominio LaCBM26 que solo llega al 38%; datos obtenidos con el servidor BestSel (<https://www.bestsel.elte.hu>; Tabla 3). Por lo que es posible que puedan tolerar mejor la inserción de cambios de aminoácidos sin perder su integridad estructural. Sin embargo, no hay que dejar de lado que todas las mutantes probadas tenían modificado el residuo Q68 y este solo hecho elimina la adsorción. En consecuencia, sería necesario restaurar este aminoácido para asegurar el rol de los otros aminoácidos y la estructura secundaria.

Respecto al despliegue de la biblioteca en el fago M13, no se han presentado limitaciones estructurales al desplegar péptidos de hasta 5 kDa fusionados a la proteína p8 del fago M13, aunque se ha encontrado que algunas inserciones de péptidos en esta proteína de despliegue pueden afectar el ensamblaje del bacteriófago, debido a mecanismos relacionados con la inserción de las proteínas de fusión en la membrana (Malik et al., 1998). Es posible que ciertas fusiones puedan ser tóxicas para *E. coli* y que su expresión limite la producción y ensamblaje del fago M13 (Sidhu et al., 2000). Además, seleccionar o enriquecer partículas hacia blancos no relacionados, por ejemplo, al plástico o BSA, conduce a falsos positivos (Fukunaga and Taki, 2012). En consecuencia, algunas de estas limitaciones podrían haber afectado el sistema de despliegue de la biblioteca del LaCBM26.

Es claro que las características estructurales que promueven la evolución de las proteínas solo se entienden parcialmente. Por lo tanto, ciertas arquitecturas de proteínas tienen menos probabilidades de compensar el impacto de la mutación de

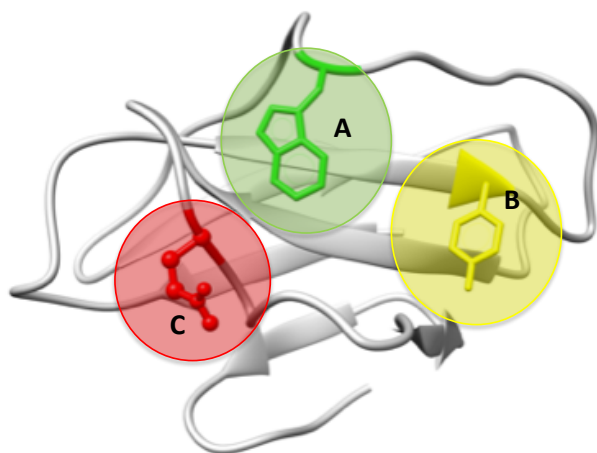


residuos importantes relacionados con la función y posiblemente menos susceptibles a la innovación funcional. Recientemente, Cohen-khait et al. (2017) observaron que tres mutaciones son suficientes para generar un nuevo sitio de unión en una proteína. Sin embargo, también mostraron que tales mutaciones desestabilizan gravemente la proteína, lo que sugiere que la estabilidad limitada de las proteínas silvestres las protege de evolucionar hacia una unión promiscua y no productiva.

## 10. Conclusiones

La caracterización de mutantes puntuales de un LaCBM26 de la  $\alpha$ -amilasa de *L. amylovorus* permitió revelar el papel de residuos polares como el de la glutamina 68, además del ya conocido papel de los aminoácidos aromáticos, ya que para la unión de  $\alpha$ -glucanos es importante tener una plataforma hidrofóbica, así como aminoácidos polares que también tienen un papel crítico en la unión.

De acuerdo con el análisis de las regiones ABC, proponemos que en el dominio LaCBM26 el W32 corresponde a la región A, la Y20 a la región B y la Q68 a la región C necesarios para llevar a cabo la unión de  $\alpha$ -glucanos, aunque hay otros residuos que también participan en la unión, como la Y18.



Por otro lado, ninguna de las mutantes puntuales obtenidas modificaron la estructura secundaria del dominio, aunque se desconoce si estas mutaciones pudieran afectar su flexibilidad; en cambio, la mutación simultánea de aminoácidos importantes en el sitio de unión del dominio LaCBM26 afecta su estructura secundaria y función, sugiriendo que este dominio podría presentar poca tolerancia al cambio.

## 11. Prospectivas

Analizar mediante dinámica molecular el efecto de la mutación de la Q68 en el dominio LaCBM26, y revisar los cambios conformacionales que dan como resultado la pérdida de la función. Así como, restaurar la mutación Q68 en el banco de mutantes para verificar su importancia en la unión de  $\alpha$ -glucanos.

Diseñar las mutaciones necesarias para incrementar la estabilidad del dominio LaCBM26, y conocer las condiciones fisicoquímicas idóneas para realizar estudios estructurales.

Explorar el papel de aminoácidos fuera o a distancia del sitio de unión del dominio LaCBM26 para el diseño de CBMs con mejor afinidad hacia almidón u otras moléculas, por ejemplo, de carácter proteico.

## 12. Referencias

- Abbott, D.W., van Bueren, A.L., 2014. Using structure to inform carbohydrate binding module function. *Curr. Opin. Struct. Biol.* 28, 32–40.  
<https://doi.org/10.1016/j.sbi.2014.07.004>
- Armenta, S., Moreno-Mendieta, S., Sánchez-Cuapio, Z., Sánchez, S., Rodríguez-Sanoja, R., 2017. Advances in molecular engineering of carbohydrate-binding modules. *Proteins Struct. Funct. Bioinforma.* 85, 1602–1617.  
<https://doi.org/10.1002/prot.25327>
- Asensio, J.L., Arda, A., Canada, F.J., Jimenez-Barbero, J., 2013. Carbohydrate - Aromatic interactions. *Acc. Chem. Res.* 46, 946–954.  
<https://doi.org/10.1021/ar300024d>
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* 44, W344–W350. <https://doi.org/10.1093/nar/gkw408>
- Bharali, S., Purama, R., Majumder, A., Fontes, C., Goyal, A., 2007. Functional characterization and mutation analysis of family 11, Carbohydrate-Binding Module (CtCBM11) of cellulosomal bifunctional cellulase from *Clostridium thermocellum*. *Indian J. Microbiol* 47, 109–118.  
<https://doi.org/10.1007/s12088-007-0023-9>
- Bolam, D.N., Xie, H., White, P., Simpson, P.J., Hancock, S.M., Williamson, M.P., Gilbert, H.J., 2001. Evidence for synergy between family 2b carbohydrate binding modules in *Cellulomonas fimi* Xylanase 11A. *Biochemistry* 40, 2468–2477. <https://doi.org/10.1021/bi002564l>

- Boraston, A.B., Bolam, D.N., Gilbert, H.J., Davies, G.J., 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769–781. <https://doi.org/10.1042/BJ20040892>
- Boraston, A.B., Creagh, A.L., Alam, M.M., Kormos, J.M., Tomme, P., Haynes, C.A., Warren, R.A.J., Kilburn, D.G., 2001. Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. *Biochemistry* 40, 6240–6247. <https://doi.org/10.1021/bi0101695>
- Boraston, A.B., Healey, M., Klassen, J., Ficko-Blean, E., Van Bueren, A.L., Law, V., 2006. A structural and functional analysis of  $\alpha$ -glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *J. Biol. Chem.* 281, 587–598. <https://doi.org/10.1074/jbc.M509958200>
- Boraston, A.B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D.R., Kilburn, D.G., Davies, G.J., 2002. Differential oligosaccharide recognition by evolutionarily-related  $\beta$ -1,4 and  $\beta$ -1,3 glucan-binding modules. *J. Mol. Biol.* 319, 1143–1156. [https://doi.org/10.1016/S0022-2836\(02\)00374-1](https://doi.org/10.1016/S0022-2836(02)00374-1)
- Buléon, A., Colonna, P., Planchot, V., Ball, S., 1998. Starch granules: Structure and biosynthesis. *Int. J. Biol. Macromol.* 23, 85–112. [https://doi.org/10.1016/S0141-8130\(98\)00040-3](https://doi.org/10.1016/S0141-8130(98)00040-3)
- Chang, H., Irwin, P.M., Nikolov, Z.L., 1998. Effects of mutations in the starch-binding domain of *Bacillus macerans* cyclodextrin glycosyltransferase. *J. Biotechnol.* 65, 191–202. [https://doi.org/10.1016/S0168-1656\(98\)00115-1](https://doi.org/10.1016/S0168-1656(98)00115-1)
- Chen, W., Erick, S., Price, J.L., Powers, D.L., Powers, E.T., Wong, C., Dyson, H.J.,

- Kelly, W., 2013. Structural and energetic basis of carbohydrate-aromatic packing interactions in proteins. *J. Am. Chem. Soc.* 135, 9877–9884.  
<https://doi.org/10.1021/ja4040472>
- Cid, M., Pedersen, H.L., Kaneko, S., Coutinho, P.M., Henrissat, B., Willats, W.G.T., Boraston, A.B., 2010. Recognition of the helical structure of  $\beta$ -1,4-galactan by a new family of carbohydrate-binding modules. *J. Biol. Chem.* 285, 35999–36009. <https://doi.org/10.1074/jbc.M110.166330>
- Cockburn, D.W., Suh, C., Medina, K.P., Duvall, R.M., Wawrzak, Z., Henrissat, B., Koropatkin, N.M., 2018. Novel carbohydrate binding modules in the surface anchored  $\alpha$ -amylase of *Eubacterium rectale* provide a molecular rationale for the range of starches used by this organism in the human gut. *Mol. Microbiol.* 107, 249–264. <https://doi.org/10.1111/mmi.13881>
- Cohen-khait, R., Dym, O., Hamer-rogotner, S., Cohen-khait, R., Dym, O., Hamer-rogotner, S., Schreiber, G., 2017. Promiscuous protein binding as a function of protein stability. *Structure* 25, 1867–1874.  
<https://doi.org/doi.org/10.1016/j.str.2017.11.002>
- Dhillon, A., Sharma, K., Rajulapati, V., Goyal, A., 2018. The multi-ligand binding first family 35 Carbohydrate Binding Module (CBM35) of *Clostridium thermocellum* targets rhamnogalacturonan I. *Arch. Biochem. Biophys.* 654, 194–208. <https://doi.org/10.1016/j.abb.2018.07.023>
- Din, N., Gilkes, N., Tekant, B., Miller, R.C., Warren, A.J., Kilburn, D.G., 1991. Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. *Nat. Biotechnol.* 9, 1096–1099.
- Ezer, A., Matalon, E., Jindou, S., Borovok, I., Atamna, N., Yu, Z., Morrison, M.,

- Bayer, E.A., Lamed, R., 2008. Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to *Ruminococcus albus*. *J. Bacteriol.* 190, 8220–8222. <https://doi.org/10.1128/JB.00609-08>
- Fernández-Alonso, M.C., Díaz, D., Berbis, M.Á., Marcelo, F., Cañada, J., Jiménez-Barbero, J., 2012. Protein-carbohydrate interactions studied by NMR: from molecular recognition to drug design. *Curr. Protein Pept. Sci.* 13, 816–830. <https://doi.org/10.2174/138920312804871175>
- Fukunaga, K., Taki, M., 2012. Practical tips for construction of custom peptide libraries and affinity selection by using commercially available phage display cloning systems. *J. Nucleic Acids* 2012. <https://doi.org/10.1155/2012/295719>
- Furtado, G.P., Lourenzoni, M.R., Fuzo, C.A., Fonseca-Maldonado, R., Guazzaroni, M.E., Ribeiro, L.F., Ward, R.J., 2018. Engineering the affinity of a family 11 carbohydrate binding module to improve binding of branched over unbranched polysaccharides. *Int. J. Biol. Macromol.* <https://doi.org/10.1016/j.ijbiomac.2018.09.022>
- Gabius, H.J., André, S., Jiménez-Barbero, J., Romero, A., Solís, D., 2011. From lectin structure to functional glycomics: Principles of the sugar code. *Trends Biochem. Sci.* 36, 298–313. <https://doi.org/10.1016/j.tibs.2011.01.005>
- García B. 2015. Estudio termodinámico y estructural de módulos de unión a carbohidratos. Tesis de Doctorado, Posgrado en Ciencias Bioquímicas, Instituto de Química, Universidad Nacional Autónoma de México.
- Georgelis, N., Yennawar, N.H., Cosgrove, D.J., 2012. Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. *Proc. Natl. Acad. Sci.* 109, 14830–14835.

<https://doi.org/10.1073/pnas.1213200109>

Gilbert, H.J., Knox, J.P., Boraston, A.B., 2013. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr. Opin. Struct. Biol.* 23, 669–677.

<https://doi.org/10.1016/j.sbi.2013.05.005>

Giraud, E., Cuny, G., 1997. Molecular characterization of the  $\alpha$ -amylase genes of *Lactobacillus plantarum* A6 and *Lactobacillus amylovorus* reveals an unusual 3' end structure with direct tandem repeats and suggests a common evolutionary origin. *Gene* 198, 149–157. [https://doi.org/10.1016/S0378-1119\(97\)00309-0](https://doi.org/10.1016/S0378-1119(97)00309-0)

Glaring, M.A., Baumann, M.J., Hachem, M.A., Nakai, H., Nakai, N., Santelia, D., Sigurskjold, B.W., Zeeman, S.C., Blennow, A., Svensson, B., 2011. Starch-binding domains in the CBM45 family - Low-affinity domains from glucan, water dikinase and  $\alpha$ -amylase involved in plastidial starch metabolism. *FEBS J.* 278, 1175–1185. <https://doi.org/10.1111/j.1742-4658.2011.08043.x>

Guillén, D., Moreno-Mendieta, S., Aguilera, P., Sánchez, S., Farres, A., Rodríguez-Sanoja, R., 2013. The starch-binding domain as a tool for recombinant protein purification. *Appl. Microbiol. Biotechnol.* 97, 4141–4148.

<https://doi.org/10.1007/s00253-013-4778-0>

Guillén, D., Moreno-mendieta, S., Pérez, R., Espitia, C., Sánchez, S., Rodríguez-sanoja, R., 2014. Starch granules as a vehicle for the oral administration of immobilized antigens. *Carbohydr. Polym.* 112, 210–215.

<https://doi.org/10.1016/j.carbpol.2014.05.089>

Guillén, D., Sánchez, S., Rodríguez-Sanoja, R., 2010. Carbohydrate-binding



- domains : multiplicity of biological roles. *Appl. Microbiol. Biotechnol.* 85, 1241–1249. <https://doi.org/10.1007/s00253-009-2331-y>
- Guillén, D., Santiago, M., Linares, L., Pérez, R., Morlon, J., Ruiz, B., Sánchez, S., Rodríguez-Sanoja, R., 2007. Alpha-amylase starch binding domains: Cooperative effects of binding to starch granules of multiple tandemly arranged domains. *Appl. Environ. Microbiol.* 73, 3833–3837. <https://doi.org/10.1128/AEM.02628-06>
- Gullfot, F., Tan, T., Schantz, L. Von, Karlsson, E.N., Ohlin, M., Brumer, H., Divne, C., 2009. The crystal structure of XG-34, an evolved xyloglucan-specific carbohydrate-binding module. *Struct. Note* 785–789. <https://doi.org/10.1002/prot.22642>
- Gunnarsson, L.C., Dexlin, L., Nordberg, E., Holst, O., Ohlin, M., 2006a. Evolution of a carbohydrate binding module into a protein-specific binder. *Biomol. Eng.* 23, 111–117. <https://doi.org/10.1016/j.bioeng.2005.12.002>
- Gunnarsson, L.C., Karlsson, E.N., Albrekt, A.S., Andersson, M., Holst, O., Ohlin, M., 2004. A carbohydrate binding module as a diversity-carrying scaffold. *Protein Eng. Des. Sel.* 17, 213–221. <https://doi.org/10.1093/protein/gzh026>
- Gunnarsson, L.C., Montanier, C., Tunnicliffe, R.B., Williamson, M.P., Gilbert, H.J., Karlsson, E.N., Ohlin, M., 2007. Novel xylan-binding properties of an engineered family 4 carbohydrate-binding module. *Biochem. J.* 214, 209–214. <https://doi.org/10.1042/BJ20070128>
- Gunnarsson, L.C., Zhou, Q., Montanier, C., Karlsson, E.N., Brumer, H., Ohlin, M., 2006b. Engineered xyloglucan specificity in a carbohydrate-binding module. *Glycobiology* 16, 1171–1180. <https://doi.org/10.1093/glycob/cwl038>

- Haas, W., MacColl, R., Banas, J.A., 1998. Circular dichroism analysis of the glucan binding domain of *Streptococcus mutans* glucan binding protein-A. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* 1384, 112–120.  
[https://doi.org/10.1016/S0167-4838\(98\)00005-3](https://doi.org/10.1016/S0167-4838(98)00005-3)
- Hashimoto, H., 2006. Recent structural studies of carbohydrate-binding modules. *Cell. Mol. Life Sci.* 63, 2954–2967. <https://doi.org/10.1007/s00018-006-6195-3>
- Hemsley, A., Arnheim, N., Toney, M.D., Cortopassi, G., Galas, D.J., 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* 17, 6545–6551.
- Hoof, R., Vriend, G., Sander, C., Abola, E., 1996. Errors in protein structure. *Nature* 381.
- Hudson, K.L., Bartlett, G.J., Diehl, R.C., Agirre, J., Gallagher, T., Kiessling, L.L., Woolfson, D.N., 2015. Carbohydrate-aromatic interactions in proteins. *J. Am. Chem. Soc.* 137, 15152–15160. <https://doi.org/10.1021/jacs.5b08424>
- Jiang, T.Y., Ci, Y.P., Chou, W.I., Lee, Y.C., Sun, Y.J., Chou, W.Y., Li, K.M., Chang, M.D.T., 2012. Two unique ligand-binding clamps of *Rhizopus oryzae* starch binding domain for helical structure disruption of amylose. *PLoS One* 7, 2–14.  
<https://doi.org/10.1371/journal.pone.0041131>
- Kauffmann, C., Shoseyov, O., Shpigel, E., Bayer, E.A., Lamed, R., Shoham, Y., Mandelbaum, R.T., 2000. Novel methodology for enzymatic removal of atrazine from water by CBD- fusion protein immobilized on cellulose. *Environ. Sci. Technol.* 34, 1292–1296. <https://doi.org/10.1021/es990754h>
- Kavoosi, M., Lam, D., Bryan, J., Kilburn, D.G., 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.  
<https://doi.org/10.1016/j.chroma.2007.07.082>
- Kelly, S.M., Jess, T.J., Price, N.C., 2005. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* 1751, 119–139.  
<https://doi.org/10.1016/j.bbapap.2005.06.005>
- Kossmann, J., Lloyd, J., 2000. Understanding and influencing starch biochemistry. *CRC. Crit. Rev. Plant Sci.* 19, 171–226.  
<https://doi.org/10.1080/07352680091139204>
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.  
<https://doi.org/10.1038/228549a0>
- Lammerts Van Bueren, A., Finn, R., Ausió, J., Boraston, A.B., 2004.  $\alpha$ -Glucan recognition by a new family of carbohydrate-binding modules found primarily in bacterial pathogens. *Biochemistry* 43, 15633–15642.  
<https://doi.org/10.1021/bi048215z>
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.  
<https://doi.org/10.1107/S0021889892009944>
- Lehtiö, J., Teeri, T.T., Nygren, P. a, 2000. Alpha-amylase inhibitors selected from a combinatorial library of a cellulose binding domain scaffold. *Proteins* 41, 316–22. [https://doi.org/10.1002/1097-0134\(20001115\)41:3<316::AID-PROT40>3.0.CO;2-K](https://doi.org/10.1002/1097-0134(20001115)41:3<316::AID-PROT40>3.0.CO;2-K)
- Lin, S., Lin, I., Chou, W., Hsieh, C., Liu, S., Huang, R., Sheu, C., Chang, M.D.,

2009. CBM21 starch-binding domain : A new purification tag for recombinant protein engineering. *Protein Expr. Purif.* 65, 261–266.  
<https://doi.org/10.1016/j.pep.2009.01.008>
- Linder, M., Lindeberg, G., Reinikainen, T., Teeri, T.T., Pettersson, G., 1995. The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. *FEBS Lett.* 372, 96–98.  
[https://doi.org/10.1016/0014-5793\(95\)00961-8](https://doi.org/10.1016/0014-5793(95)00961-8)
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, 490–495. <https://doi.org/10.1093/nar/gkt1178>
- Luís, A.S., Venditto, I., Temple, M.J., Rogowski, A., Baslé, A., Xue, J., Knox, J.P., Prates, J.A.M., Ferreira, L.M.A., Fontes, C.M.G.A., Najmudin, S., Gilbert, H.J., 2013. Understanding how noncatalytic carbohydrate binding modules can display specificity for xyloglucan. *J. Biol. Chem.* 288, 4799–4809.  
<https://doi.org/10.1074/jbc.M112.432781>
- Majzlová, K., Janeček, Š., 2014. Two structurally related starch-binding domain families CBM25 and CBM26. *Biologia (Bratisl).* 69, 1087–1096.  
<https://doi.org/10.2478/s11756-014-0415-3>
- Malik, P., Terry, T.D., Bellintani, F., Perham, R.N., 1998. Factors limiting display of foreign peptides on the major coat protein of filamentous bacteriophage capsids and a potential role for leader peptidase. *FEBS Lett.* 436, 263–266.  
[https://doi.org/10.1016/S0014-5793\(98\)01140-5](https://doi.org/10.1016/S0014-5793(98)01140-5)
- Masuko, T., Minami, A., Iwasaki, N., Majima, T., Nishimura, S.I., Lee, Y.C., 2005. Carbohydrate analysis by a phenol-sulfuric acid method in microplate format.

- Anal. Biochem. 339, 69–72. <https://doi.org/10.1016/j.ab.2004.12.001>
- McLean, B.W., Boraston, A.B., Brouwer, D., Sanaie, N., Fyfe, C.A., Warren, R.A.J., Kilburn, D.G., Haynes, C.A., 2002. Carbohydrate-binding modules recognize fine substructures of cellulose. *J. Biol. Chem.* 277, 50245–50254. <https://doi.org/10.1074/jbc.M204433200>
- Montanier, C., Flint, J.E., Bolam, D.N., Xie, H., Liu, Z., Rogowski, A., Weiner, D.P., Ratnaparkhe, S., Nurizzo, D., Roberts, S.M., Turkenburg, J.P., Davies, G.J., Gilbert, H.J., 2010. Circular permutation provides an evolutionary link between two families of calcium-dependent carbohydrate binding modules. *J. Biol. Chem.* 285, 31742–31754. <https://doi.org/10.1074/jbc.M110.142133>
- Montanier, C.Y., Correia, M.A.S., Flint, J.E., Zhu, Y., Baslé, A., McKee, L.S., Prates, J.A.M., Polizzi, S.J., Coutinho, P.M., Lewis, R.J., Henrissat, B., Fontes, C.M.G.A., Gilbert, H.J., 2011. A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization. *J. Biol. Chem.* 286, 22499–22509. <https://doi.org/10.1074/jbc.M110.217372>
- Moreno-mendieta, S.A., Guillén, D., Espitia, C., Hernández-pando, R., Sanchez, S., Rodríguez-sanoja, R., 2014. A novel antigen-carrier system : The *Mycobacterium tuberculosis* Acr protein carried by raw starch microparticles. *Int. J. Pharm.* 474, 241–248. <https://doi.org/10.1016/j.ijpharm.2014.07.041>
- Morlon-Guyot, J., Mucciolo-Roux, F., Sanoja, R.R., Guyot, J.P., 2001. Characterization of the *L. manihotivorans*  $\alpha$ -amylase gene. *Mitochondrial DNA* 12, 27–37. <https://doi.org/10.3109/10425170109042048>

- Motherway, M.O.C., Fitzgerald, G.F., Neiryneck, S., Ryan, S., Steidler, L., Van Sinderen, D., 2008. Characterization of ApuB, an extracellular type II amylopullulanase from *Bifidobacterium breve* UCC2003. *Appl. Environ. Microbiol.* 74, 6271–6279. <https://doi.org/10.1128/AEM.01169-08>
- Oliveira, C., Carvalho, V., Domingues, L., Gama, F.M., 2015. Recombinant CBM-fusion technology - Applications overview. *Biotechnol. Adv.* 33, 358–369. <https://doi.org/10.1016/j.biotechadv.2015.02.006>
- Ota, C., Ikeguchi, M., Tanaka, A., Hamada, D., 2016. Residual structures in the unfolded state of starch-binding domain of glucoamylase revealed by near-UV circular dichroism and protein engineering techniques. *Biochim. Biophys. Acta - Proteins Proteomics* 1864, 1464–1472. <https://doi.org/10.1016/j.bbapap.2016.05.002>
- Pedroza-Roldan, C., Charles-Niño, C., Saavedra, R., Govezensky, T., Vaca, L., Avannis-Aghajani, E., Gevorkian, G., Manoutcharian, K., 2009. Variable epitope library-based vaccines: Shooting moving targets. *Mol. Immunol.* 47, 270–282. <https://doi.org/10.1016/j.molimm.2009.09.024>
- Plevin, M.J., Bryce, D.L., Boisbouvier, J., 2010. Direct detection of CH/ $\pi$  interactions in proteins. *Nat. Chem.* 2, 466–471. <https://doi.org/10.1038/nchem.650>
- Ramos, R., Domingues, L., Gama, M., 2010. *Escherichia coli* expression and purification of LL37 fused to a family III carbohydrate-binding module from *Clostridium thermocellum*. *Protein Expr. Purif.* 71, 1–7. <https://doi.org/10.1016/j.pep.2009.10.016>
- Rivkin, A., Abitbol, T., Nevo, Y., Verker, R., Lapidot, S., Komarov, A., Veldhuis,

- S.C., Zilberman, G., Reches, M., Cranston, E.D., Shoseyov, O., 2015. Bionanocomposite films from resilin-CBD bound to cellulose nanocrystals. *Ind. Biotechnol.* 11, 44–58. <https://doi.org/10.1089/ind.2014.0026>
- Rodríguez-Romero, A., Arreguín, B., Hernández-Arana, A., 1989. Unusual far-ultraviolet circular dichroism of wheat germ agglutinin and hevein originated from cystine residues. *Biochim. Biophys. Acta* 998, 21–24.
- Rodríguez-Sanoja, R., Oviedo, N., Escalante, L., Ruiz, B., Sánchez, S., 2009. A single residue mutation abolishes attachment of the CBM26 starch-binding domain from *Lactobacillus amylovorus*  $\alpha$ -amylase. *J. Ind. Microbiol. Biotechnol.* 36, 341–346. <https://doi.org/10.1007/s10295-008-0502-y>
- Rodríguez-Sanoja, R., Oviedo, N., Sánchez, S., 2005. Microbial starch-binding domain. *Curr. Opin. Microbiol.* 8, 260–267. <https://doi.org/10.1016/j.mib.2005.04.013>
- Rodríguez Sanoja, R., Morlon-Guyot, J., Jore, J., Pintado, J., Juge, N., Guyot, J.P., 2000. Comparative characterization of complete and truncated forms of *Lactobacillus amylovorus*  $\alpha$ -amylase and role of the c-terminal direct repeats in raw-starch binding. *Appl. Environ. Microbiol.* 66, 3350–3356. <https://doi.org/10.1128/AEM.66.8.3350-3356.2000>
- Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Salama-Alber, O., Jobby, M.K., Chitayat, S., Smith, S.P., White, B.A., Shimon, L.J.W., Lamed, R., Frolow, F., Bayer, E.A., 2013. Atypical cohesin-dockerin

- complex responsible for cell surface attachment of cellulosomal components: Binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.* 288, 16827–16838. <https://doi.org/10.1074/jbc.M113.466672>
- Schantz, L. Von, Gullfot, F., Scheer, S., Filonova, L., Gunnarsson, L.C., Flint, J.E., Daniel, G., Nordberg-karlsson, E., Brumer, H., Ohlin, M., 2009. Affinity maturation generates greatly improved xyloglucan-specific carbohydrate binding modules. *BMC Biotechnol.* 9, 92. <https://doi.org/10.1186/1472-6750-9-92>
- Sidhu, S.S., Weiss, G.A., Wells, J.A., 2000. High copy display of large proteins on phage for functional selections. *J. Mol. Biol.* 296, 487–95. <https://doi.org/10.1006/jmbi.1999.3465>
- Simpson, P.J., Jamieson, S.J., Abou-Hachem, M., Nordberg Karlsson, E., Gilbert, H.J., Holst, O., Williamson, M.P., 2002. The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. *Biochemistry* 41, 5712–5719. <https://doi.org/10.1021/bi012093i>
- Skerra, A., 2000. Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* 13, 167–187. [https://doi.org/10.1002/1099-1352\(200007/08\)13:4<167::AID-JMR502>3.0.CO;2-9](https://doi.org/10.1002/1099-1352(200007/08)13:4<167::AID-JMR502>3.0.CO;2-9)
- Smith, G.P., Patel, S.U., Windass, J.D., Thornton, J.M., Winter, G., Griffiths, a D., 1998. Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage. *J. Mol. Biol.* 277, 317–332. <https://doi.org/10.1006/jmbi.1997.1621>
- Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33,



244–248. <https://doi.org/10.1093/nar/gki408>

Sugimoto, N., Igarashi, K., Samejima, M., 2012. Cellulose affinity purification of fusion proteins tagged with fungal family 1 cellulose-binding domain.

PROTEIN Expr. Purif. 82, 290–296. <https://doi.org/10.1016/j.pep.2012.01.007>

Tomme, P., Boraston, A., Kormos, J.M., Warren, R.A.J., Kilburn, D.G., 2000.

Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules. Enzyme Microb. Technol. 27, 453–458. [https://doi.org/10.1016/S0141-0229\(00\)00246-5](https://doi.org/10.1016/S0141-0229(00)00246-5)

Tsuzuki, S., Honda, K., Uchimarui, T., Mikami, M., Tanabe, K., 2000. The

magnitude of the CH/ $\pi$  interaction between benzene and some model hydrocarbons. J. Am. Chem. Soc. 122, 3746–3753.

<https://doi.org/10.1021/ja993972j>

Vaaje-Kolstad, G., Horn, S.J., Van Aalten, D.M.F., Synstad, B., Eijsink, V.G.H.,

2005. The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. J. Biol. Chem. 280, 28492–28497. <https://doi.org/10.1074/jbc.M504468200>

Valdez, H.A., Busi, M. V., Wayllace, N.Z., Parisi, G., Ugalde, R.A., Gomez-Casati,

D.F., 2008. Role of the N-terminal starch-binding domains in the kinetic properties of starch synthase III from *Arabidopsis thaliana*. Biochemistry 47, 3026–3032. <https://doi.org/10.1021/bi702418h>

Valk, V., Eeuwema, W., Sarian, F.D., van der Kaaij, R.M., Dijkhuizen, L., 2015.

Degradation of granular starch by the bacterium *Microbacterium aurum* strain B8.A involves a modular  $\alpha$ -amylase enzyme system with FNIII and CBM25 domains. Appl. Environ. Microbiol. 81, 6610–6620.

<https://doi.org/10.1128/AEM.01029-15>

Valk, V., Lammerts van Bueren, A., van der Kaaij, R.M., Dijkhuizen, L., 2016.

Carbohydrate-binding module 74 is a novel starch-binding domain associated with large and multidomain  $\alpha$ -amylase enzymes. *FEBS J.* 283, 2354–2368.

<https://doi.org/10.1111/febs.13745>

Vamadevan, V., Bertoft, E., 2015. Structure-function relationships of starch components. *Starch/Staerke* 67, 55–68.

<https://doi.org/10.1002/star.201400188>

Wernéus, H., Lehtiö, J., Teeri, T., Nygren, P.Å., Ståhl, S., 2001. Generation of metal-binding *Staphylococci* through surface display of combinatorially engineered cellulose-binding domains. *Appl. Environ. Microbiol.* 67, 4678–4684. <https://doi.org/10.1128/AEM.67.10.4678-4684.2001>

Williamson, G., Belshaw, N.J., 1993. Specificity of the binding domain of glucoamylase 1. *Eur. J. Biochem.* 211, 717–724.

Wiseman, T., Williston, S., Brandts, J.F., Lin, L.N., 1989. Rapid measurement of binding constants and heats of binding using a new titration calorimeter.

*Anal. Biochem.* 179, 131–137. [https://doi.org/10.1016/0003-2697\(89\)90213-3](https://doi.org/10.1016/0003-2697(89)90213-3)

Woody, R.W., 1994. Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins. *Eur Biophys J* 23, 253.

<https://doi.org/10.1007/BF00213575>

Xie, H., Gilbert, H.J., Charnock, S.J., Davies, G.J., Williamson, M.P., Simpson, P.J., Raghothama, S., Fontes, C.M.G.A., Dias, F.M. V, Ferreira, L.M.A., Bolam, D.N., 2001. *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: the role of conserved amino acids in ligand binding.

- Biochemistry 40, 9167–9176. <https://doi.org/10.1021/bi0106742>
- Xu, X., Visser, R.G.F., Trindade, L.M., 2014. Starch modification by biotechnology: state of art and perspectives. *Starch Polym. From Genet. Eng. to Green Appl.* 79–102. <https://doi.org/10.1016/B978-0-444-53730-0.00021-X>
- Yamaguchi, R., Arakawa, T., Tokunaga, H., Ishibashi, M., Tokunaga, M., 2012. Distinct characteristics of single starch-binding domain SBD1 derived from tandem domains SBD1-SBD2 of halophilic *Kocuria varians* alpha-amylase. *Protein J.* 31, 250–258. <https://doi.org/10.1007/s10930-012-9400-2>
- Yang, J., Yan, R., Roy, A., Xu, D., J, P., Zhang, Y., 2015. The I-TASSER Suite: Protein structure and function prediction. *Nat Methods* 12, 7–8. <https://doi.org/10.1038/nmeth.3213>
- Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9, 40. <https://doi.org/10.1186/1471-2105-9-40>
- Zhang, Y., Chen, S., Xu, M., Cavoco-Paulo, A.P., Wu, J., Chen, J., 2010. Characterization of *Thermobifida fusca* cutinase-carbohydrate-binding module fusion proteins and their potential application in bioscouring. *Appl. Environ. Microbiol.* 76, 6870–6876. <https://doi.org/10.1128/AEM.00896-10>
- Zhang, Y., Wang, L., Chen, J., Wu, J., 2013. Enhanced activity toward PET by site-directed mutagenesis of *Thermobifida fusca* cutinase-CBM fusion protein. *Carbohydr. Polym.* 97, 124–129. <https://doi.org/10.1016/j.carbpol.2013.04.042>

# ANEXOS

## **Anexo 1.** Composición de amortiguadores y medios de cultivo.

### **Medio 2YT**

Cantidad en g para 1 L.

17 g      Triptona  
10 g      Extracto de Levadura  
5 g      NaCl

### **PEG/NaCl [20% / 2.5M]**

20 g de PEG 8000

14.6 g de NaCl

Disolver por separado, mezclar y aforar a 100 mL.

Esterilizar por filtración o en autoclave por 15 min/121 °C/15 lb.

Almacenar a 4 °C.

### **Amortiguador TBS 1X**

50 mM      Tris-HCl

150 mM     NaCl

### **Amortiguador PBS 1X**

10 mM      Na<sub>2</sub>HPO<sub>4</sub>

2 mM      KH<sub>2</sub>PO<sub>4</sub>

137 mM     NaCl

2.7 mM     KCl

## **Anexo 2.** Protocolo para la construcción y despliegue de la biblioteca del dominio LaCBM26.

### Construcción de la biblioteca combinatoria.

1. Primero se realizó el ensamblaje adicionando 20 pmol de cada uno de siete oligonucleótidos (Tabla A1). Los oligonucleótidos fueron ensamblados mediante una reacción de PCR (95°C por 1 min, 56°C por 2 min, 68°C por 2 min, ocho ciclos) usando la DNA polimerasa Platinum® (Thermo Scientific).
2. De la reacción anterior se tomaron como templado alícuotas de 0.5 µL, 1.0 µL y 3 µL. La reacción de PCR (98°C por 30 s, 59°C por 30 s, 72°C por 30 s, 30 ciclos) se realizó con 25 pmol de los oligos Forward: 5'-CAGCAGCCATGGAACAACAGAAAC-3' y Reverse: 5'-TTAAGTTAGGATCCTACGTATTTTAC-3' que contiene los sitios de las enzimas de restricción *NcoI* y *BamHI*, respectivamente. El amplificado más limpio se obtuvo adicionando como templado una alícuota de 0.5 µL. Se realizaron 10 reacciones de PCR más para obtener suficiente cantidad de amplificado.
3. Posteriormente, se digirió el amplificado de la biblioteca con 200 U de cada una de las enzimas *NcoI* y *BamHI* en un volumen total de 400 µL. La reacción se incubó por 2h a 37 °C. Se siguió la reacción mediante la visualización en gel de agarosa al 2%. Después se purificó la reacción de digestión con el Kit DNA Clean and Concentrator-S™.
4. Simultáneamente, el fagémido pG8SAET fue purificado y digerido como se mencionó anteriormente. Para la limpieza del fagémido digerido se realizó el método de fenol-cloroformo.
5. Previo a la clonación de la biblioteca en el fagémido, se transformaron 50 µL de células DH5α electrocompetentes con 100ng de pG8SAET digerido para revisar la correcta digestión del vector. No se obtuvieron colonias al incubar las cajas de medio LB-agar por 24 h a 37 °C.
6. Para la clonación de la biblioteca en el fagémido pG8SAET se utilizaron 1.5 µg de fragmento digerido, purificado y revisado en un gel de agarosa al 2% en una relación 3:1 con el fagémido, utilizando 5 µg de vector digerido y purificado. La ligación se realizó con la enzima ligasa T4 (NE), la reacción se incubó a temperatura ambiente por 30 min y posteriormente toda la noche a 16 °C. Enseguida se desactivó la enzima incubando la reacción a 65 °C por 10 min.

7. Posteriormente se limpió la reacción de ligación con el kit Promega (Wizard®SV Gel and PCR Clean-up System), se obtuvieron 100 µL.
8. A partir de la solución anterior, que contiene el producto de la ligación, se realizaron varias pruebas de electroporación con células de *E. coli* TGI electrocompetentes. Se transformaron 45 µL de células con 0.5 µL, 1 µL, 2 µL, 3 µL y 4 µL.  
De acuerdo con la ecuación:

$$\begin{aligned}
 (\# \text{ de Colonias}) \left( \frac{\mu\text{L de cultivo}}{\mu\text{L plaqueados}} \right) \left( \frac{\mu\text{L de rxn de ligación}}{\mu\text{L de la ligación para transformar}} \right) \\
 = \text{total de transformantes.}
 \end{aligned}$$

se determinó que realizar 16 electroporaciones con 3 µL de la reacción de ligación era suficiente para obtener una librería teórica de  $3.2 \times 10^6$  transformantes.

9. Por lo tanto, se realizaron 16 electrotransformaciones adicionando 3 µL de la reacción de ligación purificada a 45 µL de células electrocompetentes de *E. coli* TGI. Cada tubo se incubó en hielo 1 min, se colocaron los 48 µL en una celda de electroporación nueva y fría. Después de dar un pulso de 1250 V se agregó 1 mL de medio SOC. Se colectó cada transformación en cuatro tubos falcon en grupos de 4 mL para incubar de forma secuencial a 37 °C durante 40 min. Colectando al final en un solo tubo donde se tiene la librería completa. Para estimar la variabilidad y el tamaño se tomaron 3 µL y 30 µL y se diluyen a 150 µL con medio SOC y se plaquean en cajas de medio LB-Amp (100 µg/mL). Las cajas se incuban a 37 °C por 24 h. Se cuantifican las colonias obtenidas.
10. Al volumen restante del punto anterior se le adicionó un mismo volumen de medio 2YT fresco y se incubó sin antibiótico a 37 °C por 1h y 250 rpm de agitación.

Posteriormente se adicionan Ampicilina a una concentración final de 100 µg/mL y se incubó por 2 h más en las condiciones anteriores.

11. Del tubo anterior se tomaron 0.5 mL y se adicionaron a 6 mL de medio 2YT-Amp (100 mg/mL). EL tubo se incubó a 37 °C y agitación de 250 rpm por 16 h.
12. De este tubo se realizó la extracción del DNA plasmídico de la librería, de acuerdo con las indicaciones del fabricante (QIAGEN) y se almacenó a -20 °C.

Tabla A1. Oligonucleótidos utilizados en la construcción de la biblioteca combinatoria del dominio CBM26 de *Lactobacillus amylovorus*.

Rv1	5'-TGCTACCCCAGCTTGACGGCTTTTCGAAGTAAACCTTTTTAGTTTCTGTTGTTCCATGGCTGCTG
Fw2a	5'-CAAGCTGGGGTAGCAGAGTGTATGCCNNKGTANNKAATAAGAACACGAACAAGG <sup>a</sup>
Rv2a	5'-CTAAAGCGGTCATTTTTTTGCCAGGCCAAGCTGAAGTTATAGCCTTGTTCTGTGTTTC
Fw3	5'-AAATGACCGCTTTAGGTAACGACAAGTATGAATTGGATCTCGACACTGATGAAG
Rv3a	5'-GTCCCATCGGTGAATATAACAGCTAAATCAGAGTCATCTTCATCAGTGTCGA
Fw4b	5'-ATTCACCGATGGGACTAAGNNKACACCAGCAGCTAATNNKGCAGGTNNKACATTCACAGCAGA <sup>a</sup>
Rv4b	5'-TTAAGTTAGGATCCTACGTATTTTACGACACCATTTTGGTCATAAGTGGCATCTGCTGTGAATGT

#### Producción de partículas.

##### Día 1.

1. A partir de un glicerol, inocular 5 mL de medio 2-YT/Amp (100 µg/mL) con 10 µL o bien picar una colonia, e incubar OVN a 37 °C.

##### Día 2.

2. Del cultivo anterior tomar 50 µL e inocular 5 mL de medio 2-YT/Amp (100 µg/mL) fresco (Dilución 1:100), e incubar a 37 °C hasta alcanzar una D.O = 0.7 aproximadamente.
3. Al cultivo anterior, agregar fago helper (M13K07) a 10 MOI diluído en 100 µL de medio 2-YT/Amp (100 µg/mL).
4. Incubar por 30 min a 37 °C SIN AGITACIÓN y 30 min más CON AGITACIÓN a 250 rpm.
5. Tomar 2 mL del cultivo anterior e inocular a 48 mL de medio 2-YT/Amp (100 µg/mL) / kanamicina (50 µg/mL) fresco (en matraces de 250 mL) e incubar a 37 °C durante 16-20 h con agitación (250 rpm).

Nota: A partir del cultivo anterior se realiza el rescate y titulación del fago, para ser utilizado en la siguiente ronda de selección.

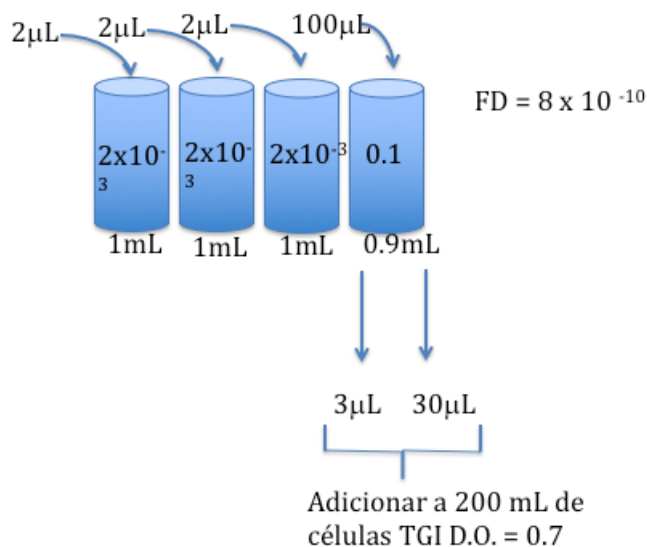
Día 3.

6. El volumen de cada matraz se pasa a dos tubos para centrifuga estériles de 50 mL. Centrifugar a 8000 rpm a 4 °C por 20 min. Posteriormente vaciar el sobrenadante en un tubo limpio y adicionar 1/6 del volumen de PEG 20%/NaCl 2.5 M para concentrar las partículas de fagos, mezclar por inversión suave de 4-6 veces.
7. Incubar a 4 °C durante 2 h o toda la noche.
8. Centrifugar a 10 000 rpm a 4 °C durante 10 min y decantar los tubos. “Ahora las partículas están en el pellet”.
9. Adicionar a cada tubo 1 mL de amortiguador TBS y resuspender las partículas, los tubos son lavados con 0.5 mL más de amortiguador. Los 1.5 mL se depositan en un tubo microfuga de 2 mL limpio y estéril.
10. A cada tubo se le adicionan 0.3 mL de PEG 20%/NaCl 2.5 M y se mezcla por inversión suave. Los tubos se incuban a 4 °C por 1 h.
11. Centrifugar a 15 000 rpm durante 15 min a 4 °C. Decantar el sobrenadante, repetir centrifugación para eliminar el total del líquido.
12. Resuspender las partículas con 0.25 mL de amortiguador TBS y 0.1 mL más para lavar el tubo y recuperar el volumen total sólo en un tubo, obteniendo 600 µL.
13. Finalmente centrifugar a 4000 rpm por 5 min a 4 °C para eliminar cualquier agregado de partículas o células.



Título de los fagos.

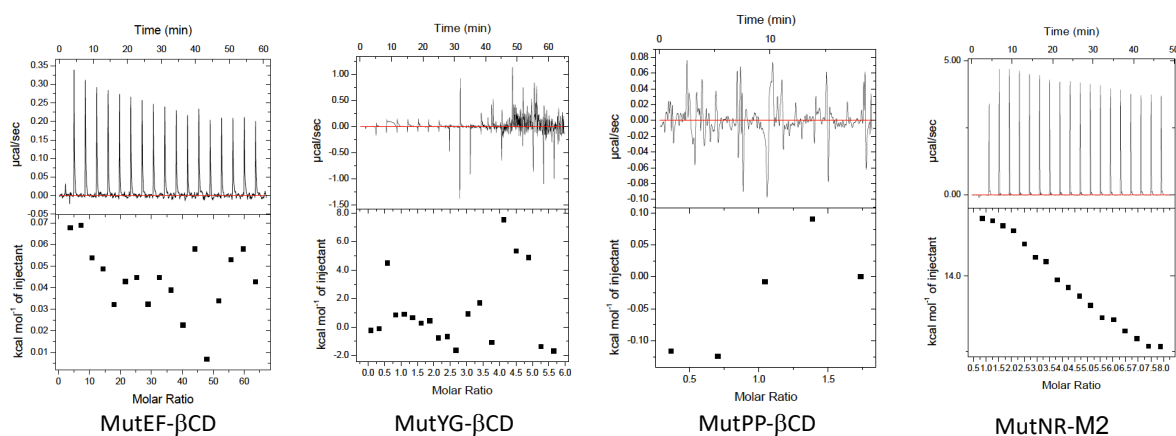
1. Realizar diluciones de acuerdo con el siguiente ejemplo:



2. Después de realizar las diluciones, incubar durante 40 min a 37 °C.

3. Plaquear todo el volumen en cajas de LB-Amp (100  $\mu$ g/mL) e incubar toda la noche a 37 °C. Se cuantifican el número de colonias y se calcula el total de pfc/ $\mu$ L.

**Anexo 3.** ITC de variantes del dominio LaCBM26 sin la capacidad de unir almidón.



**Anexo 4.** Documentos publicados en revistas indexadas.



## The role of conserved non-aromatic residues in the *Lactobacillus amylovorus* $\alpha$ -amylase CBM26-starch interaction

Silvia Armenta<sup>a,b</sup>, Zaira Sánchez-Cuapio<sup>a,c</sup>, Maria Elena Munguia<sup>a</sup>, Nancy O. Pulido<sup>d</sup>, Amelia Farrés<sup>e</sup>, Karen Manoutcharian<sup>a</sup>, Alejandra Hernandez-Santoyo<sup>f</sup>, Silvia Moreno-Mendieta<sup>g</sup>, Sergio Sánchez<sup>a</sup>, Romina Rodríguez-Sanoja<sup>a,\*</sup>

<sup>a</sup> Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Circuito Mario de la Cueva s/n, Ciudad Universitaria, Ciudad de México 04510, Mexico

<sup>b</sup> Programa de Doctorado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México (UNAM), A.P.70228, Ciudad Universitaria, México, DF 04510, Mexico

<sup>c</sup> Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), A.P.70228, Ciudad Universitaria, México, DF 04510, Mexico

<sup>d</sup> Centro de Investigaciones Químicas, Universidad Nacional Autónoma del Estado de Morelos, Av. Universidad s/n, Cuernavaca, Mor 62210, Mexico

<sup>e</sup> Facultad de Química, Universidad Nacional Autónoma de México, Circuito Exterior s/n Ciudad Universitaria, Ciudad de México 04510, Mexico

<sup>f</sup> Instituto de Química, Universidad Nacional Autónoma de México, Circuito Exterior s/n Ciudad Universitaria, Ciudad de México 04510, Mexico

<sup>g</sup> CONACYT, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Circuito Mario de la Cueva s/n, Ciudad Universitaria, Ciudad de México 04510, Mexico

### ARTICLE INFO

#### Article history:

Received 12 May 2018

Received in revised form 12 October 2018

Accepted 14 October 2018

Available online 15 October 2018

#### Keywords:

Carbohydrate-binding module

Starch

$\alpha$ -Glucans recognition

### ABSTRACT

It is generally accepted that carbohydrate binding modules (CBMs) recognize their carbohydrate ligands by hydrophobic and CH- $\pi$  interactions. Point mutations of one CBM26 of the *Lactobacillus amylovorus*  $\alpha$ -amylase starch-binding domain (*LaCBM26*) showed that conserved non-aromatic residues are essential in the starch recognition function of the domain, as the mutation of a single glutamine (Q68L) eliminates binding to starch and  $\beta$ -cyclodextrin, even in the presence of aromatic amino acids necessary for ligand binding. The secondary structure of mutated proteins was verified and showed no differences from the wild-type domain. However, random mutations of five residues involved in binding (Y18, Y20, Q68, E74, and F77) did cause change in the secondary structure of the protein, which also causes loss of function. Much of the diversity introduced in the *LaCBM26* was probably incompatible with the appropriate folding of these proteins, suggesting that the domain has little tolerance to change.

© 2018 Elsevier B.V. All rights reserved.

### 1. Introduction

Starch is one of the most abundant carbohydrates in nature. It is generally found in the form of insoluble granules that can be degraded only by enzymes containing substrate-binding regions on their surface, or adjacent carbohydrate-binding modules (CBM) that allow the adsorption of the enzyme by the insoluble substrate [1–3]. CBMs are protein domains that have no catalytic activity and can specifically recognize and bind not only to starch but also to almost any other known carbohydrate in nature. CBMs are classified into 83 families on the basis of amino acid sequence. Specifically, the starch-binding domains (SBDs) are included in 15 families (CBM20, CBM21, CBM25, CBM26, CBM34, CBM41, CBM45, CBM48, CBM53, CBM58, CBM68, CBM69, CBM74, CBM82, CBM83) known for their binding preference to starch and/or maltooligosaccharides [4,5].

The primary structures of the SBDs contain few consensus residues, even among members of the same family. For example, in family

CBM26, <30% of residues are conserved, however the three essential aromatic positions are present in all members, normally two tyrosines and one tryptophan. Other residues indicated as functionally important are conserved, but not in any member of the family. Nevertheless, most SBDs are folded into a highly conserved three-dimensional structure, with  $\beta$ -sandwich folding as the most reported structure [6,7]. Usually, the substrate binding site is a groove located in a concave portion of the protein, which fits linear  $\alpha$ -glucans [8,9].

Abbot et al. [9] proposed three spatially and functionally conserved regions (A–C) for  $\alpha$ -glucan-binding CBMs. A and B form a hydrophobic platform in which A is invariably a tryptophan, while B may be a residue of tryptophan or tyrosine. By contrast, C is a polar residue that forms hydrogen bonds with the substrate. The CBM26 structure (PDB:2C3H) from the *Bacillus halodurans* C125  $\alpha$ -amylase (*BhCBM26*) shows six residues interacting with maltose: W36, Y23, Y25, Q71, G76 and E77. The binding is based on van der Waals interactions with residues Y25 and W36, stabilized by 5 hydrogen bonds with the aromatic residue (Y23), the polar residues (Q71 and E77) and the hydrophobic residue (G76) [10]. In *BhCBM26*, W36 corresponds to region A, Y25 to region B, and E77 to region C; however, the importance of all residues involved in

\* Corresponding author.

E-mail address: [romina@biomedicas.unam.mx](mailto:romina@biomedicas.unam.mx) (R. Rodríguez-Sanoja).

the binding interaction is unclear [9,10]. Recently, the tandem structure of two CBM26 from the  $\alpha$ -amylase of *Eubacterium rectale* was determined (PDB:6B15; with maltoheptaose PDB: 6B3P). Both domains interact equally with the substrate through an aromatic platform formed by two aromatic residues (*Er*CBM26-1: W208 and Y198; *Er*CBM26-2: W314 and W299) and hydrogen bonds formed by 3–5 amino acids (*Er*CBM26-1: Q247, Y196, D250, N202, and E200; *Er*CBM26-2: Q352, Y297, and N355) [5].

*Lactobacillus amylovorus*  $\alpha$ -amylase SBD (*La*SBD) has an unusual modular arrangement, composed of five identical CBM26 tandem modules [3]. Previous site-directed mutagenesis studies conducted in one of these identical modules (*La*CBM26) suggests that its starch-binding capacity relies on 3 aromatic residues: Y18, Y20 and W32; where W32 is essential, since the W32 L mutated protein loses the ability to bind soluble starch, while Y18L and Y20L still bind starch but do so less effectively than the WT domain [11]. This and other works have emphasized the importance of aromatic amino acids as mediators of CBM-carbohydrate interactions through hydrophobic and CH- $\pi$  interactions with the substrate, while polar amino acids help stabilize the binding by forming hydrogen bonds with the carbohydrate [12–14]. However, to the best of our knowledge, there is no information regarding the functional importance of the hydrogen-bond-forming residues in the binding site.

For SBDs without known three-dimensional structure, it is difficult to predict the position and functional importance of binding residues. The *La*CBM26 domain undergoes a process of self-proteolysis that hinders its crystal formation (unpublished results); consequently, a possible approximation to understand the interaction of this SBD with its substrate is the point or simultaneous substitution of the residues that may be important for substrate interaction. In this study, we report the importance of the binding residues in the structure and function of a *La*CBM26.

## 2. Materials and methods

### 2.1. Strains and plasmids

*E. coli* strain DH5 $\alpha$  was used as host for pQ-1mod, a preexisting construct containing the gene encoding one CBM26 from the *L. amylovorus*  $\alpha$ -amylase starch-binding domain [15]. *E. coli* strains XL10 Gold and Rosetta 2 (DE3) were used for production of WT-CBM26, and their mutated proteins. *E. coli* TGI was used as host for construction of the library and for phage display work. The pG8SAET phagemid vector was used for CBM26 library construction fused to *gene VIII*, which codes for protein VIII of phage M13. Genes coding for selected clones were inserted in the expression vector pET-22b(+) (Novagen, Madison, WI, USA).

### 2.2. Protein modeling of the *L. amylovorus* $\alpha$ -amylase CBM26 and model validation

Protein modeling was performed using Iterative Threading ASSEmily Refinement server (I-TASSER, <http://zhanglab.cmb.med.umich.edu/I-TASSER>) based on the *ab initio*/threading method [16,17]. The model was subjected to conjugate gradient energy minimizations using the Amber 16 package with Amber ff99SB-ILDN protein force field (<http://ambermd.org/AmberModels.php>). The structural validation of the three-dimensional model was done using different validation tools: PROCHECK (<https://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>) [18], WHAT\_CHECK (<https://swift.cmbi.umcn.nl/gv/whatcheck/>) [19] and ERRAT online servers (<http://servicesn.mbi.ucla.edu/ERRAT/>). The overall quality of the protein structure was obtained by analyzing the percentage of residues in the most favorable or core regions in the Ramachandran plot. In the generated model, >85% of the residues were in correct regions in preferred/allowed regions, showing

that the model has high quality. The model was deposited in the Protein Model Database (PMDb) [20] with PMDB identifier PM0081297.

### 2.3. Site-directed mutagenesis of the *L. amylovorus* $\alpha$ -amylase CBM26

Point mutations were introduced to change selected amino acids to leucine. Mutations were generated using Quick Change Mutagenesis Kit (Stratagene, La Jolla, CA) with two antiparallel primers containing the desired mutation and pQ-1mod as DNA template. The primers containing the desired mutation are shown in Table 1. PCRs were performed using Phire Hot Start II DNA polymerase (F122S Thermo Scientific).

After PCR, the products were treated with *Dpn I* to digest the parental DNA template and then transformed into *E. coli* XL10-Gold cells. All constructs were validated by DNA sequencing (Laboratorio de Secuenciación Genómica de la Biodiversidad y de la Salud, UNAM).

### 2.4. Protein production and purification

Cells were grown in LB medium with 100 mg/L ampicillin for 6 h at 37 °C for a DO<sub>600</sub> ~0.8. Protein expression was induced with the addition of isopropyl- $\beta$ -D-thiogalactopyranoside (0.4 mM IPTG) (Sigma-Aldrich, St Louis, MO), and incubation was continued for 2 h at 37 °C. Cells were harvested by centrifugation (8000  $\times$ g for 10 min at 4 °C) and suspended to approximately 1/100 of the original culture volume in binding buffer (20 mM sodium phosphate pH 7.4, 0.5 M NaCl and 25 mM imidazole) with a protease inhibitor cocktail for use in purification of histidine-tagged proteins (P8849; Sigma-Aldrich, St. Louis, MO).

Cells were disrupted by sonication with pulses of 30 s at 60 Hz and pause of 30 s during 30 min (Vibra-Cell; Sonics & Materials, Inc). Protein was purified from the clarified extracts by immobilized metal affinity chromatography according to the manufacturer's protocols (ProBond Resin 46-0019, Invitrogen), followed by anion exchange chromatography according to the manufacturer's protocols (HiTrap™ DEAE FF 17-5055-01, GE Healthcare).

### 2.5. CBM library construction

The gene encoding the CBM26 from *L. amylovorus*  $\alpha$ -amylase was initially amplified from the pQ-1mod vector [15] using the gene specific primers: 5'-CAGCAGCCATGGAAACAACAGAAAC-3' and 5'-TTAAGT TAGGATCCTACGTATTTTAC-3' carrying *NcoI* and *BamHI* restriction sites (underlined in oligos), to generate corresponding DNA fragments for

**Table 1**  
Oligonucleotide primers used for site-directed mutagenesis or library construction.

Name	Primers
pQ-Q68L	P1f: 5'-CCGATGGGACAAAGCTTACACCAGCAGCTAATG P2r: 5'-CATTAGCTGCTGGTGAAGCTTTGCCATCCGG
pQ-E74L	P1f: 5'-GCAAACACCAGCTGCTAATCTGGCTGTTTACC P2r: 5'-GGTAAAACAGCCAGATTAGCAGCTGGTGTITGC
pQ-F77L	P1f: 5'-GAGGCTGGTTAACTTTACTGCAGATGCTACTTATG P2r: 5'-CATAAGTAGCATCTGCAGTAAAGGTTAAACCAGCTC
1/Rv1	5'-TGCTACCCAGCTTGACGGCTTTTCGAAGTAAACCTTTTGTGTTCTGTTGTTCCATGGCTGCTG
2/Fw2a	5'-CAAGCTGGGGTAGCAGAGTGTATGCCNNKGTANNKAATAAGAACACGAACAAG <sup>a</sup>
3/Rv2a	5'-CTAAAGCGGTCAATTTTTTGGCAGCCAAGCTGAAGTTATAGCCTTGTTCGTGTC
4/Fw3	5'-AAATGACCGCTTTAGGTAACGACAAGTATGAATGGATCTCGACACTGATGAAG
5/Rv3a	5'-GTCCATCCGGTGAATATAACAGCTAAATCAGAGTCATCTTCATCAGTGTCTGA
6/Fw4b	5'-ATTCACCGATGGGACTAAGNNKACACCAGCAGCTAATNNKGCAGGTNNKACATTCACAGCAGA <sup>a</sup>
7/Rv4b	5'-TTAAGTTAGGATCCTACGTATTTTACGACACCAATTTGGTCATAAGTGGCATCTGCTGTAATGT

<sup>a</sup> Degenerate codon: K = T or G; N = A, C, G or T.

their cloning in the phagemid vector pG8SAET as fusions with major phage coat protein (pVIII) of M13 phage.

For library construction, mutations were introduced at five positions of the *LaCBM26* (Y18, Y20, Q68, E74 and F77). Mutations were introduced by amplification of seven gene fragments using the degenerate oligonucleotides shown in Table 1, where N is any of 20 natural amino acids and K is T or G. Oligos were assembled by an overlap-extension PCR (eight cycles, 95 °C for 1 min, 56 °C for 2 min, 68 °C for 2 min) using Platinum® DNA Polymerase (Thermo Scientific). An aliquot (0.5 µL) of the resulting PCR product from this reaction was amplified using the gene-specific primers (30 cycles, 98 °C for 30 s, 59 °C for 30 s, 72 °C for 30 s) using Phire Hot Start II DNA polymerase (F122S Thermo Scientific). This allowed generation of fragments for cloning into the phagemid vector, previously digested with *NcoI* and *BamHI*. The size of the library after transformation into electrocompetent *E. coli* TGI was calculated to be  $1.1 \times 10^6$  clones. To verify the random insertion of the mutations, the DNA insert of ten clones from the CBM26 library were sequenced.

## 2.6. Phage stocks

An exponentially growing culture of cells from the constructed library or a selected clone (2YT with 100 mg/L ampicillin at OD<sub>600</sub> ~0.7) was infected with helper phages M13 K07 (Invitrogen) with a multiplicity of infection of ~10 for 30 min at 37 °C without stirring and 30 min more at 250 rpm stirring. The medium was changed to 2YT, 100 µg/L ampicillin, 50 µg/L kanamycin and the cultures were grown at 30 °C overnight at 250 rpm of stirring. After centrifugation of cell culture (8000 ×g, 10 min), the supernatants (containing phage particles) were concentrated by polyethyleneglycol/NaCl precipitation. After centrifugation (14,810 ×g, 15 min at 4 °C), the precipitated phages were dissolved in Tris-buffered saline (TBS, 1×) and stored at 4 °C. As a negative control we used a non-related phage expressing HIV-1 Gag-derived epitope/peptide, and as a positive control we used WT *LaCBM26*. Both controls were displayed fused to pVIII protein on surface phage.

## 2.7. Biopanning

Phage selections were performed on soluble and insoluble starch (Sigma-Aldrich). Soluble starch was immobilized on costar 96 well polystyrene plates (3370) coated with 100 µL of 10% soluble starch as described in [21]. For soluble starch selection, 10<sup>11</sup> phage particles diluted in blocking buffer (phosphate-buffered saline containing 0.2% bovine serum albumin (PBS-BSA) and 0.05% (v/v) Tween-20; pre-incubated for 40 min) were added to starch-coated plates and incubated overnight at 4 °C. The plates were washed four times with PBS plus 0.05% (v/v) Tween-20 (PBS-5 T) and once with PBS. Bound phages were eluted using 200 µL of 0.1 M triethylamine (TEA) by incubating for 30 min at room temperature (RT) and neutralized by the addition of 200 µL of Tris-HCl 1 M (pH 7.4). Fifty microliters of the elution solution were saved for titration experiments, and the rest was used to infect exponentially growing *E. coli* TGI for preparation of new phage stocks. Three rounds of biopanning were performed.

For insoluble starch selections, 500 µL of phage stock containing 10<sup>11</sup> phage particles, displaying different variants of the CBM26, was added to 500 µL of a 5% starch suspension in PBS-BSA and 0.1% (v/v) Tween-20 (PBS-T). The mixture was incubated overnight using a roller rotator at 4 °C. After centrifugation (8330 ×g, 10 min), the substrate was subsequently washed four times with PBS-T and once with PBS. Bound phages were eluted using 300 µL of 9 mg/mL β-cyclodextrin by incubating for 40 min at 4 °C using a roller rotator, or 200 µL of 0.1 M triethylamine (TEA) by incubating for 30 min at RT and neutralized by the addition of 200 µL of Tris-HCl 1 M (pH 7.4). The elution solution was saved following the instructions described above. Two rounds of biopanning were performed, and the eluate from each round was

amplified and used for the subsequent round or for characterization of clones (first and second rounds).

## 2.8. Affinity gel electrophoresis (AGE)

The binding capacity of the mutant and wild-type proteins was determined by affinity gel electrophoresis. Purified proteins were separated in native gels containing either no carbohydrates or 1% of potato starch (Prolabo, Fontenay-sous-Bois, France). Electrophoresis was run at 4 °C at 20 mA (Bio-Rad Mini-Protean III system). Albumin was used as a negative control. Proteins were detected by staining with Coomassie Brilliant Blue R-250 [22].

## 2.9. Circular dichroism (CD)

CD spectroscopy was carried out at 20 °C on a J-710 spectropolarimeter (Jasco Inc., Easton, MD, USA) in a 1 mm cuvette with a protein concentration between 13 and 20 µM in 10 mM sodium phosphate buffer (pH 7.4). Spectra were collected three times from 260 to 190 nm at a scan speed of 10 nm/min, with a time constant of 10 s, a bandwidth of 1 nm, resolution of 1 nm and sensitivity of 20 mdeg. The raw data were converted to molar ellipticity ([θ]) using the formula:  $[\theta] = \theta / (l \times C \times N_r)$  where θ is ellipticity in millidegrees, *l* is the cell path length in millimeters, *C* is the molar concentration of protein and *N<sub>r</sub>* is the number of residues [23].

## 2.10. Isothermal titration calorimetry (ITC)

ITC was performed using an ITC<sub>200</sub> (Malvern) in 50 mM citrate-phosphate pH 5.5 at 25 °C using a protein concentration of 50–100 µM in the reaction cell and 1–10 mM β-cyclodextrin or maltoheptaose in the syringe. Protein samples were extensively dialyzed against the buffer. Ligand solutions were prepared in buffer saved from the final protein dialysis step. Both, protein and sugar solutions, were centrifuged (16,110 ×g) prior to use. Protein concentrations were determined by UV absorbance at 280 nm. The dilution heat of the ligand was obtained via addition of the ligand to the buffer solution under identical conditions to those of the protein sample. The binding constant (*K<sub>a</sub>*), and the enthalpy change (Δ*H*) were determined by nonlinear fitting of normalized titration data using a one-binding-site model [24] in the MicroCal Origin 7.0 software. Data were also analyzed using AFFINImeters software (website: <https://www.affinimeter.com/>). The standard Gibbs energy change (Δ*G*<sup>o</sup>) and the standard entropy change (Δ*S*<sup>o</sup>) were calculated from  $\Delta G^o = -RT \ln K_a$  and  $\Delta G^o = \Delta H^o - T \Delta S^o$ , where *R* (1.987207 cal/mol K) is the ideal gas constant and *T* (K) the temperature.

## 2.11. Thermal shift assay (thermofluor)

Protein unfolding was reviewed by thermal shift assay (thermofluor) in the CFX96 Touch Real Time PCR Detection System (Bio-Rad Laboratories, Inc.) with the hydrophobic fluorophore SYPRO orange as described in Ref. [21].

## 3. Results

### 3.1. Sequence analysis and homology modeling of *LaCBM26*

Structural prediction of the *LaCBM26* suggests that the domain adopts a β-sandwich fold, similar to most CBMs [6,7,9]. Most of the residues (>85%) fall in the allowed phi/psi region in the Ramachandran plot, making the model sufficiently reliable. The secondary structure prediction tool PDBsum (<http://www.ebi.ac.uk/pdbsum>) estimate that *LaCBM26* contains about 31.9% of β-sheets and 68.1% of other structures that include turns, and disordered structures. An amino acid sequence alignment of the *LaCBM26* domain model (PM0081297) and the three



other domains with solved structures CBM26 domain from *Bacillus halodurans* (*BhCBM26*, PDB: 2C3G), CBM26-1 and CBM26-2 domains from *Eubacterium rectale* (*ErCBM26-1* y *ErCBM26-2*, PDB: 6B15) was performed. The amino acid sequence of *LaCBM26* is 28% identical to *BhCBM26*; 26% identical to *ErCBM26-1* and 15% identical to *ErCBM26-2*. Fig. 1 shows four conserved residues related to starch binding, which, in the *LaCBM26* domain, are the amino acids W32, Y18, Y20, and Q68. These residues are conserved in the majority of the domains belonging to the family CBM26 [13]. Other residues participate in the ligand interaction such as G76 and E77 in the *BhCBM26* domain [10]; D250, N201, and E200 in the *ErCBM26-1* domain, or N355 in the *ErCBM26-2* domain [5], but these residues are not conserved among all CBM26 domains.

Structural superposition of the *LaCBM26* model (PM0081297), *BhCBM26* (PDB: 2C3G), *ErCBM26-1*, and *ErCBM26-2* (PDB: 6B15) was performed using Chimera software (<http://www.rbvi.ucsf.edu/chimera>). The *BhCBM26*, *ErCBM26-1* and *ErCBM26-2* overlap, with respect to *LaCBM26*, has a root mean square deviation of 1.3 Å, 2.5 Å and 2.1 Å, respectively, indicating structural similarity (DALI server, <http://ekhidna2.biocenter.helsinki.fi/dali/>) [25]. The structures show the location of the aromatic and polar residues in the binding cleft (Fig. 1). In addition, 3D structural analysis located the other residues that participate in starch binding in the vicinity of the binding site.

As expected, in the structural comparison, the greatest changes were observed in the loops. The separation between the *LaCBM26* Q68 Cα with respect to the corresponding Cα in Q71 (*BhCBM26*), Q247 (*ErCBM26-1*) and Q352 (*ErCBM26-2*) is 1.96 Å; 2.78 Å and 2.04 Å, respectively. Comparing the orientation of the side chain, there is a 4.43 Å, 7.36 Å, and 4.18 Å Cδ atom separation, respectively, indicating greater changes between these residues localized on the loop regions. There are only minor differences between W36/W208/W314 from *BhCBM26*/

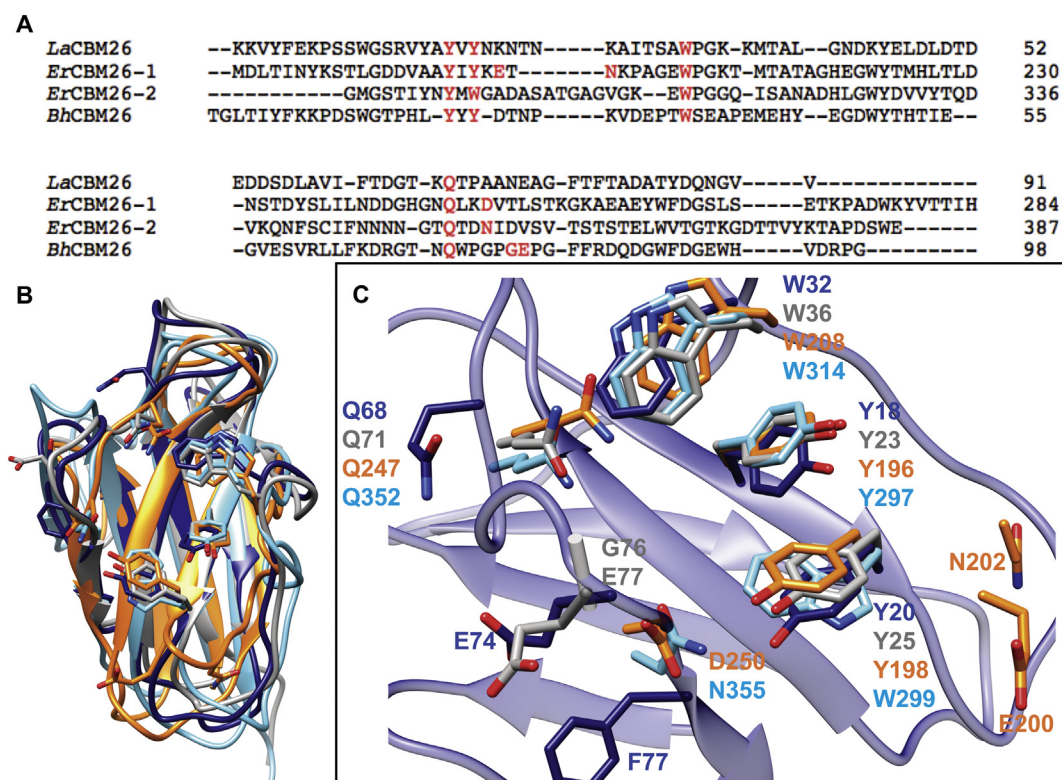
*ErCBM26-1/ErCBM26-2* and W32 from *LaCBM26*, since there is only a 1.51 Å, 2.87 Å, and 1.22 Å Cα atom separation, respectively, and 2.46 Å, 3.42 Å, and 1.88 Å in Nε1 atom separation, respectively. Conversely, two tyrosines overlap perfectly, and both are located in a β-strand. Cα atoms from Y23/Y196/Y297 from *BhCBM26/ErCBM26-1/ErCBM26-2* are separated by 0.47 Å, 0.78 Å, and 1.02 Å, respectively from the Cα of *LaCBM26*, with OH- separation of 1.21 Å, 3.18 Å, and 2.36 Å, respectively. Similarly, Cα atoms from Y25/Y198/W299 from *BhCBM26/ErCBM26-1/ErCBM26-2* show a separation of 0.29 Å, 1.0 Å, and 0.59 Å from the Cα atoms of Y20 from *LaCBM26* and 0.54 Å, 1.27 Å, and 0.35 Å Cγ atom separation, respectively.

To establish the importance of the above amino acids, we obtained several mutated proteins of the *LaCBM26* domain and constructed a combinatorial library, taking into account five of the binding residues (Y18, Y20, Q68, E74, and F77) but respecting W32 because its deletion avoids starch binding [11]. Phenylalanine 77 was selected for its aromatic character and proximity to the binding site [10].

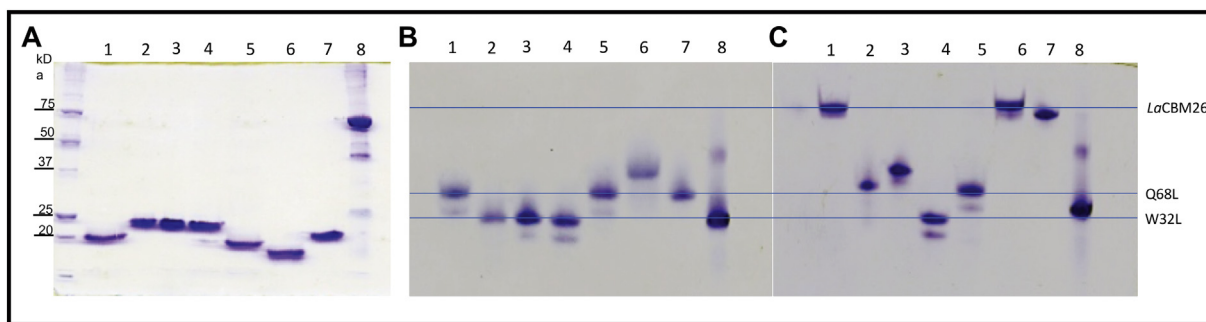
### 3.2. Point mutation impact in α-glucan-binding

The binding capability of six mutated proteins to soluble starch was examined by affinity electrophoresis. In previous work, we reported that the W32 residue contributes largely to the starch binding, since its mutation precludes attachment to soluble starch [11], while other mutations such as Y18L and Y20L weaken the interaction. Unexpectedly, the mutation of Q68L led to the complete loss of the binding capacity to starch, suggesting that both amino acids (Q68 and W32) were equally important for this interaction. Other mutations such as E74L and F77L did not affect the binding to soluble starch (Fig. 2).

The thermodynamic parameters for *LaCBM26* domain were determined using β-cyclodextrin and maltoheptaose as ligand (Fig. 3).



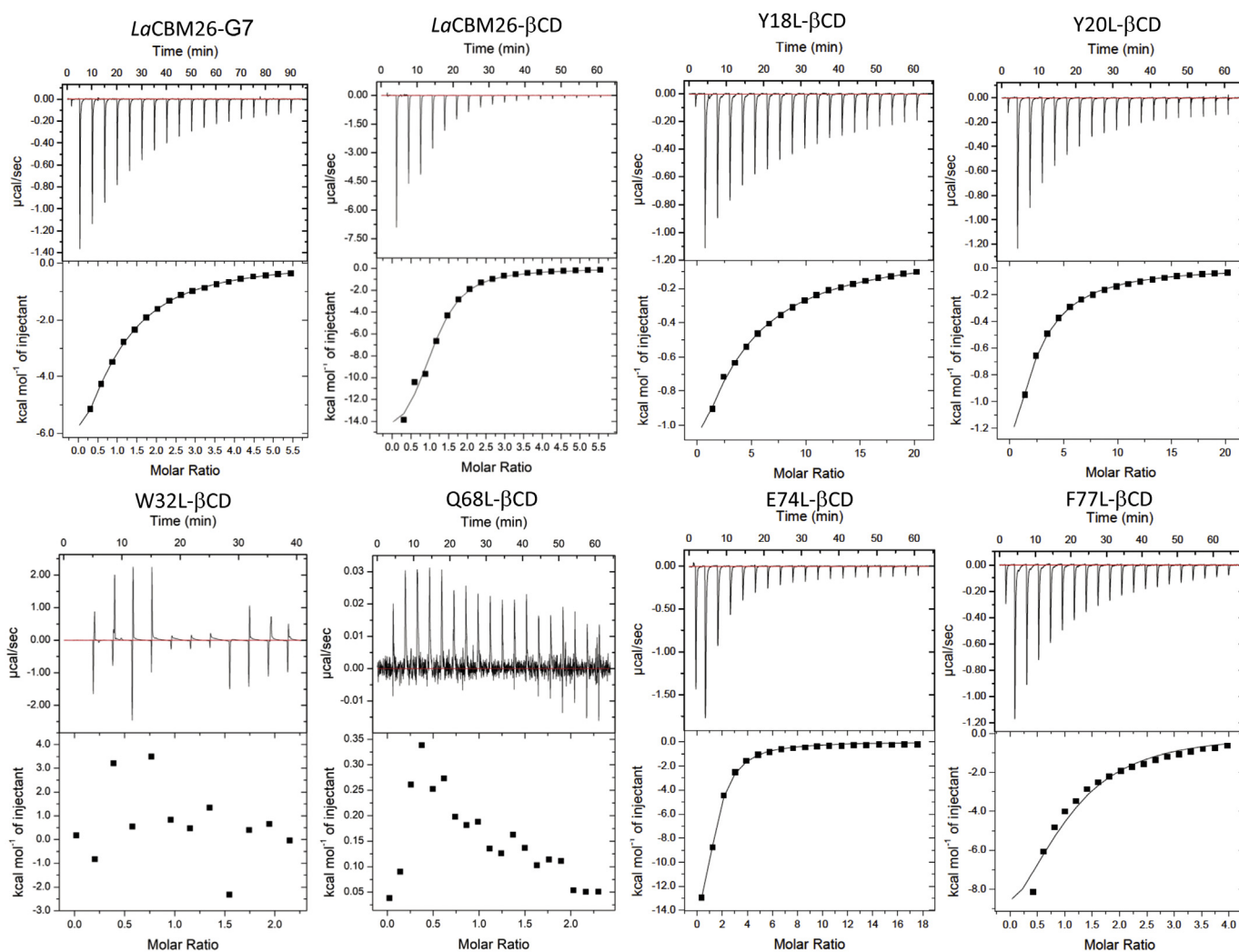
**Fig. 1.** Structural comparison between four carbohydrate binding domains of the family 26. A) Alignment of the primary structure obtained by Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>); in red are the conserved residues between these domains that participate in the ligand interaction according to the structure studies or by site-directed mutagenesis. B) Structural superimposition of tertiary structure of the CBM26 domain from *Bacillus halodurans* maltohexaose-forming amylase (PDB: 2c3g; gray), CBM26-1 domain (orange) and CBM26-2 domain (light blue) (PDB: 6B15) from α-amylase of *Eubacterium rectale*, and CBM26 domain model from α-amylase of *Lactobacillus amylovorus* (PM0081297; blue). The amino acids that are bound to the ligand are displayed as sticks. C) Close-up view of the binding site of the four CBM26 domains, showing binding residues. The backbone β-sandwich is *LaCBM26* displayed as a blue cartoon. Chimera software (<https://www.cgl.ucsf.edu/chimera/>) was used for structure superimposition.



**Fig. 2.** Affinity Gel Electrophoresis of *LaCBM26* and different mutated proteins. (A) SDS-PAGE showing purified *LaCBM26* mutated proteins (B) separated by electrophoresis in native conditions of the *LaCBM26* variants in the absence and (C) presence of soluble starch. 1: *LaCBM26*, 2: Y18L, 3: Y20L, 4: W32 L, 5: Q68L, 6: E74L, 7: F77L, 8: BSA (negative control). Horizontal lines in gel indicate the slowing of migration, in presence of starch, for WT *LaCBM26*, Q68L and W32 L proteins.

Similar to others CBMs (CBM26-1 and CBM26-2 of *E. rectale* [5]; CBM25 and CBM26 of *B. halodurans* [10]; CBM41 the *Thermotoga maritima* [26]; CBM45 of *Arabidopsis thaliana* [27]), *LaCBM26* shows a stoichiometry of 1:1 and a binding constant ( $K_d$ ) of an order of magnitude of  $10^4 \text{ M}^{-1}$  for both ligands (Table 2). However, the affinity was slightly better for  $\beta$ -cyclodextrin, probably due to the additional entropic penalty associated with the stabilization of the conformation of the linear ligand upon binding [27].

Stoichiometry and binding constants for mutated proteins E74L and F77L are on the same order of magnitude as those observed with the WT domain, which corroborates the result observed in AGE (Fig. 2). On the other hand, Y18L and Y20L mutated proteins have binding constants 50 and 20 times lower than that of the wild-type domain. The inability of W32L and Q68L to bind the carbohydrate was also corroborated, as the titration of these proteins in ITC did not show any signal (Fig. 3, Table 2).



**Fig. 3.** Isothermal Titration Calorimetry profiles of the *LaCBM26* and mutated proteins. ITC was performed in 50 mM citrate-phosphate pH 5.5 at 25 °C using a protein concentration of 50–100  $\mu\text{M}$  in the reaction cell and 1–10 mM  $\beta$ -cyclodextrin or maltoheptaose in the syringe.

**Table 2**

Parameters of *LaCBM26* and the mutated proteins binding to  $\alpha$ -glucans as determined by isothermal titration calorimetry.

	$K_a \times 10^{-4} \text{ M}^{-1}$	$\Delta G$ kcal/mol	$\Delta H$ kcal/mol	$T\Delta S$ kcal/mol
Maltoheptaose				
CBM26	$2.41 \pm 0.4$	-5.98	$-11.4 \pm 1.0$	-5.46
$\beta$ -Cyclodextrin				
CBM26	$5.69 \pm 0.9$	-6.49	$-16.8 \pm 0.9$	-10.31
Y18L	$0.11 \pm 0.01$	-4.15	$-7.8 \pm 0.01$	-3.65
Y20L	$0.25 \pm 0.04$	-4.64	$-8.01 \pm 0.03$	-3.37
W32L	NB			
E74L	$2.71 \pm 0.3$	-6.04	$-11.32 \pm 0.16$	-5.27
F77L	$3.66 \pm 0.3$	-6.22	$-14.4 \pm 1.1$	-8.17
Q68L	NB			

NB: not binding. Data analyzed with AFFINImeter software.

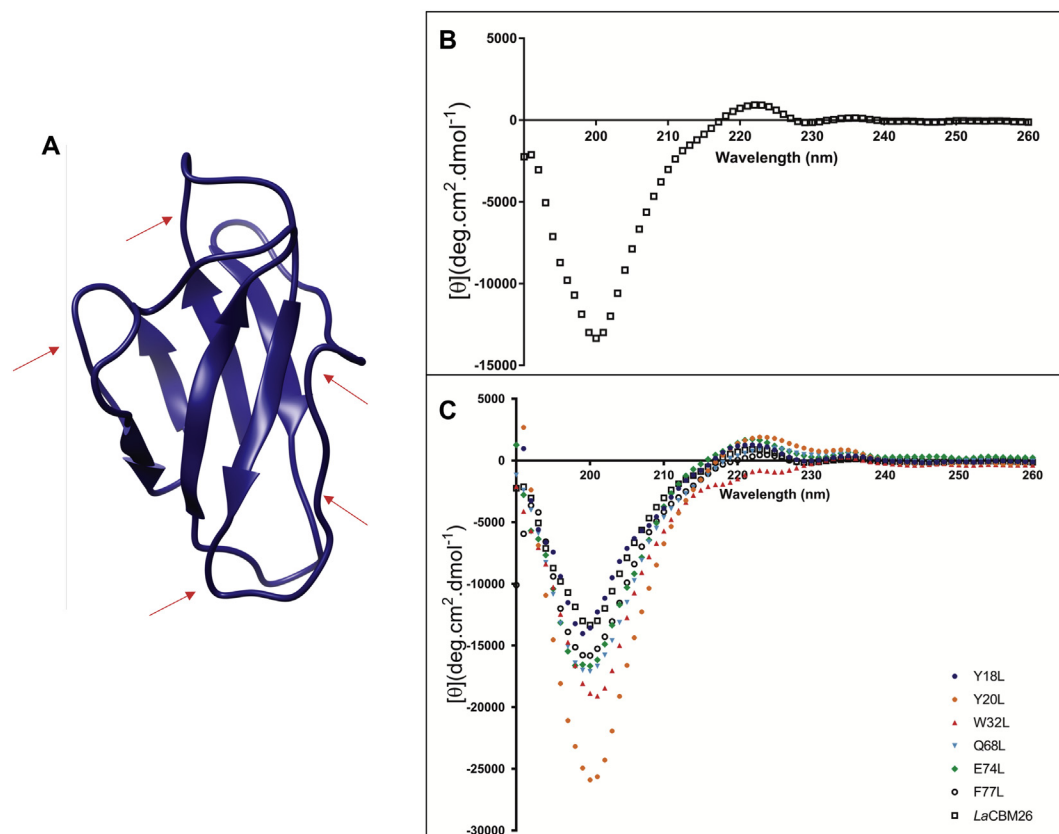
The thermodynamic parameters showed that the interaction of the domain *LaCBM26* and the mutated domains with the substrate was favored for a negative change in the enthalpy, ( $\Delta H$ ) and partially compensate unfavorable entropy ( $T\Delta S$ ), a common characteristic of the interaction of the CBMs with carbohydrates [6,10,28]. Particularly, changes in enthalpy values were observed when tyrosines (18 and 20) were mutated to leucine, since mutant proteins showed a decrease in affinity accompanied by a penalization in the enthalpy and an increase in the entropy of the system. In the case of the E74L and F77L, the obtained data were similar to data of the wild type domain (Table 2).

Since the far-UV CD provides information about the backbone conformation of a protein [23], the effect of point mutations on the secondary structure of the proteins was reviewed. Fig. 4 shows the obtained circular dichroism spectra in the far UV range (190–260 nm) for the

different mutated proteins of the *LaCBM26* domain. It is clear that *LaCBM26* and the modified proteins share same structural characteristics in relation to the folding of their polypeptide chains (Fig. 4). In particular, negative bands located approximately 198–202 nm suggest atypical spectral due to few elements of secondary structure and  $\beta$ -turns [23,29]. Circular Dichroism analysis for secondary structure estimation by BeStSel server (<http://bestsel.elte.hu/index.php>) confirmed that the *LaCBM26* structure has few secondary structures elements comprising 38.7%  $\beta$ -sheets and 61.3% of others structures that includes turns and disordered regions which are related with the data obtained from 3D-structure analysis by PDBsum (Table 3) [30]. Besides, two positive bands are observed, the first approximately 218–227 nm and an intensity of the  $921 \text{ deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$ , the second with a lower intensity,  $150 \text{ deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$ , approximately 231–239 nm. The positive ellipticity observed at 223 nm has been proposed as result of the presence of aromatic patches and may change due to local conformations around tryptophan and tyrosine in the structure [31,32]. In particular, the mutation W32L completely eliminated the positive peak approximately 218–227 nm, corroborating the aforementioned. On the other hand, the spectra obtained from Y18L, Y20L, Q68L, E74L, and F77L mutated proteins, showed no evident change in secondary structure (Fig. 4), which was confirmed through CD analysis since no changes were observed in secondary structure estimation (Table 3).

### 3.3. Simultaneous random mutagenesis on *LaCBM26* binding residues

To assess the impact of random mutagenesis of the *LaCBM26* residues implicated in binding, we constructed a combinatorial library by modifying the conserved amino acids related to starch binding in *LaCBM26* domain. Residues Y18, Y20, Q68, E74, and F77 were changed randomly for any of the twenty amino acids through the use of degenerate primers (see Materials and methods), except for W32 that was



**Fig. 4.** Far-UV CD spectra of *LaCBM26* and generated mutated proteins. A) *LaCBM26* model in blue cartoon showing the number of coils in the structure (red arrows). B) Spectra for WT *LaCBM26* and C) *LaCBM26* mutated proteins in 10 mM sodium phosphate pH 7.4, at 20 °C.



**Table 3**

Estimation of the Secondary structure of *LaCBM26* and mutated proteins by the BestSel server (<http://bestsel.elte.hu/index.php>).

Protein	Antiparallel				Turn	Others
	Sum	Left-twisted	relaxed	Right-twisted		
<i>LaCBM26</i>	38.7	2.6	18.1	18.0	13.3	48.0
W32L	41.0	2.1	18.0	21.0	13.5	45.5
Q68L	42.5	2.4	19.5	20.6	13.7	43.8
Y18L	44.6	2.7	20.8	21.1	14.0	41.4
Y20L	42.8	2.1	19.4	21.4	13.8	43.4
E74L	42.6	2.1	19.3	21.1	13.9	43.5
F77L	41.0	2.5	18.7	19.7	13.7	45.3

Sum: total content of antiparallel  $\beta$ -strands, divided as left-twisted, relaxed or right-twisted.

conserved because of its importance in the hydrophobic platform and its relevance in the secondary structure (Fig. 4). The size of the resulting library after the transformations in *E. coli* was  $1.1 \times 10^6$  variants; 15 out of 20 colonies selected at random showed the insert in the correct reading frame. The sequence of these proteins showed that 17 of the 20 amino acids were inserted in each position, which illustrates the diversity of the library (Fig. 5).

Selection of mutated proteins from the library was carried out in the starch immobilized in polystyrene plates or directly on insoluble starch. In the case of immobilized starch, its availability to specifically interact with *LaCBM26* and *LaSBD* was reviewed, using the inactive mutant W32L as negative control (see Fig. 1 in Ref. [21]).

Phage-displayed library was subjected to several rounds of selection in soluble or insoluble starch. After two or three selections rounds,

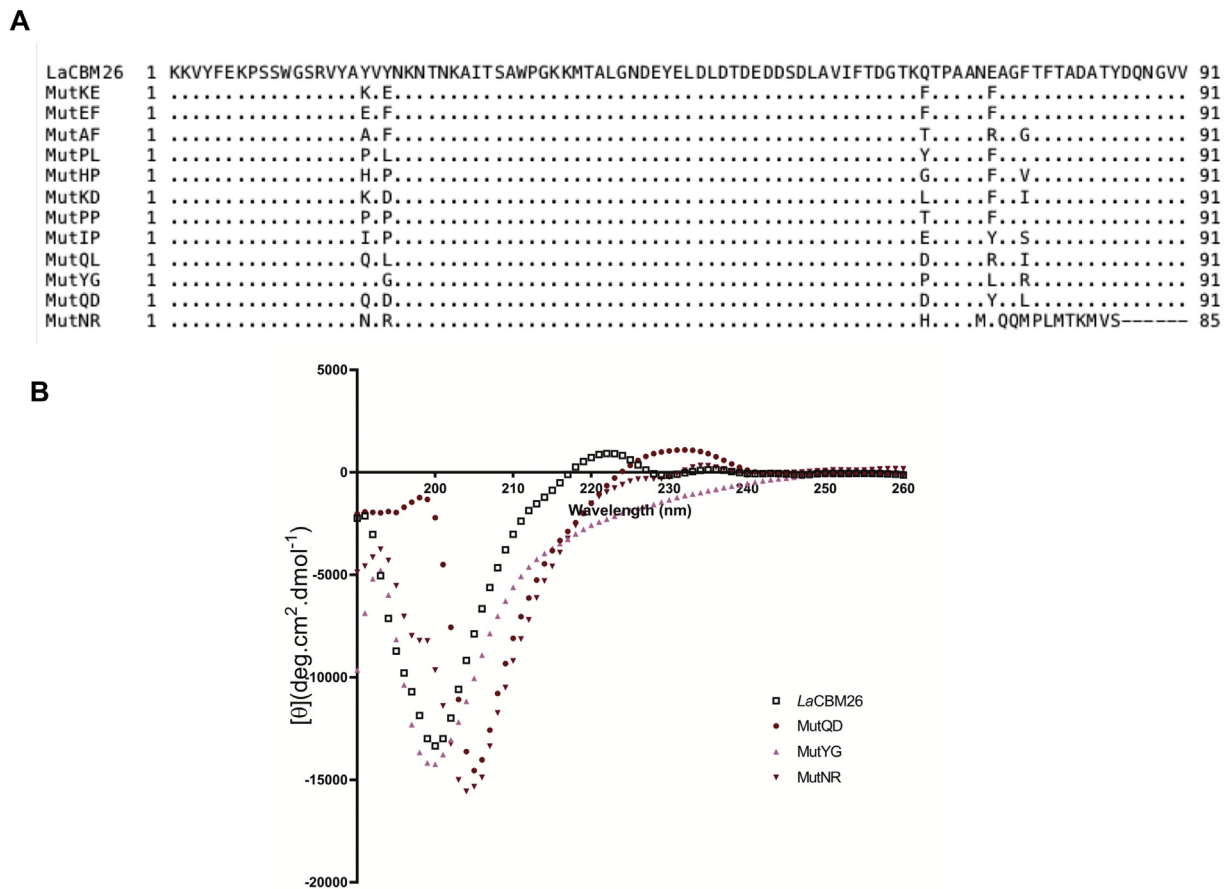
thirteen variants were selected, expressed and purified by nickel affinity chromatography to analyze their substrate binding capacity and secondary structure.

Interestingly, none of the analyzed proteins showed affinity for starch or maltooligosaccharides. Fig. 5 shows the differences found in the primary sequence as well as the far-UV CD spectra of three different mutated proteins, MutYG, MutQD and MutNR. The inserted mutations in these variants caused changes at the secondary structure level and affected their function since they were incapable to bind starch (see Fig. 2 in Ref. [21]). Therefore, the obtained results with the generated *LaCBM26* library suggested little tolerance of the domain to change.

#### 4. Discussion

The CBMs are auxiliary domains with a wide variety of biological functions. Many crystallographic structures of these domains have been solved, observing that the preferred structure is that of a  $\beta$ -sandwich and demonstrating, in all of them, the importance of aromatic amino acids as mediators of CBM-carbohydrate interactions [12–14].

The tertiary structure model of the *LaCBM26* domain predicts a  $\beta$ -sandwich fold (Fig. 1), where the amino acids related to binding are in the concave face of the structure model as with most CBMs [6]. Thus, we expected to obtain a CD spectrum associated with the type of secondary structure related to presence of antiparallel  $\beta$ -sheets [23]. However, the data from far-UV CD showed very little ordered secondary structure which suggest the presence of a significant number of coils, as it has been previously observed for other CBMs such as the starch-binding domain-1 (SBD1) from *Kocuria varians*  $\alpha$ -amylase [33] and hevein [34]. Likewise, we examined whether the presence of the



**Fig. 5.** Amino acid sequence alignment of *LaCBM26* and selected mutated proteins. A) Primary structure alignment of clones obtained by random mutagenesis on *LaCBM26*, only the amino acids that were modified are indicated in the alignment. B) Far-UV CD spectra of wild-type *LaCBM26* and mutated proteins MutYG, MutQD and MutNR. Spectra were obtained in 10 mM sodium phosphate pH 7.4, at 20 °C.



substrate could cause changes in domain folding. However, no changes were observed in the presence of  $\beta$ -cyclodextrin at the secondary structure level beyond a slight change between 218 and 230 nm wavelength (see Fig. 4 in Ref. [21]).

In the same way, we check *LaCBM26* thermal stability through a fluorescence-based thermal shift (ThermoFluor) assay, which allows distinction between folded and unfolded states of a protein [35]. The fluorophore binds to exposed hydrophobic areas in an unfolded protein, leading to fluorescence. Fluorescence curves for *LaCBM26* showed high fluorescence even at low temperature, which suggests that the protein has the hydrophobic areas exposed where the fluorophore may interact. Consequently, the thermofluor experiments corroborate that *LaCBM26* has few secondary structure elements, although the *LaSBD* (tandem of five CBM26, [3]) developed an optimal thermal denaturation curve. The addition of the substrate, in this case maltoheptaose, did not modify the lack of signal (see Fig. 3 in Ref. [21]).

Despite the suggestive low level of structure, the *LaCBM26* domain is functional since it is able to bind soluble and insoluble starch. According to the thermodynamic data presented here, the *LaCBM26* domain has an affinity on the order of  $\sim 10^4 \text{ M}^{-1}$  (Table 2) for substrates analogous to starch such as  $\beta$ -cyclodextrin and maltoheptaose; this domain presents a stable interaction ( $\Delta G = -6.49 \text{ kcal/mol}$ ), enthalpically favored but entropically disadvantaged, which is characteristic of type B-CBMs [36,37].

Highest affinity was observed with  $\beta$ -cyclodextrin, a ligand that mimics a single turn of the helix in amylose, while the maltoheptaose in aqueous solution are not in a double-helical conformation, but it has been suggested they form transient single helices [38].

Structure-function relationships of conserved amino acids in the binding site were reviewed by site-directed mutagenesis. As previously established, in the CBM-carbohydrate interaction, aromatic amino acids are indispensable and govern such interaction [6,7,10,39–42]. For example, in the CBM21 domain of the *Rhizopus oryzae* glucoamylase, two aromatic amino acids, W47 and Y83, form the binding platform [41], in the CBM41 from the *Thermotoga maritima* pullulanase Pula, three tryptophan (W27, W29, and W73) are necessary for binding [42]. In the *BhCBM26*, W36 and Y25 are responsible for stacking interaction against pyranose rings of maltose sugar residues. Y23 contributes to a hydrogen bond formation with oxygen 6 of the sugar, stacking against W36 [10]. However, in the interaction of the *LaCBM26* domain to the starch, the hydrogen bond formed by Q68 is equally important than the hydrophobic interaction of W32, since the elimination of any of them completely abolishes the binding (Fig. 2). In contrast, mutations in Y18L and Y20L, residues which theoretically form a hydrogen bond and hydrophobic interaction, respectively, still allowed starch binding. In this case, it is possible that the loss of hydrogen bonding leads to redistribution of interactions that compensate for such linkages although thermodynamic parameters suggested that the performed mutations decreased the interaction stability [43]. However, it is difficult to accurately assess the effect of mutations without having three-dimensional structural information. Mutagenesis of other amino acids such as E74L or F77L does not affect the binding or the thermodynamics parameters (Table 2). Taken together, all these observations suggest that the movement proposed for *BhCBM26* in the loop P73–G79 (P70–G78 in *LaCBM26*), which allows the formation of the hydrogen bond between glutamic acid and glucose [10], may not occur in *LaCBM26*.

Far-UV CD spectra of the analyzed proteins showed that the secondary structure of the *LaCBM26* domain was not affected although it may exhibit some local conformational changes. A major effect was observed in the spectrum obtained with the W32 L mutation. The maximum between 218 and 227 nm possibly disappeared because the contribution given by the aromatic ring of the amino acid no longer existed (Fig. 4) [31]. Observations were previously reported for other proteins [33,44,45], such as glucan-binding protein A (GBP-A) of *Streptococcus mutants* [46] and starch-binding domain of glucoamylase from *Aspergillus niger* [32].

Molecular engineering of the *LaCBM26* affected the secondary structure of selected proteins.

Few CBMs have been mutated and used as evolution platforms for designing domains with novel features of specificity and affinity for different substrates [1]. However, given the various biotechnological applications in which different CBMs [47] have been used, a library of the *LaCBM26* domain was generated to ascertain which modifications in the primary structure could change the affinity towards starch and to obtain successor domains with improved characteristics.

Simultaneous mutations of residues Y18, Y20, Q68, E74 and F77 resulted in the loss of function and a significant change in the secondary structure (Fig. 5). None of the obtained combinations at random was well tolerated and probably much of the diversity introduced into *LaCBM26* during library design was not compatible with the appropriate folding of the proteins.

These observations suggest that the *LaCBM26* domain has not robustness or structural stability to introduce mutations in the different residues selected. However, this may not be general for other CBMs. For example, the CBM4-2 domain of xylanase 10A from the thermophilic bacterium *Rhodothermus marinus* has been proposed as a protein scaffold [48], from which mutated proteins with different specificity towards carbohydrates such as xylan [49], xyloglucan [50], mannan [48], and even specificity for other proteins [51], have been obtained; although this is the only highly successful case of evolution in carbohydrate binding modules so far reported.

Thus, it is possible that structural differences could explain that the *LaCBM26* domain presented greater resistance to change than the *RmCBM4-2* domain. As mentioned earlier, the circular dichroism results suggested that the *LaCBM26* domain presents an important proportion of turns and disordered areas that could reduce its structural stability decreasing the probability of finding functional mutated proteins (Figs. 4, 5). In contrast, the CBM4-2 domain presents a spectrum corresponding to a  $\beta$ -sandwich fold [48] with 50%  $\beta$ -sheets and 5% helices, forming an eleven-stranded  $\beta$ -sheet (PDB:1K45) [52] that could make the core rigid, enabling tolerance to amino acid changes without loss of structural integrity.

It is clear that the structural characteristics that promote the evolution of proteins are only partially understood. Thus, certain protein architectures are less likely to compensate for the impact of mutation of important residues related to function and possibly less susceptible to functional innovation. Recently, Cohen-Khait et al. [53] observed that three mutations are sufficient to generate a new binding site on a protein. However, they also shown that such mutations severely destabilize the protein, suggesting that the limited stability of natural proteins protects them from evolving towards a promiscuous and non-productive binding and that proteins in nature are stable enough for their own needs.

## 5. Conclusions

The characterization of mutated proteins of the *LaCBM26* domain disclosed the role of each conserved amino acid of the binding site; interestingly, polar amino acids have a more critical role than has generally been recognized. Specifically, the glutamine Q68 in the *LaCBM26* domain was essential for the binding process. In our analysis, the established regions A, B and C would correspond to residues W32, Y20 and Q68, respectively.

On the other hand, the simultaneous mutation of five of the six residues reported as participants in binding was enough to cause the destabilization of the structure and consequently the loss of function, suggesting that, in a domain where random coiling prevails, the presence of each amino acid is important for the overall function.

## Acknowledgments

This work was part of the doctoral thesis of Armenta S. within the Doctoral Program in Biochemical Sciences, UNAM. Armenta S. and Sánchez-Cuapio Z. were supported by personal grants from Consejo

Nacional de Ciencia y Tecnología Conacyt, México. This work was supported by DGAPA, UNAM grant IN209316. ITC and DC experiments were performed at the LabDP-UAEM and at the Chemistry Institute, UNAM. Authors thank to Patricia de la Torre for the sequencing of mutants (LSGBS, UNAM), to Dr. Daniel Guillén for technical support, and to Dr. Sergio Romero-Romero for the validations of the LaCBM26 model and for critical reading of the manuscript.

## References

- [1] S. Armenta, S. Moreno-Mendieta, Z. Sánchez-Cuapio, S. Sánchez, R. Rodríguez-Sanoja, Advances in molecular engineering of carbohydrate-binding modules, *Proteins: Struct., Funct., Bioinf.* 85 (2017) 1602–1617, <https://doi.org/10.1002/prot.25327>.
- [2] D. Cockburn, M.M. Nielsen, C. Christiansen, J.M. Andersen, J.B. Rannes, A. Blennow, B. Svensson, Surface binding sites in amylase have distinct roles in recognition of starch structure motifs and degradation, *Int. J. Biol. Macromol.* 75 (2015) 338–345, <https://doi.org/10.1016/j.ijbiomac.2015.01.054>.
- [3] R. Rodríguez-Sanoja, J. Morlon-Guyot, J. Jore, J. Pintado, N. Juge, J.P. Guyot, Comparative characterization of complete and truncated forms of *Lactobacillus amylovorus*  $\alpha$ -amylase and role of the c-terminal direct repeats in raw-starch binding, *Appl. Environ. Microbiol.* 66 (2000) 3350–3356, <https://doi.org/10.1128/AEM.66.8.3350-3356.2000>.
- [4] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, *Nucleic Acids Res.* 42 (2014) 490–495, <https://doi.org/10.1093/nar/gkt1178>.
- [5] D.W. Cockburn, C. Suh, K.P. Medina, R.M. Duvall, Z. Wawrzak, B. Henrissat, N.M. Koropatkin, Novel carbohydrate binding modules in the surface anchored  $\alpha$ -amylase of *Eubacterium rectale* provide a molecular rationale for the range of starches used by this organism in the human gut, *Mol. Microbiol.* 107 (2018) 249–264, <https://doi.org/10.1111/mmi.13881>.
- [6] A.B. Boraston, D.N. Bolam, H.J. Gilbert, G.J. Davies, Carbohydrate-binding modules: fine-tuning polysaccharide recognition, *Biochem. J.* 382 (2004) 769–781, <https://doi.org/10.1042/BJ20040892>.
- [7] H. Hashimoto, Recent structural studies of carbohydrate-binding modules, *Cell. Mol. Life Sci.* 63 (2006) 2954–2967, <https://doi.org/10.1007/s00018-006-6195-3>.
- [8] H.J. Gilbert, J.P. Knox, A.B. Boraston, Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules, *Curr. Opin. Struct. Biol.* 23 (2013) 669–677, <https://doi.org/10.1016/j.sbi.2013.05.005>.
- [9] D.W. Abbott, A.L. van Bueren, Using structure to inform carbohydrate binding module function, *Curr. Opin. Struct. Biol.* 28 (2014) 32–40, <https://doi.org/10.1016/j.sbi.2014.07.004>.
- [10] A.B. Boraston, M. Healey, J. Klassen, E. Ficko-Blean, A.L. Van Bueren, V. Law, A structural and functional analysis of  $\alpha$ -glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition, *J. Biol. Chem.* 281 (2006) 587–598, <https://doi.org/10.1074/jbc.M509958200>.
- [11] R. Rodríguez-Sanoja, N. Oviedo, L. Escalante, B. Ruiz, S. Sánchez, A single residue mutation abolishes attachment of the CBM26 starch-binding domain from *Lactobacillus amylovorus*  $\alpha$ -amylase, *J. Ind. Microbiol. Biotechnol.* 36 (2009) 341–346, <https://doi.org/10.1007/s10295-008-0502-y>.
- [12] C. Christiansen, M. Abou Hachem, Š. Janeček, A. Viksø-Nielsen, A. Blennow, B. Svensson, The carbohydrate-binding module family 20 – diversity, structure, and function, *FEBS J.* 276 (2009) 5006–5029, <https://doi.org/10.1111/j.1742-4658.2009.07221.x>.
- [13] K. Majzlová, Š. Janeček, Two structurally related starch-binding domain families CBM25 and CBM26, *Biologia* 69 (2014) 1087–1096, <https://doi.org/10.2478/s11756-014-0415-3>.
- [14] Š. Janeček, K. Majzlová, B. Svensson, E.A. MacGregor, The starch-binding domain family CBM41—an *in silico* analysis of evolutionary relationships, *Proteins: Struct., Funct., Bioinf.* 85 (2017) 1480–1492, <https://doi.org/10.1002/prot.25309>.
- [15] D. Guillén, M. Santiago, L. Linares, R. Pérez, J. Morlon, B. Ruiz, S. Sánchez, R. Rodríguez-Sanoja, Alpha-amylase starch binding domains: cooperative effects of binding to starch granules of multiple tandemly arranged domains, *Appl. Environ. Microbiol.* 73 (2007) 3833–3837, <https://doi.org/10.1128/AEM.02628-06>.
- [16] J. Yang, R. Yan, A. Roy, D. Xu, P. J. Y. Zhang, The I-TASSER suite: protein structure and function prediction, *Nat. Methods* 12 (2015) 7–8, <https://doi.org/10.1038/nmeth.3213>.
- [17] Y. Zhang, I-TASSER server for protein 3D structure prediction, *BMC Bioinf.* 9 (2008) 40, <https://doi.org/10.1186/1471-2105-9-40>.
- [18] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291, <https://doi.org/10.1107/S0021889892009944>.
- [19] R. Hoof, G. Vriend, C. Sander, E. Abola, Errors in protein structure, *Nature* 381 (1996).
- [20] T. Castrignanò, P.D. De Meo, D. Cozzetto, I.G. Talamo, A. Tramontano, The PMDB protein model database, *Nucleic Acids Res.* 34 (2006) D306–D309, <https://doi.org/10.1093/nar/gkj105>.
- [21] S. Armenta, Z. Sánchez-Cuapio, M.E. Munguía, N. Pulido, A. Farrés, K. Manoutcharian, A. Hernández-Santoyo, S. Moreno-Mendieta, S. Sánchez, R. Rodríguez-Sanoja, Data of secondary structure and  $\alpha$ -glucan-binding capacity of the LaCBM26, *Int. J. Biol. Macromol.* (2018) (Data in Brief Submitted for publication).
- [22] P. Tomme, A. Boraston, J.M. Kormos, R.A.J. Warren, D.G. Kilburn, Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules, *Enzym. Microb. Technol.* 27 (2000) 453–458, [https://doi.org/10.1016/S0141-0229\(00\)00246-5](https://doi.org/10.1016/S0141-0229(00)00246-5).
- [23] S.M. Kelly, T.J. Jess, N.C. Price, How to study proteins by circular dichroism, *Biochim. Biophys. Acta, Proteins Proteomics* 1751 (2005) 119–139, <https://doi.org/10.1016/j.bbapap.2005.06.005>.
- [24] T. Wiseman, S. Williston, J.F. Brandts, L.N. Lin, Rapid measurement of binding constants and heats of binding using a new titration calorimeter, *Anal. Biochem.* 179 (1989) 131–137, [https://doi.org/10.1016/0003-2697\(89\)90213-3](https://doi.org/10.1016/0003-2697(89)90213-3).
- [25] L. Holm, L.M. Laakso, Dali server update, *Nucleic Acids Res.* 44 (2016) W351–W355, <https://doi.org/10.1093/nar/gkw357>.
- [26] A. Lammerts Van Bueren, R. Finn, J. Ausiò, A.B. Boraston,  $\alpha$ -Glucan recognition by a new family of carbohydrate-binding modules found primarily in bacterial pathogens, *Biochemistry* 43 (2004) 15633–15642, <https://doi.org/10.1021/bi048215z>.
- [27] M.A. Glaring, M.J. Baumann, M.A. Hachem, H. Nakai, N. Nakai, D. Santelia, B.W. Sigurskjold, S.C. Zeeman, A. Blennow, B. Svensson, Starch-binding domains in the CBM45 family – low-affinity domains from glucan, water dikinase and  $\alpha$ -amylase involved in plastidial starch metabolism, *FEBS J.* 278 (2011) 1175–1185, <https://doi.org/10.1111/j.1742-4658.2011.08043.x>.
- [28] A.L. Van Bueren, M. Higgins, D. Wang, R.D. Burke, A.B. Boraston, Identification and structural basis of binding to host lung glycogen by streptococcal virulence factors, *Nat. Struct. Mol. Biol.* 14 (2007) 76–84, <https://doi.org/10.1038/nsmb1187>.
- [29] S.M.M.M. Kelly, N.C.C.C. Price, The use of circular dichroism in the investigation of protein structure and function, *Curr. Protein Pept. Sci.* 1 (2000) 349–384, <https://doi.org/10.2174/1389203003381315>.
- [30] A. Miconnai, F. Wien, É. Bulyáki, J. Kun, É. Mousong, Y.H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra, *Nucleic Acids Res.* 46 (2018) W315–W322, <https://doi.org/10.1093/nar/gky497>.
- [31] R.W. Woody, Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins, *Eur. Biophys. J.* 23 (1994) 253, <https://doi.org/10.1007/BF00213575>.
- [32] C. Ota, M. Ikeguchi, A. Tanaka, D. Hamada, Residual structures in the unfolded state of starch-binding domain of glucoamylase revealed by near-UV circular dichroism and protein engineering techniques, *Biochim. Biophys. Acta, Proteins Proteomics* 1864 (2016) 1464–1472, <https://doi.org/10.1016/j.bbapap.2016.05.002>.
- [33] R. Yamaguchi, T. Arakawa, H. Tokunaga, M. Ishibashi, M. Tokunaga, Distinct characteristics of single starch-binding domain SBD1 derived from tandem domains SBD1-SBD2 of halophilic *Kocuria varians* alpha-amylase, *Protein J.* 31 (2012) 250–258, <https://doi.org/10.1007/s10930-012-9400-2>.
- [34] A. Rodríguez-Romero, B. Arreguin, A. Hernández-Arana, Unusual far-ultraviolet circular dichroism of wheat germ agglutinin and hevein originated from cystine residues, *Biochim. Biophys. Acta* 998 (1989) 21–24.
- [35] L. Reinhard, H. Mayerhofer, A. Geerluf, J. Mueller-Dieckmann, M.S. Weiss, Optimization of protein buffer cocktails using thermofluor, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 69 (2013) 209–214, <https://doi.org/10.1107/S1744309112051858>.
- [36] A.S. Luís, I. Venditto, M.J. Temple, A. Rogowski, A. Baslé, J. Xue, J.P. Knox, J.A.M. Prates, L.M.A. Ferreira, C.M.G.A. Fontes, S. Najmudin, H.J. Gilbert, Understanding how noncatalytic carbohydrate binding modules can display specificity for xyloglucan, *J. Biol. Chem.* 288 (2013) 4799–4809, <https://doi.org/10.1074/jbc.M112.432781>.
- [37] A. Viegas, J. Sardinha, F. Freire, D.F. Duarte, A.L. Carvalho, C.M.G.A. Fontes, M.J. Romão, A.L. Macedo, E.J. Cabrita, Solution structure, dynamics and binding studies of a family 11 carbohydrate-binding module from *Clostridium thermocellum* (Ct CBM11), *Biochem. J.* 451 (2013) 289–300, <https://doi.org/10.1042/BJ20120627>.
- [38] G. Williamson, N.J. Belshaw, Specificity of the binding domain of glucoamylase 1, *Eur. J. Biochem.* 211 (1993) 717–724.
- [39] M.P. Williamson, M.F. Le Gal-Coëffet, K. Sorimachi, C.S.M. Furniss, D.B. Archer, G. Williamson, Function of conserved tryptophans in the *Aspergillus niger* glucoamylase 1 starch binding domain, *Biochemistry* 36 (1997) 7535–7539, <https://doi.org/10.1021/bi9702896>.
- [40] A.L. Van Bueren, A.B. Boraston, The Structural Basis of  $\alpha$ -Glucan Recognition by a Family 41 Carbohydrate-binding Module from *Thermotoga maritima*, 2007 555–560, <https://doi.org/10.1016/j.jmb.2006.10.018>.
- [41] Y. Liu, Y. Lai, W. Chou, M.D. Chang, P. Lyu, Solution structure of family 21 carbohydrate-binding module from *Rhizopus oryzae* glucoamylase, *Biochem. J.* 403 (2007) 21–30, <https://doi.org/10.1042/BJ20061312>.
- [42] A.L. van Bueren, A.B. Boraston, The structural basis of  $\alpha$ -glucan recognition by a family 41 carbohydrate-binding module from *Thermotoga maritima*, *J. Mol. Biol.* 365 (2007) 555–560, <https://doi.org/10.1016/j.jmb.2006.10.018>.
- [43] A.B. Boraston, A.L. Creagh, M.M. Alam, J.M. Kormos, P. Tomme, C.A. Haynes, R.A.J. Warren, D.G. Kilburn, Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A, *Biochemistry* 40 (2001) 6240–6247, <https://doi.org/10.1021/bi0101695>.
- [44] P.S. Satheshkumar, P. Gayathri, K. Prasad, H.S. Savithri, “Natively unfolded” VPg is essential for Sesbania mosaic virus serine protease activity, *J. Biol. Chem.* 280 (2005) 30291–30300, <https://doi.org/10.1074/jbc.M504122200>.
- [45] T.Y. Jiang, Y.P. Ci, W.I. Chou, Y.C. Lee, Y.J. Sun, W.Y. Chou, K.M. Li, M.D.T. Chang, Two unique ligand-binding clamps of *Rhizopus oryzae* starch binding domain for helical structure disruption of amylose, *PLoS One* 7 (2012) 2–14, <https://doi.org/10.1371/journal.pone.0041131>.
- [46] W. Haas, R. MacColl, J.A. Banas, Circular dichroism analysis of the glucan binding domain of *Streptococcus mutans* glucan binding protein-A, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 1384 (1998) 112–120, [https://doi.org/10.1016/S0167-4838\(98\)00005-3](https://doi.org/10.1016/S0167-4838(98)00005-3).
- [47] C. Oliveira, V. Carvalho, L. Domingues, F.M. Gama, Recombinant CBM-fusion technology – applications overview, *Biotechnol. Adv.* 33 (2015) 358–369, <https://doi.org/10.1016/j.biotechadv.2015.02.006>.

- [48] L.C. Gunnarsson, E.N. Karlsson, A.S. Albrekt, M. Andersson, O. Holst, M. Ohlin, A carbohydrate binding module as a diversity-carrying scaffold, *Protein Eng. Des. Sel.* 17 (2004) 213–221, <https://doi.org/10.1093/protein/gzh026>.
- [49] L. Filonova, L.C. Gunnarsson, G. Daniel, M. Ohlin, Synthetic xylan-binding modules for mapping of pulp fibres and wood sections, *BMC Plant Biol.* 7 (2007) 54, <https://doi.org/10.1186/1471-2229-7-54>.
- [50] L.C. Gunnarsson, Q. Zhou, C. Montanier, E.N. Karlsson, H. Brumer, M. Ohlin, Engineered xyloglucan specificity in a carbohydrate-binding module, *Glycobiology* 16 (2006) 1171–1180, <https://doi.org/10.1093/glycob/cwl038>.
- [51] L.C. Gunnarsson, L. Dexlin, E. Nordberg, O. Holst, M. Ohlin, Evolution of a carbohydrate binding module into a protein-specific binder, *Biomol. Eng.* 23 (2006) 111–117, <https://doi.org/10.1016/j.bioeng.2005.12.002>.
- [52] P.J. Simpson, S.J. Jamieson, M. Abou-Hachem, E. Nordberg Karlsson, H.J. Gilbert, O. Holst, M.P. Williamson, The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase, *Biochemistry* 41 (2002) 5712–5719, <https://doi.org/10.1021/bi012093i>.
- [53] R. Cohen-Khait, O. Dym, S. Hamer-Rogotner, R. Cohen-Khait, O. Dym, S. Hamer-Rogotner, G. Schreiber, Promiscuous protein binding as a function of protein stability, *Structure* 25 (2017) 1867–1874, <https://doi.org/10.1016/j.str.2017.11.002>.



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

Data concerning secondary structure and alpha-glucans-binding capacity of the *LaCBM26*

Silvia Armenta<sup>a,b</sup>, Zaira Sánchez-Cuapio<sup>a,c</sup>, Amelia Farrés<sup>d</sup>,  
Karen Manoutcharian<sup>a</sup>, Alejandra Hernandez-Santoyo<sup>e</sup>,  
Sergio Sánchez<sup>a</sup>, Romina Rodríguez-Sanoja<sup>a,\*</sup>

<sup>a</sup> Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México (UNAM), A.P. 70228, Ciudad Universitaria, Ciudad de México 04510, México

<sup>b</sup> Programa de Doctorado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México (UNAM), A.P. 70228, Ciudad Universitaria, México DF 04510, Mexico

<sup>c</sup> Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM), A.P. 70228, Ciudad Universitaria, México DF 04510, Mexico

<sup>d</sup> Facultad de Química, Universidad Nacional Autónoma de México, Circuito Exterior s/n Ciudad Universitaria, Ciudad de México 04510, México

<sup>e</sup> Instituto de Química, Universidad Nacional Autónoma de México, Circuito Exterior s/n Ciudad Universitaria, Ciudad de México 04510, México

## ARTICLE INFO

## Article history:

Received 16 October 2018

Received in revised form

9 November 2018

Accepted 12 November 2018

Available online 14 November 2018

## Keywords:

Carbohydrate-binding module

Starch

 $\alpha$ -Glucans recognition

## ABSTRACT

Carbohydrate-binding modules (CBMs) are auxiliary domains into glycoside-hydrolases that allow the interaction between the insoluble substrate and the solubilized enzyme, through hydrophobic, CH- $\pi$  interactions and hydrogen bonds. Here, we present the data article related to the interaction of one *LaCBM26* and some mutated proteins with soluble  $\alpha$ -glucans determined by enzyme-linked carbohydrate-binding assay, isothermal titration calorimetry (ITC), and affinity gel electrophoresis (AGE). The data of the behavior of proteins in presence and absence of substrate analyzed by circular dichroism CD and thermofluor are also presented. These results are complementary to the research article "The role of conserved non-aromatic residues in the *Lactobacillus amylovorus*  $\alpha$ -amylase CBM26-starch interaction" (Armenta et al., 2019).

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

DOI of original article: <https://doi.org/10.1016/j.ijbiomac.2018.10.061>

\* Corresponding author.

E-mail address: [romina@biomedicas.unam.mx](mailto:romina@biomedicas.unam.mx) (R. Rodríguez-Sanoja).

<https://doi.org/10.1016/j.dib.2018.11.056>

2352-3409/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications table

Subject area	Biochemistry
More specific subject area	Protein characterization
Type of data	Graph, figure
How data was acquired	Multiskan FC (Thermo Scientific); J-710 spectropolarimeter (Jasco Inc., Easton, MD, USA); ITC <sub>200</sub> (Malvern); CFX96 Real Time System C1000 Touch Thermal cycler (Bio-Rad Laboratories, Inc).
Data format	Analyzed
Experimental factors	Recombinant proteins were produced in <i>E. coli</i> and purified by immobilized metal affinity chromatography (ProBond Resin 46-0019, Invitrogen), followed by anion exchange chromatography (HiTrapTM DEAE FF 17–5055-01, GE Healthcare).
Experimental features	Enzyme-linked carbohydrate binding assay, CD in the far-UV, Fluorescence Thermal Shift Assay.
Data source location	Ciudad de México, México
Data accessibility	All data are shown within this article
Related research article	S. Armenta, Z. Sánchez-Cuapio, M.E. Munguia, N. Pulido, A. Farrés, K. Manoutcharian, A. Hernández-Santoyo, S. Moreno-Mendieta, S. Sánchez, R. Rodríguez-Sanoja, The role of non-aromatic residues in the <i>Lactobacillus amylovorus</i> $\alpha$ -amylase CBM26-starch interaction, Int J Biol Macromol. 121 (2019) 829–838 [1].

## Value of the data

- These data offer a system for the study of carbohydrate–protein interaction in an enzyme-linked carbohydrate-binding assay. The data obtained with this methodology agree with those observed with more used methodologies such as ITC and AGE.
- The data contribute to the study of proteins with atypical CD spectra that are rarely shown.
- Although active in binding, fluorescence signal plots (thermofluor) suggest an unfolding state of the LaCBM26 domain.

## 1. Data

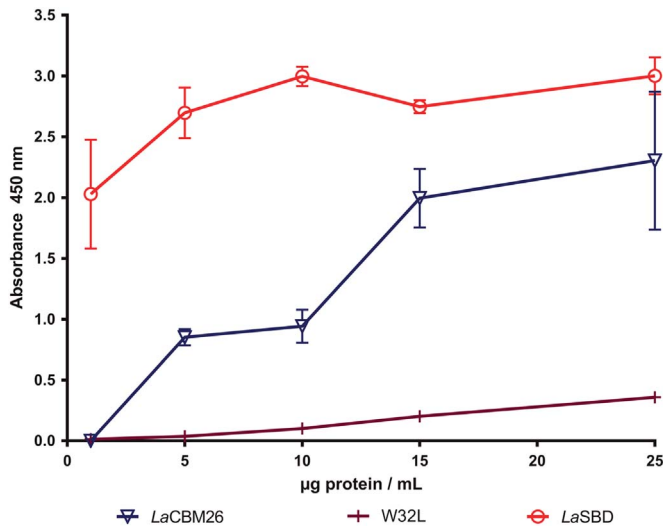
In these data, the binding capacity of *Lactobacillus amylovorus*  $\alpha$ -amylase Starch Binding Domain (LaSBD) and one of the five identical modules (LaCBM26) that constitutes the LaSBD domain was determined by an adapted system of the enzyme-linked carbohydrate-binding assay (Fig. 1). In parallel, the preference of LaCBM26 for linear substrates (glucose  $\alpha$ -1,4) or branched at  $\alpha$ -1,6 was investigated by affinity gel electrophoresis (Fig. 2). The thermostability of LaCBM26 was estimated by measuring the fluorescence emission by thermally-induced protein melting [2]. Unlike the LaSBD, LaCBM26 seems to have exposed his hydrophobic patches and no change was observed in the presence of ligand (maltoheptaose) (Fig. 3). Similarly, there is no change in the circular dichroism spectrum due to the presence of the ligand (Fig. 4).

## 2. Experimental design, materials, and methods

### 2.1. Protein production and purification

LaCBM26 and the mutated proteins were constructed and purified as previously described [1,3].





**Fig. 1.** Protein interaction to immobilized soluble starch. *LaSBD*: starch binding domain from *Lactobacillus amylovorus*  $\alpha$ -amylase formed by five CBM26 in tandem. *LaCBM26*: only one CBM26 domain of *LaSBD*. *W32L*: CBM26 with a mutation in W32 that disables it to bind to starch or derived  $\alpha$ -glucans. The error bars represent the standard deviation of triplicates.

## 2.2. Enzyme-linked carbohydrate-binding assay

Approximately 100  $\mu$ L of 10% soluble starch (Prolabo Fontenay-sous-Bois, France) was used to coat 96-well microplates (Costar 3370) in phosphate-buffered saline (PBS) buffer, pH 7.4 for 22 h at room temperature (RT). Wells were washed three times with PBS, then blocked with 0.2% bovine serum albumin (BSA) in PBS-0.05%Tween 20 (PBS-5T) for 60 min at room temperature and wash once with PBS-5T.

The accessibility of the starch for the interaction was validated by measuring the adsorption of the *LaSBD*, one *LaCBM26* and the null mutant protein *W32* (CBM26 *W32L*) that does not interact with starch; at three different concentrations: 1  $\mu$ g/mL, 10  $\mu$ g/mL, and 25  $\mu$ g/mL. Plates were incubated for 3 h at RT, followed by a wash step. The coated plate was then incubated 2 h at 37 °C with anti-His antibody (GeneTex) diluted 1:1,000 in PBS-BSA. Then, the plate was washed and incubated for 1.5 h at RT with anti-mouse IgG alkaline peroxidase (Invitrogen) diluted 1:5,000 in PBS-BSA. After washing, the plate was incubated for 30 min at RT with the FAST OPD substrate (SIGMA). Plates were read at 450 nm in a Multiskan FC (Thermo Scientific).

## 2.3. Affinity gel electrophoresis (AGE)

Affinity of mutated proteins and wild-type *LaCBM26* and derived mutated proteins for different carbohydrates was determined by affinity gel electrophoresis. The AGE method was performed with the Bio-Rad Mini-Protean III system (Bio-Rad, Richmond, CA). Purified proteins were separated in native gels containing 1% of potato starch (Prolabo, Fontenay-sous-Bois, France), amylose, amylopectin, pullulan, glycogen, cellulose, and birchwood xylan (all purchased from Sigma-Aldrich, St Louis Mo). Electrophoresis was run at 4 °C at 20 mA. Albumin was used as a negative, non-interacting control. Proteins were detected by staining with Coomassie brilliant blue [4].

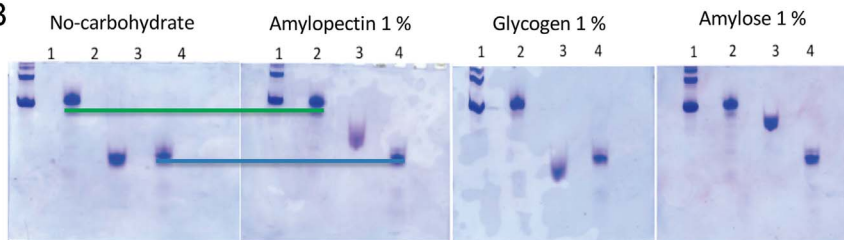
## 2.4. Isothermal titration calorimetry (ITC)

The quantitative binding capacity of the wild-type *LaCBM26* and derived mutated proteins was determined by ITC using an ITC<sub>200</sub> (Malvern) in 50 mM citrate-phosphate pH 5.5 at 25 °C as described

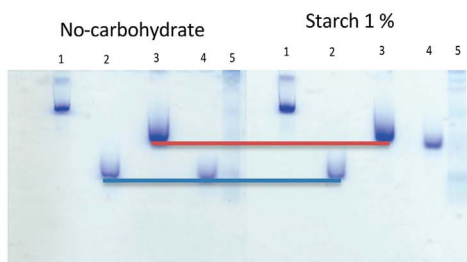
A

LaCBM26	1	KKVYFEKPSWGSRVYAYVYNKNTNKAITSAWPGKKMTALGNDKYELDLDTDEDDSDLAVIFTDGKTQTPAANEAGFTFTADATYDQNGVV	91
W32L	1	.....L.....	91
MutQD	1	.....Q.D.....D.....Y.L.....	91
MutNR	1	.....N.R.....H.....M.QQMPLMTKMVS-----	85

B

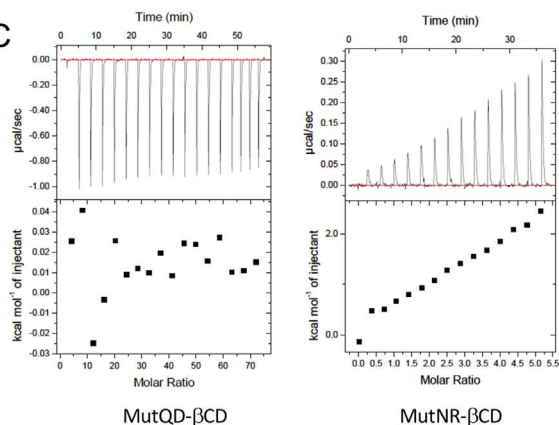


1) BSA, 2) MutNR (—), 3) CBM26, and 4) W32L (—)



1) BSA, 2) W32L (—), 3) MutQD (—), and 4) CBM26

C



MutQD-βCD

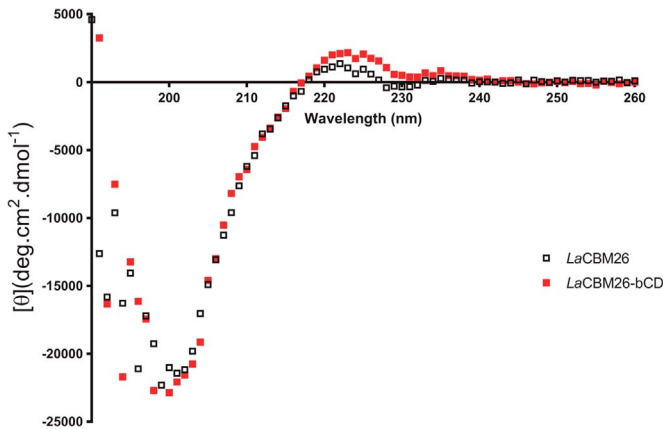
MutNR-βCD

**Fig. 2.** Glucan-binding of the *LaCBM26* mutated proteins. (A) Alignment of primary structure of the *LaCBM26* mutated proteins. (B) Affinity gel electrophoresis of representative variants. Image show the loss of interaction of the variants towards  $\alpha$ -glucans. (C) Non-interaction observed by ITC calorimetry of selected variants.

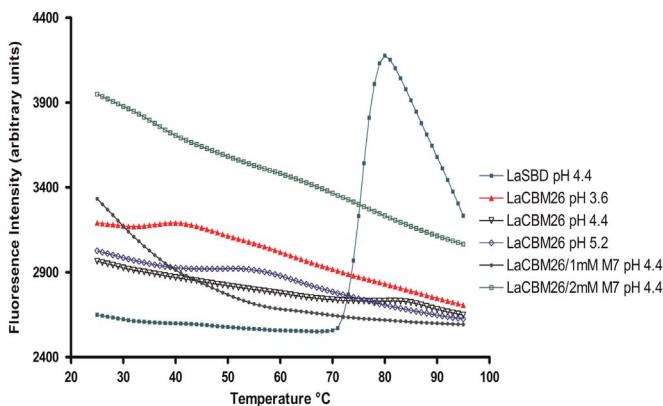
in [1]. The data show no interaction between the protein and the carbohydrate; thus, no data analysis was done.

### 2.5. Thermal shift assay (*ThermoFluor*)

To monitor protein unfolding, the hydrophobic fluorophore SYPRO orange was used. The thermal shift assay was conducted in the CFX96 Touch Real Time PCR Detection System (Bio-Rad Laboratories, Inc.).



**Fig. 3.** Thermofluor-based protein unfolding curves of the Starch Binding Domain from *Lactobacillus amylovorus*  $\alpha$ -amylase (*LaSBD*), and one CBM26 domain (*LaCBM26*) with and without maltoheptose (M7). The data were acquired in 1 M sodium lactate/HCl buffer at the pH indicated in the legend.



**Fig. 4.** Far-UV CD spectra of *LaCBM26* with and without  $\beta$ -cyclodextrin (bCD). Spectra were obtained in 10 mM sodium phosphate buffer pH 7.4, at 20 °C.

The buffer system used was 1 M sodium lactate/HCl buffer at different pH values (from 2.4 to 5.2), 1  $\mu$ L of SYPRO orange (SYPRO orange in DMSO, Sigma-Aldrich; San Luis MO, USA) was added to 1 mL of *LaCBM26* 10  $\mu$ M. Ten microliters of the protein mix and 10  $\mu$ L of the buffer solution were pipetted into the 96-well plate. The final volume of each reaction was 20  $\mu$ L. The plate was heated from 25 to 95 °C with a heating rate of 1 °C/min. Fluorescence changes in the wells of the plate are monitored simultaneously using excitation and emission wavelengths of 492 and 516 nm, respectively.

## 2.6. Circular dichroism (CD)

CD spectroscopy was carried out at 20 °C on a J-710 spectropolarimeter (Jasco Inc., Easton, MD, USA) with purified proteins as described in [1]. The effect of the addition of the ligand on the secondary structure was checked by adding to *LaCBM26* 13  $\mu$ M,  $\beta$ -cyclodextrin to a final concentration of 39  $\mu$ M. The raw data were converted to molar ellipticity ( $[\theta]$ ) using the formula:  $[\theta] = \theta / (l \times C \times N_r)$  where  $\theta$  is ellipticity in millidegrees,  $l$  is the cell path length in millimeters,  $C$  is the molar concentration of protein, and  $N_r$  is the number of residues [5].



## Acknowledgments

Armenta S. and Sánchez-Cuapio Z. were supported by personal grants from Consejo Nacional de Ciencia y Tecnología, Conacyt 215144 and 295830 respectively, México. This work was supported by UNAM-DGAPA grant IN209316. ITC and DC experiments were performed at the LabDP-UAEM and at the Chemistry Institute, UNAM.


## Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2018.11.056>.

## References

- [1] S. Armenta, Z. Sánchez-Cuapio, M.E. Munguia, N. Pulido, A. Farrés, K. Manoutcharian, A. Hernández-Santoyo, S. Moreno-Mendieta, S. Sánchez, R. Rodríguez-Sanoja, The role of non-aromatic residues in the *Lactobacillus amylovorus*  $\alpha$ -amylase CBM26-starch interaction, *Int. J. Biol. Macromol.* 121 (2019) 829–838.
- [2] M.W. Pantoliano, E.C. Petrella, J.D. Kwasnoski, V.S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B.A. Springer, P. Lane, F. R. Salemme, High-density miniaturized thermal shift assay as a general strategy for drug discovery, *J. Biomol. Screen.* 6 (2001) 429–440.
- [3] R. Rodríguez-Sanoja, N. Oviedo, L. Escalante, B. Ruiz, S. Sánchez, A single residue mutation abolishes attachment of the CBM26 starch-binding domain from *Lactobacillus amylovorus*  $\alpha$ -amylase, *J. Ind. Microbiol. Biotechnol.* 36 (2009) 341–346. <https://doi.org/10.1007/s10295-008-0502-y>.
- [4] P. Tomme, A. Boraston, J.M. Kormos, R.A.J. Warren, D.G. Kilburn, Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules, *Enzym. Microb. Technol.* 27 (2000) 453–458. [https://doi.org/10.1016/S0141-0229\(00\)00246-5](https://doi.org/10.1016/S0141-0229(00)00246-5).
- [5] S.M. Kelly, T.J. Jess, N.C. Price, How to study proteins by circular dichroism, *Biochim. Biophys. Acta - Proteins Proteom.* 1751 (2005) 119–139. <https://doi.org/10.1016/j.bbapap.2005.06.005>.

# Advances in molecular engineering of carbohydrate-binding modules

Silvia Armenta <sup>1</sup>, Silvia Moreno-Mendieta,<sup>2</sup> Zaira Sánchez-Cuapio,<sup>1</sup> Sergio Sánchez,<sup>1</sup> and Romina Rodríguez-Sanoja<sup>1\*</sup>

<sup>1</sup> Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Circuito Mario de la Cueva s/n Ciudad Universitaria, Ciudad de México 04510, México

<sup>2</sup> CONACYT, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Circuito Mario de la Cueva s/n Ciudad Universitaria, Ciudad de México 04510, México

## ABSTRACT

Carbohydrate-binding modules (CBMs) are non-catalytic domains that are generally appended to carbohydrate-active enzymes. CBMs have a broadly conserved structure that allows recognition of a notable variety of carbohydrates, in both their soluble and insoluble forms, as well as in their alpha and beta conformations and with different types of bonds or substitutions. This versatility suggests a high functional plasticity that is not yet clearly understood, in spite of the important number of studies relating protein structure and function. Several studies have explored the flexibility of these systems by changing or improving their specificity toward substrates of interest. In this review, we examine the molecular strategies used to identify CBMs with novel or improved characteristics. The impact of the spatial arrangement of the functional amino acids of CBMs is discussed in terms of unexpected new functions that are not related to the original biological roles of the enzymes.

Proteins 2017; 85:1602–1617.  
© 2017 Wiley Periodicals, Inc.

**Key words:** molecular evolution; protein engineering; mutation; carbohydrate binding; protein–carbohydrate interaction.

## INTRODUCTION

Carbohydrate-binding modules (CBMs) are auxiliary domains usually attached to carbohydrate-active enzymes (CAZymes), which are proteins involved in the synthesis, transport, and metabolism of carbohydrates. CBMs typically display carbohydrate specificity similar to that of the catalytic module of the parent enzyme and are capable of recognizing a wide variety of carbohydrates found in nature. Functionally, CBMs might improve catalysis by positioning the substrate closer to the catalytic domain and by inducing conformational changes in polysaccharide chains, such as cellulose, chitin, or starch.<sup>1</sup>

With respect to enzyme architecture, CBMs might be present as isolated units or as composite tandem units of the same or different CBM families; in addition, tandem organization increases the substrate affinity, and cooperative binding effects have also been observed.<sup>2</sup> In addition, composite tandems of different CBM families increase the repertoire of polysaccharides recognizable by the enzyme.<sup>3</sup>

CBMs have a broadly conserved structure, but the topology of the ligand-binding site might vary between different families, allowing the binding of substrates with a broad range of

structures. This is the main reason why CBMs are ideal targets in protein engineering projects focused on changing or improving specificity toward substrates of interest.

Current knowledge of their structural features and proposed mechanisms of action cannot completely explain the broad diversity of carbohydrates recognized by the CBMs. Here, we review the main information available regarding the process of carbohydrate recognition by CBMs and their use as evolutionary platforms that are ideal for the development of new biotechnological applications.

## CBM-CARBOHYDRATE INTERACTION

The three-dimensional structure of CBMs is conserved in most cases, despite their low sequence identity. The

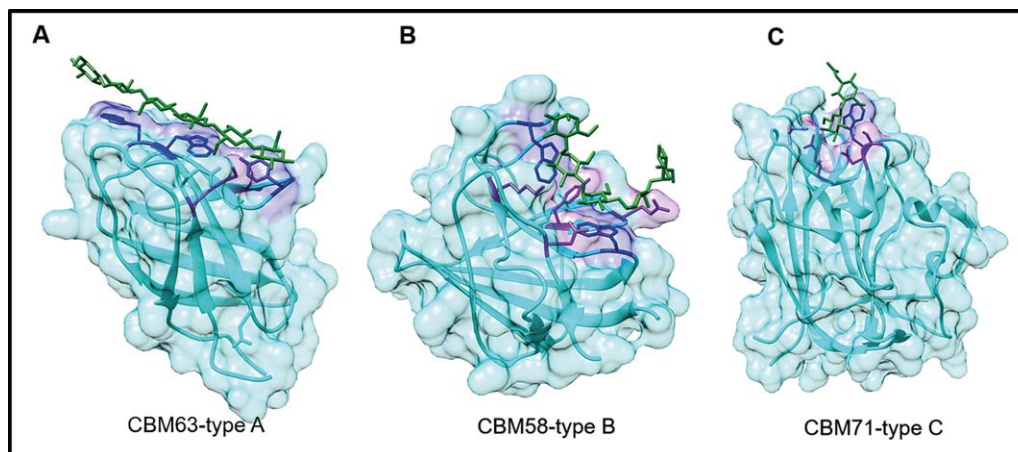
Grant sponsor: Consejo Nacional de Ciencia y Tecnología Conacyt, México; Grant sponsor: UNAM-DGAPA; Grant number: IN209316.

\*Correspondence to: R. Rodríguez-Sanoja, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Circuito Mario de la Cueva s/n Ciudad Universitaria, Ciudad de México 04510, México.

E-mail: romina@biomedicas.unam.mx

Received 3 March 2017; Revised 4 May 2017; Accepted 20 May 2017

Published online 25 May 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.25327



**Figure 1**

Topology of the binding site of three types of CBMs. (A) CBM63 from the *Bacillus subtilis* expansin (EXLX1) in complex with cellobiose (PDB ID 4FER). (B) CBM58 from the *Bacteroides thetaiotaomicron*  $\alpha$ -amylase (susG-D498N) with maltopentaose (PDB ID 3K8L). (C) CBM71 from the *Streptococcus pneumoniae*  $\beta$ -galactosidase (BgaA) in complex with  $\beta$ -D-galactopyranosyl-1,4-N-acetyl-D-glucosamine (PDB ID 4CUB). The backbone is displayed as a cyan cartoon with a transparent solvent-accessible surface. Aromatic amino acids establishing CH- $\pi$  interactions are shown in blue color, amino acids forming hydrogen bonds are shown in purple (with the side chains displayed in sticks), and the substrate is shown in green. The visualizations were generated using the UCSF Chimera package (<https://www.cgl.ucsf.edu/chimera>).

main fold is a  $\beta$ -sandwich, although other types of folds, such as the  $\beta$ -trefoil, cysteine knot, OB fold, hevein fold and hevein-like fold, have been observed.<sup>4,5</sup>

According to the topology of the ligand-binding site, CBMs can be classified into types A, B, or C (Fig. 1).<sup>1,3,4</sup> In all cases, substrate binding is mediated by aromatic amino acids, particularly tryptophan and tyrosine through hydrophobic and CH- $\pi$  interactions, resulting in the formation of hydrophobic platforms at the CBM-binding site. Hydrogen bonds formed by tyrosines or polar amino acids stabilize the interaction, although their contribution is more important for type B and C topologies.<sup>6,7</sup>

Because this type of topology mirrors that of the substrate, the main feature of type A CBMs is a planar and hydrophobic surface capable of binding insoluble carbohydrates, such as crystalline cellulose and chitin, through entropically driven CH- $\pi$  interactions between the enzyme's aromatic amino acids and the sugar residues. It has been reported that type A CBMs present little or no affinity for oligosaccharides.<sup>8,9</sup>

In contrast, the type B (or endo-type) binding site topology has the shape of a slit or cavity that accommodates individual substrate strands.<sup>3,8</sup> The binding site can be located in two regions of the three-dimensional structure. The first region, called the variable loop site (VLS), is located toward the end of the structure, where multiple loops can be found. The second region, which is called the concave face site (CFS), is located on the concave side of the structure.<sup>1</sup> In most cases, the binding site is located in the CFS region, but there are families with two binding sites, involving both regions (i.e.,

CBM6, CBM20, CBM21, and CBM34); in such cases, the binding site located in the VLS region could have a more specialized function for binding larger, branched polysaccharides.<sup>1</sup>

In general, type B CBMs cannot interact with planar substrates, such as crystalline cellulose, but can bind to a wide variety of glycans, such as xylans, mannans, galactans, and starch.<sup>7,9-11</sup> Similar to type A CBMs, these interactions are governed by aromatic amino acids through hydrophobic and CH- $\pi$  interactions. Hydrogen bonds established mainly by polar residues are determinants in the selectivity and affinity of CBMs for their substrates.<sup>12-15</sup> We should note that similar hydrogen bonds can also be found but are less determinant in type A CBMs. In general, the type B binding reaction is dominated by favorable changes in enthalpy that are partially compensated by unfavorable entropy changes.<sup>6,16</sup>

The type C (or exo-type) binding site displays a pocket topology that is able to recognize exposed or short-chain carbohydrates. Type C CBMs show higher affinity toward monosaccharides, disaccharides, and trisaccharides, and it is apparent that hydrogen bonds are more essential in this topology than in the type B topology.<sup>3,4</sup> It is possible for domains belonging to this group to recognize more complex polysaccharides, but this recognition relies on the ability to bind the carbohydrates exposed on the surface of complex substrates.

Some of the known CBMs are quite specific, whereas others tolerate some slight structural differences. Therefore, not only CBM structure but binding site topology are very important factors for understanding the mechanism of recognition of these proteins. Table I classifies

**Table 1**  
Classification of the CBM Families According to the Topology of Their Binding Site

Topology	CBM family <sup>a</sup>	CBM origin (representative example)	Ligand binding function	Reference	
A	1	Cellobiohydrolase I, <i>Trichoderma reesei</i>	Cellulose, hemicellulose, chitin	17	
	2	Chitinase, <i>Pyrococcus furiosus</i> ; Xylanase, <i>Cellulomonas fimi</i>	Cellulose, xylan, chitin	18,19	
	3	Endo- $\beta$ -1,4-glucanase, <i>Clostridium cellulolyticum</i> ; endo/exo- $\beta$ -1,4-glucanase, <i>Thermobifida fusca</i> YX	Cellulose, chitin	20,21	
	5	Chitinase, <i>Serratia marcescens</i> ; endo- $\beta$ -1,4-glucanase, <i>Dickeya dadantii</i>	Chitin	22	
	63	Expansin, <i>Bacillus subtilis</i>	Cellulose	9	
	64	4- $\beta$ -D-glucan 4-glucanohydrolase, <i>Spirochaeta thermophila</i>	Cellulose	23	
	79	Celulosome from <i>Ruminococcus flavefaciens</i> FD-1	Cellulose, $\beta$ -glucans, xyloglucan, HEC	24	
	B	2	Xylanase, <i>C. fimi</i>	Xylan	10,25
		4	Endo- $\beta$ -1,4-glucanase, <i>C. fimi</i> ; Xylanase, <i>Rhodothermus marinus</i> ; Laminarinase, <i>Thermotoga maritima</i> MSB8	Xylan, $\beta$ -1,3-glucan, $\beta$ -1,3-1,4-glucan, $\beta$ -1,6-glucan, amorphous cellulose	26,27
		6	Xylanase, <i>C. stecorarium</i> , <i>Ruminiclostridium thermocellum</i> ; $\beta$ -1,3-gluconase/laminarinase, <i>B. halodurans</i> C-125	Amorphous cellulose, $\beta$ -1,4-xylan, $\beta$ -1,3-glucan and $\beta$ -1,4-glucan	28–30
11		Endo- $\beta$ -1,4-glucanase, $\beta$ -1,4-xylanase, <i>R. thermocellum</i>	$\beta$ -1,4-glucan, $\beta$ -1,3-1,4-glucans	31	
15		Xylanase, <i>Cellvibrio japonicus</i>	Xylan, xylo-oligosaccharides	32	
16		$\beta$ -mannanase A, <i>Caldanaerobius polysaccharolyticus</i>	Cellulose, glucomannan	14	
17		Endo- $\beta$ -1,4-glucanase, <i>C. cellulovorans</i>	Amorphous cellulose, oligosaccharides	33	
20		$\beta$ -amylase, <i>B. cereus</i>	Starch	34	
21		Glucoamylase, <i>Rhizopus oryzae</i>	Starch	35	
22		Xylanase, <i>Paenibacillus barcinonensis</i> BP-23, <i>R. thermocellum</i>	Xylan, $\beta$ -1,3/1,4-glucans	36,37	
25		$\alpha$ -amylase, <i>B. halodurans</i>	Starch, amylose, amylopectin	11	
26		$\alpha$ -amylase, <i>B. halodurans</i>	Starch, amylose, amylopectin	11	
27		$\beta$ -mannanase, <i>Caldicellulosiruptor</i> sp., <i>T. maritima</i>	Mannan	38,39	
28		Endo- $\beta$ -1,4-glucanase, <i>B. akibai</i> , <i>C. josui</i>	Amorphous cellulose, $\beta$ -1,3-1,4-glucans	40,41	
29		Non-catalytic protein 1, <i>Piromyces equi</i>	Mannan, glucomannan	12	
31		Xylanase, <i>Alcaligenes</i> sp. XY234	$\beta$ -1,3-xylan	42	
34		$\alpha$ -amylase, <i>Thermoactinomyces vulgaris</i> R-47; neopullulanase, <i>Geobacillus stearothermophilus</i>	Starch	43,44	
35		Xylanases, <i>C. japonicas</i> ; CITase, <i>B. circulans</i>	Xylan/ $\beta$ -galactan, $\alpha$ -1,6-glucan	45,46	
36		Xylanase, <i>Paenibacillus polymyxa</i>	Xylan, xylo-oligosaccharides	47	
41		Pullulanase, <i>Klebsiella pneumonia</i> , <i>T. maritima</i> MSB8, <i>Streptococcus pneumonia</i> , <i>S. pyogenes</i>	Amylose, amylopectin, starch, pullulan, malto-oligosaccharides	13,48,49	
44		Endoglucanase J, <i>Clostridium thermocellum</i>	Cellulose, xyloglucan, $\beta$ -glucan, lichenan	50	
48		Branching enzyme, <i>Cyanothece</i> sp.; Pullulanase, <i>B. acidopullulyticus</i>	Glycogen, pullulan	51,52	
58		$\alpha$ -amylase (SusG), <i>Bacteroides thetaiotaomicron</i>	Starch, amylopectine, pullulan, $\beta$ -cyclodextrin,	53	
61		Endo- $\beta$ -1,4-galactanase, <i>Thermotoga maritima</i>	$\beta$ -1,4-galactan, pectin	54	
65		Endoglucanase <i>Eubacterium cellulosolvens</i> 5	Xyloglucan, $\beta$ -glucan, HEC	15	
70		Hyaluronate lyase, <i>S. pneumoniae</i>	Hyaluronan	55	
75		Celulosome, <i>R. flavefaciens</i> FD-1	Xyloglucan	24	
76		Celulosome, <i>R. flavefaciens</i> FD-1	Xyloglucan, glucomannan, $\beta$ -glucan, HEC	24	
77	Celulosome, <i>R. flavefaciens</i> FD-1	Pectin	24		
78	Celulosome, <i>R. flavefaciens</i> FD-1	Xyloglucan, glucomannan, $\beta$ -glucan, HEC,	24		
80	Celulosome, <i>R. flavefaciens</i> FD-1	Cellulose, $\beta$ -mannan	24		
81	Endoglucanase, sugar cane soil metagenome	$\beta$ -glucan, xyloglucan, amorphous cellulose	7		
C	9	Xylanase A, <i>T. maritima</i> MSB8	Glucose, cellobiose	56	
	13	Actinohivin, <i>Actinomyces</i> K97; Xylanase 10A, <i>S. lividans</i>	$\alpha$ (1-2)mannobiose/lactose, galactose	57,58	
	14	Chitinase/chitotriosidase, <i>Homo sapiens</i>	Chitotriose	59	
	32	Sialidase, <i>Micromonospora viridifaciens</i> /PGA-BP, <i>Yersenia enterocolitica</i> /LacNAc'ase, <i>C. perfingens</i>	Galactose, lactose/polygalacturonic acid, $\beta$ -D-galactosyl-1,4- $\beta$ -D-N-acetylglucosamine	60–62	
	42	Exo- $\alpha$ -1,5-L-arabinofuranosidase, <i>S. avermitilis</i> MA-4680; $\alpha$ -L-arabinofuranosidase B, <i>A. kawachii</i> IFO 4308	Arabinofuranose present in arabinoxylan	63	

**Table 1**

(Continued)

Topology	CBM family <sup>a</sup>	CBM origin (representative example)	Ligand binding function	Reference
	62	Xylanase, <i>Ruminiclostridium thermocellum</i> ATCC 27405	Galactose present in xyloglucan, arabinogalactan and galactomannan	64
	66	Levanase, <i>Bacillus subtilis</i> subsp. <i>Subtilis</i> str. 168	Fructose present in fructans	65
	71	$\beta$ -galactosidase (BgaA), <i>Streptococcus pneumoniae</i> R6	Lactose, LacNAc	66

<sup>a</sup>Some CBM families are not included because the structural information is not sufficient for their classification.

HEC, Hydroxyethyl cellulose; CItase, Cycloisomaltooligosaccharide glucanotransferase; PGA-BP, polygalacturonic acid-binding protein; LacNAc,  $\beta$ -D-galactosyl-1,4- $\beta$ -D-N-acetylglucosamine; LacNAc'ase,  $\beta$ -D-galactosyl-1,4- $\beta$ -D-N-acetylglucosaminidase.

the families present in the CAZY database according to the topology of their binding site.

## IMPORTANCE OF AROMATIC AMINO ACID RESIDUES FOR CARBOHYDRATE BINDING

The roles that aromatic amino acids play in CBM-carbohydrate interactions have been widely studied through structural imaging or directed mutagenesis experiments. For example, the cellulose-binding CBM2b mutants W570A and W602A demonstrate the importance of these amino acids because the removal of their side chains causes a twofold decrease in affinity.<sup>10</sup>

For CBM61 from the *Thermotoga maritima* endo- $\beta$ -1,4-galactanase, the W594A mutation abolishes binding to  $\beta$ -1,4-galactan, and mutation of either tryptophan, W508A or W563A, results in a 10-fold reduction in affinity. These amino acids are also important for adopting the curvature of the ligand, giving them a critical role in galactan binding.<sup>54</sup>

In addition, removal of the aromatic ring of the W32L mutant of CBM26 of the *Lactobacillus amylovorus*  $\alpha$ -amylase causes complete loss of substrate binding, whereas the Y18L and Y20L mutations result in a 50% loss in its affinity to starch.<sup>67</sup>

The same behavior was observed with the three aromatic amino acids (Y22, W65, and Y129) that form the hydrophobic platform in the binding site of the CBM11 of the cellulosomal bifunctional cellulase from *Clostridium thermocellum*. The replacement of any of these residues by alanine resulted in decreased binding of carboxymethylcellulose (CMC),  $\beta$ -glucan, glucomannan, and xylan. The highest effect was observed with the W65A mutant, whereas lesser impacts are observed with the Y129A and Y22A mutants.<sup>68</sup>

CBM20 of *Aspergillus awamori* var. *kawachi* glucoamylase (GAI) has two binding sites formed by three tryptophans: W589 and W542 in binding site 1 (SB1) and W562 in binding site 2 (SB2). The effects of the substitution of different amino acids at the same residue, that is, mutating tryptophan 562 to leucine (W562L), phenylalanine (W562F), and glycine (W562G), were investigated.

The W562G mutant did not bind  $\beta$ -cyclodextrin and showed no enzymatic activity, whereas the W562L and W562F mutants maintained their substrate-binding capabilities. For residue 589, the W589Q mutant retained its capacity for binding  $\beta$ -cyclodextrin and showed enzymatic activity similar to that of the wild type.<sup>69</sup>

Each of the above examples provides experimental evidence for the importance of aromatic amino acids in the binding sites of CBMs. Although they are crucial for proper recognition of the substrate by the enzyme, interactions established by aromatic amino acids are complemented by hydrogen bonding, a factor that contributes decisively to the stability of binding and likely to the plasticity of recognition.

## IMPORTANCE OF POLAR AMINO ACIDS FOR CARBOHYDRATE BINDING

Polar amino acids are important for hydrogen bonds that stabilize the carbohydrate-CBM interaction.<sup>4</sup> Discrete modifications within the binding site of CBMs can affect the orientation, form and distribution of charges, influencing their substrate specificity and affinity. In *Cel lulomonas fimi*, the CBM2b-2 domain of xylanase 11 A has an affinity constant for soluble xylan of  $1.2 \times 10^4 \text{ M}^{-1}$  and cannot bind insoluble cellulose. However, changing arginine 573 to glycine modifies the specificity of the domain from xylan to cellulose, highlighting the importance of polar amino acids in this interaction. Researchers have observed that modification of the side-chain residues allows rearrangement of the aromatic amino acids in the binding site to contribute to the change in specificity.<sup>10</sup>

A certain degree of promiscuity in CBMs could be caused by flexibility in the formation of hydrogen bonds at the binding site. For example, the *Thermoanaerobacterium polysaccharolyticum* CBM16-1 of mannanase binds  $\beta$ -1,4-glucose and  $\beta$ -1,4-mannose polymers, and this binding is driven by tryptophans 20 and 125. However, glutamine 81 also plays an important role because this residue is able to interact with the C-2 hydroxyl at site 3 in either the axial or equatorial configuration. These



different configurations of the C-2 hydroxyl are accommodated within the protein scaffold by compensatory hydrogen bond interactions.<sup>14</sup>

The same strategy has been reported for the CBM29 domain in the *Piromyces equi* non-catalytic protein 1. The orientation of the aromatic residues in this enzyme complements the conformation of the target sugar polymer, while accommodation of manno- and gluco-configured oligosaccharides is made possible by the plasticity of arginine 112 and glutamine 116, which allow the formation of new hydrogen bonds.<sup>12</sup>

Similarly, CBM families that recognize  $\alpha$ -1,4-glucans can tolerate the  $\alpha$ -1,6 links that are present in substrates such as pullulan. For example, CBM41 from *Thermotoga maritima* pullulanase tolerates  $\alpha$ -1,6 linkages, but only in the context of a sufficient number of  $\alpha$ -1,4-linked glucose residues. The CBM41 structure in complex with maltotetraose (M4) and 63- $\alpha$ -D-glucosyl-maltotriose (GM3) shows that the domain establishes eight and seven direct hydrogen bonds, respectively. Of these, two hydrogen bonds with M4 are lost in the interaction with GM3, but one new hydrogen bond with glutamine 36 and two hydrogen bonds mediated by water molecules are added. The authors propose that the observed preference is due to the nature of the enzyme; thus, CBM41 would target regions rich in  $\alpha$ -1,4 linkages, which would likely leave  $\alpha$ -1,6 linkages exposed and susceptible to cleavage by the catalytic domain of pullulanase.<sup>13,48</sup>

Therefore, the orientation and establishment of bonds between the substrate and both aromatic and polar amino acids define the relevant enzymatic features, such as the affinity, specificity, and activity. Thus, aromatic residues and hydrogen bonding together form the chemical environment necessary for outlining, limiting, or expanding the CBM-carbohydrate interaction.

## APPLICATIONS OF CBMS IN BIOTECHNOLOGY

Because the main functions of CBMs are recognition and binding to diverse carbohydrates, CBM-carrying proteins play important biological roles in the processing of complex macromolecular structures, such as cell walls in plants, and the recognition of cellular glycans from pathogenic bacteria during pathogenesis or colonization.<sup>70–73</sup>

Recombinant DNA technology has enabled the development of new applications for CBMs. For example, CBMs are used to reveal the carbohydrate composition and distribution in the cell walls of plants, which allow the detection of specific carbohydrate structures related to biological processes.<sup>45,74,75</sup>

Similarly, CBMs have biotechnological potential in the treatment of agricultural wastes for back-end use in diverse applications. For biomass degradation, chimeric proteins have been designed by fusing CBMs to catalytic

domains from multiple enzymes, which could diversify the range of carbohydrate targets<sup>76</sup> and increase the enzymatic activity.<sup>77,78</sup> For example, the fusion of CBM2a to two thermophilic endoglucanases enhances their catalytic activity on insoluble substrates up to threefold with respect to that of the WT.<sup>78</sup>

Biomass degradation is necessary as a pretreatment process for the saccharification of lignocellulosytic material, which would aid bioethanol or biofuel production. In addition to their use as components of chimeric enzymes, CBMs have been used as additives in reactions during the saccharification process of crystalline cellulose, increasing the production of glucose up to 30%.<sup>79</sup>

Lignocellulosytic material is not easily digested into sugars mainly due to the presence of lignin. Low-lignin rice was developed through the expression of a laccase-cellulose-binding domain (CBD) fusion protein, and this protein increases saccharification efficiency (up to 1.5-fold) and modifications in the cell wall composition.<sup>80</sup> Similarly, changes in the cell wall composition of *Arabidopsis thaliana* plants can be achieved through overexpression of the starch-binding domain (SBD123) of *A. thaliana* starch synthase III, which increases the cell wall degradability while increasing biomass, resulting in higher levels of fermentable sugars and hydrolyzed cellulose.<sup>81</sup>

Four tandem proteins, comprising two CBMs, have been used in the paper industry.<sup>35</sup> The CBM3–CBM3 protein shows promising potential for the paper industry because it can enhance the mechanical properties of paper, such as folding endurance (27.4%) and tensile strength (15.5%). Conversely, the CBM1–CBM1 tandem shows increased tensile (12.6%) and burst (8.8%) strengths, paving the way for an upgrade in secondary pulp processing, which is suitable for fiber recycling.<sup>82</sup>

Additionally, CBMs are present in the growing field of biomaterials development, including those enzymes that use cellulose as building blocks. Rivkin *et al.* (2015) developed bionanocomposite films by binding the recombinant resilin-like protein (res) from *Drosophila melanogaster* fused to a cellulose-binding domain (CBD) over cellulose nanocrystals (CNCs). After testing different formulations (with 0.5–5% glycerol W/V), these researchers found that the res-CBD-CNCs films with 0.5% glycerol displayed improved mechanical properties compared with neat CNC films.<sup>83</sup>

Furthermore, a wide variety of biomedical applications require polymers made from starch and cellulose, such as partially biodegradable acrylic bone cements (SCA),<sup>84</sup> or controlled drug release hydrogels,<sup>85</sup> for which the construction of a cellulose/starch cross bridging protein (CSCP) could be of great use.<sup>86</sup>

Another biomedical application of great interest is the production of new biomaterials to replace cardiovascular vessels. Recently, a protein containing a CBD and an adhesion peptide (RGD or GRGDY) was used to

improve adhesion to human microvascular endothelial cells on bacterial cellulose.<sup>87</sup>

One of the most direct industrial applications of CBMs is their use as purification tags to catch unrelated proteins. CBMs from different families, such as those that are able to recognize cellulose<sup>88,89</sup> and starch,<sup>90,91</sup> have been used.

Similarly, CBM fusion is considered an excellent strategy for enzyme immobilization on a cellulose surface. Lu *et al.* (2012) used microcrystalline cellulose for the immobilization of  $\beta$ -galactosidase from *Lactobacillus bulgaricus* L3 fused with CBD as a great method for the industrial production of galactooligosaccharides, which are attractive prebiotics.<sup>92,93</sup>

In the same way, fusion of a starch-binding protein (SBDtag) to antigens such as the  $\alpha$ -crystalline protein from *Mycobacterium tuberculosis* or fragment C of the tetanus toxin has allowed their immobilization on starch, and the resulting product can be used as a potential mucosal vaccine vehicle for generating useful subunit vaccines against several diseases.<sup>94,95</sup>

In contrast, the use of CBMs as analytical reagents is limited by their low affinity toward glycans (millimolar).<sup>96</sup> Thus, the characteristics of CBMs can be improved through protein engineering to expand the catalog of biotechnological applications.

## STRATEGIES FOR GENERATING NOVEL CBMS WITH DIVERSE FUNCTIONAL PROPERTIES

The great diversity of carbohydrates recognized by CBMs has led to different strategies for improving the understanding of the binding mechanism. Due to the structural features of CBMs, it is believed that they could function as scaffolds.

Ideally, a scaffold is a protein with intrinsic conformational stability and a structure that separates the determinants of stability from the active site or the recognition determinants. However, few proteins present both of these characteristics. After the analysis of experimental data, Skerra (2000) proposed that a scaffold for engineering should have two important characteristics: (1) a well-defined hydrophobic core that must importantly contribute to the free energy of folding and (2) a highly conserved and stable fold with a solvent-accessible binding site that is separated from the core and capable of recognizing different substrates.<sup>97,98</sup>

Due to their structural characteristics, CBMs can potentially function as protein scaffolds because they present a highly conserved three-dimensional structure, predominantly the  $\beta$ -sandwich fold, between the different members of different families.<sup>4,5,99</sup> In addition, the binding site of these domains is usually located toward

regions exposed to the solvent with a generally extended topology.<sup>1,4</sup>

With respect to their intrinsic conformational stability, there is little information regarding whether CBMs can tolerate the insertion of random mutations without affecting their structural stability. However, bioinformatics tools might aid the prediction of the effects of mutations on protein stability by allowing a rapid pre-experimental evaluation of different target regions. Wiederstein and Sippl (2005) developed an efficient computational method for the quantitative evaluation of the impact of mutations on protein stability and proposed a helpful guide for the design of mutagenesis experiments.<sup>100</sup>

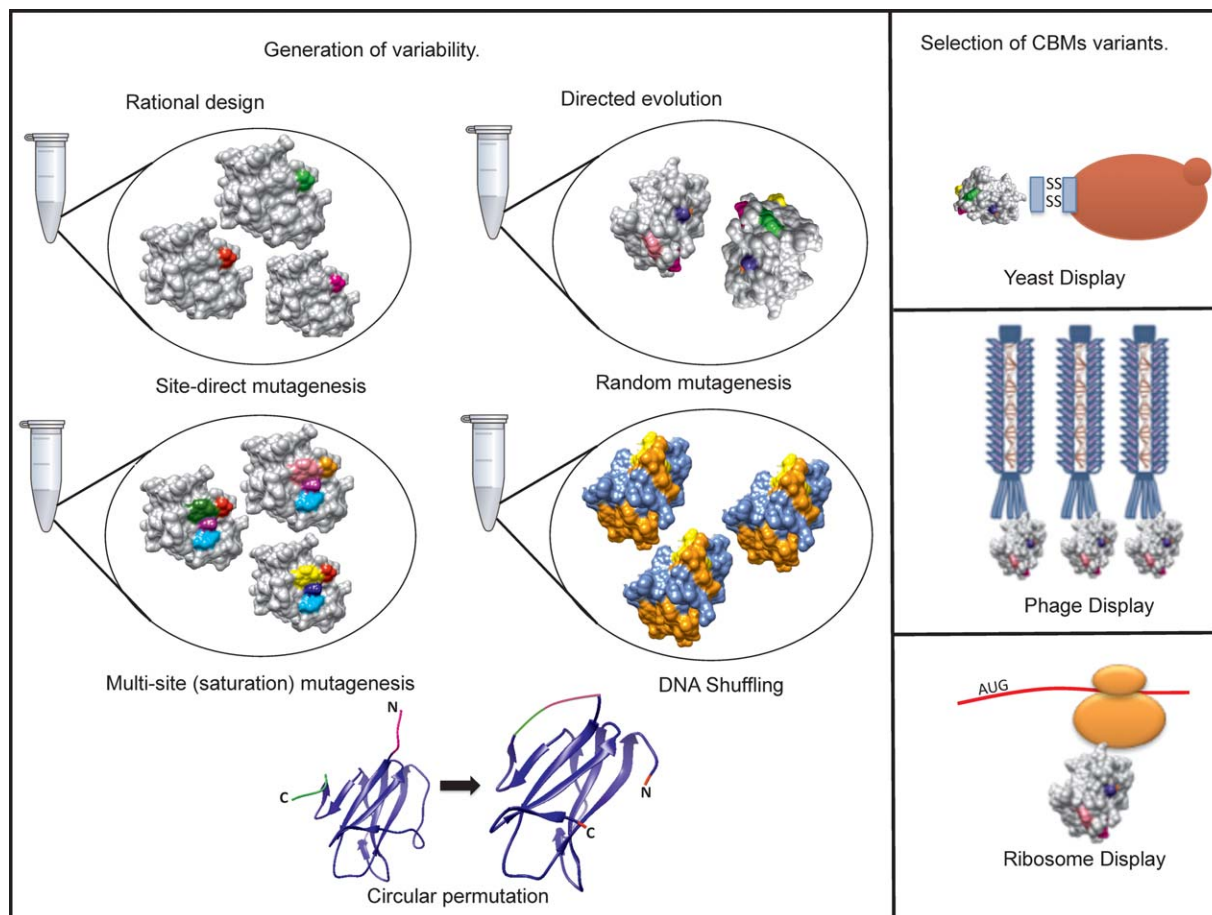
The main method used for generating variability in CBMs has been rational design. This method requires knowledge of the relationship between protein structure and function and a computational analysis that indicates which amino acids should be modified. Site-directed mutagenesis and site-directed saturation mutagenesis, as well as the circular permutation technique, are subsequently typically used to construct the modifications (Fig. 2).<sup>101</sup>

Circular permutation is a strategy for generating novel CBMs through ligation of the *N*- and *C*-termini of the protein and subsequent cleavage at another site to generate novel amino- and carboxyl-terminal ends. The CPred algorithm predicts possible circular permutation sites in diverse proteins.<sup>102</sup>

Stephen *et al.* (2012) predicted circular permutation sites in CBM21 from *Rhizopus oryzae* glucoamylase and found some interesting sites at the loops. These researchers proposed that the relocation of aromatic amino acids in these loops to obtain a major solvent-exposed surface area could improve carbohydrate binding. One of these mutant proteins, CP90, showed a substantial increase (up to 15-fold) in affinity toward soluble starch compared with the wild-type. This increase also shows selectivity because its binding to soluble starch was better than that to  $\beta$ -cyclodextrin and maltose.<sup>103</sup>

Site-directed mutagenesis is an important method for the modification and study of the structural and functional properties of a protein.<sup>104</sup> This strategy has been successfully used for the optimization of CBMs,<sup>105–107</sup> allowing researchers to not only know the role of certain amino acids in the recognition of specific substrates but also modify the affinity or specificity of the domains.

For example, Linder *et al.* (1995) studied the effect of site-directed mutations on the binding capacity of the CBM1 from *Trichoderma reesei* cellobiohydrolase I and found that replacement of tyrosine 5 with a tryptophan resulted in a twofold improvement of binding affinity.<sup>105</sup> Zhang *et al.* (2013) fused the *Cellulomonas fimi* CenA CBM2 to a cutinase and discovered that the mutation of the cellulose-binding residues W68Y and W68L modified the CBM specificity, nullifying its original binding

**Figure 2**

Strategies for CBM engineering. (A) Generation of variability through rational design (site-direct mutagenesis, saturation mutagenesis, and circular permutation) or direct evolution (random mutagenesis and DNA shuffling). (B) Different display systems (yeast display, phage display, and ribosome display) used for the selection process. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

capacity for cellulose and increasing its binding to PET fiber by 17% and 14%, respectively.<sup>107</sup>

Mutating the Y634 residue in the starch-binding domain (CBM20) of the *Bacillus macerans* cyclodextrin glycosyltransferase (CGTase) resulted in changes in thermostability and enzyme activity. Mutants were less stable ( $T_{50\%} = 49^{\circ}\text{C}$ ) than the WT enzyme ( $T_{50\%} = 57^{\circ}\text{C}$ ). Interestingly, the starch hydrolysis activity of the Y634G mutant was increased by approximately 25%. Although the effect of the mutation on CBM20 binding was not investigated, the authors proposed that the increase in activity could be caused by an increase in substrate affinity that results in a more processive enzyme. The effect of the Y634G mutation on thermostability is likely caused by an increase in flexibility, tyrosine has a bulky benzene ring, and glycine has only a hydrogen for a side-chain, allowing greater flexibility in the peptide chain.<sup>106</sup>

Another strategy for CBM design is oligonucleotide-directed randomization, which involves a semi-rational

approach through which specific residues are subjected to random mutagenesis.<sup>108</sup> This can be an important advantage because it is possible to randomize only the amino acids that are known to participate in substrate recognition without affecting other features, such as folding. For this strategy, it is necessary to know the three-dimensional structure of the domain or to use a structural model for the selection of amino acids. Regardless of the analytical tool used, it is not possible to predict the effect of the substitution on the affinity or other functional properties.

This strategy has been applied to three families of carbohydrate binding domains, specifically CBM1,<sup>109–111</sup> CBM4,<sup>112</sup> and CBM21.<sup>113</sup> In every case, the authors used assembly PCR and the introduction of mutations with degenerate codons, such as NNN or NNK, which allow incorporation of any of the four bases or partly randomized bases by restricting the mixture in some positions, for example, NNK ( $N = A/C/G/T$ ,  $K = G/T$ ).



Together, these techniques allow the insertion of any of the 20 amino acids, resulting in the generation of thousands of variations with potentially novel features.

Oligonucleotide-directed randomization has allowed improvements in substrate specificity by generating novel domains from some promiscuous CBMs, such as domain CBM4-2 from *Rhodothermus marinus* xylanase that binds xylan, xyloglucan,  $\beta$ -glucans, and amorphous cellulose. These novel domains include the variant of CBM4-2, namely X-2, which preferentially binds xylan, and the domain XG-34, which specifically binds xyloglucan.<sup>112,114</sup> Other examples include altering the cellulose-binding CBM1 from *T. reesei* cellobiohydrolase I to bind other molecules, such as  $\beta$ -glucuronidase<sup>109</sup> and porcine alpha-amylase proteins,<sup>110</sup> or altering CBM4-2 from *Rhodothermus marinus* xylanase to bind proteins such as human immunoglobulin G-4.<sup>115</sup>

Moreover, directed evolution methods such as random mutagenesis and gene shuffling can be used to alter protein properties. The main advantage of directed evolution over rational protein design is that a detailed understanding of the complex relationship between protein sequence, structure and function is not necessary.

Random mutagenesis causes mutations in any position of the CBM gene; therefore, this strategy mimics the natural evolution process and does not require previous knowledge of the protein structure. Due to the minimal control we have over the location of mutations, it is possible to create protein versions with changes in stability and folding. Random mutagenesis by error-prone polymerase chain reaction (epPCR) is the most commonly used technique for creating diversified gene libraries. This technique is performed under mutagenic reaction conditions (e.g., by the addition of small amounts of  $Mn^{2+}$  or increased amounts of  $Mg^{2+}$ ) to accentuate the natural error-prone properties of DNA polymerases.<sup>116</sup>

From the CBM4-2 library,<sup>112</sup> the variant XG-34 protein was selected and subjected to random mutagenesis. The obtained mutants showed increased affinity (by two orders of magnitude) for galactose-decorated xyloglucan. Based on the results of this work, the domain CBM4-2 from *R. marinus* xylanase and its variants were proposed as a protein scaffold.<sup>117</sup>

Given the data in the preceding paragraphs, it is possible to perform modifications in CBMs to improve their affinity and selectivity to different carbohydrates. The ease of obtaining CBMs with novel features or functional properties relies heavily on the selection methods and library construction.

## METHODS FOR SELECTING AND DISPLAYING ENGINEERED CBM VARIANTS WITH NOVEL FUNCTIONAL PROPERTIES

The search for novel and useful versions of CBMs among libraries of hundreds of millions of mutations

can be a tedious task. First, libraries are cloned and expressed in a display system such as ribosome display,<sup>118,119</sup> mRNA display,<sup>120,121</sup> cellular display,<sup>122</sup> yeast display,<sup>123</sup> or phage display.<sup>124</sup>

Yeast display technology requires fusing the library to the Aga2p agglutinin subunit protein, which is anchored on the yeast surface by the formation of disulfide bridges with the Agap1 protein.<sup>125</sup> Shiraga *et al.* (2004) expressed a combinatorial library of CBM21 of the *Rhizopus oryzae* glucoamylase on a yeast surface and selected mutants with improved glucoamylase activity.<sup>113</sup>

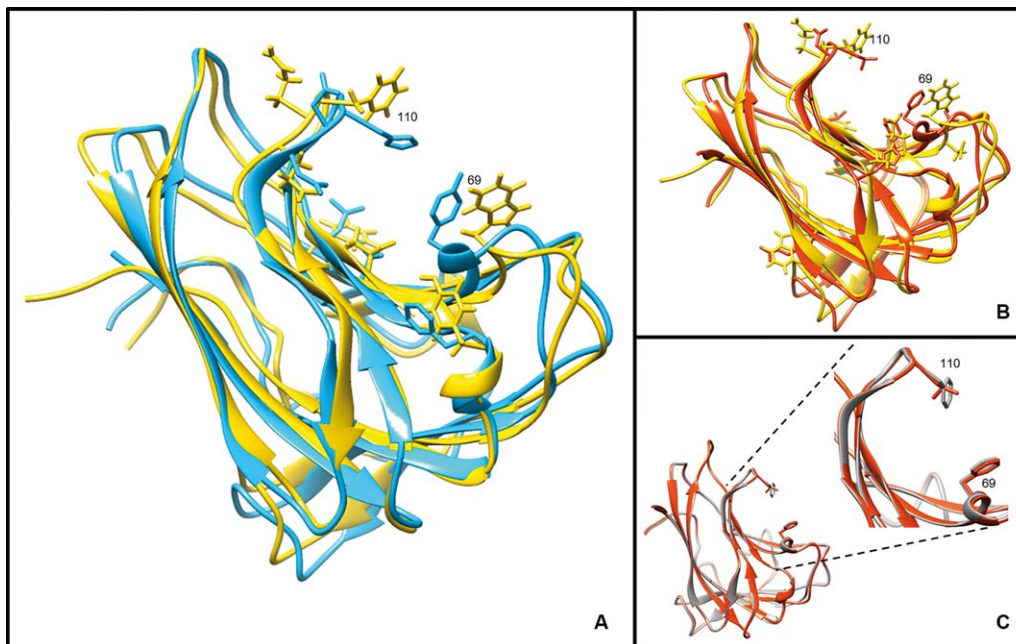
Phage display technology has been the most widely used approach for CBM engineering. This system is a robust technology that consists of expressing randomized sequences on the phage surface and selecting the improved affinity or specific selectivity for a given ligand through repeated rounds. The coat proteins of phage M13, pIII (five copies) and pVIII (approximately 2700 copies), are the most widely used for the deployment of recombinant proteins. The main system used for CBM display is the pIII protein, mainly because it supports large-sized proteins with fewer copies, resulting in the identification of variants with higher affinity toward a specific ligand.

Therefore, phage display has enabled researchers to find new proteins with different qualities. Some CBMs that have been displayed using this technique are CBM1 from the *T. reesei* cellobiohydrolase I (Cel7A) and CBM4-2 from the *R. marinus* xylanase, and both of these were fused to the pIII coat protein of phage M13.

The CBM1 from the *T. reesei* cellobiohydrolase I (Cel7A) was the first domain used as a scaffold for the engineering of novel binding proteins with distinct functions. This CBM is folded in a compact triple-stranded antiparallel  $\beta$ -sheet that constitutes a framework for the functionally important residues on the flat face (PDB 1AZ6).<sup>126</sup>

Smith *et al.* (1998) used cellulose and bovine alkaline phosphatase as platforms for the selection of CBM1 variants. These researchers isolated four variants (AP1-AP4) without the ability to bind cellulose but did bind bovine alkaline phosphatase without a change in phosphatase activity. Hence, these proteins appear to bind to the enzyme at a different site from the catalytic center. The AP1 mutant has important changes in the mutation sites, such as H4L, Y5G, N29R, Y31W, Y32F, and Q34M, which might cause structural changes and nullify the cellulose interaction.<sup>109</sup>

From the above-mentioned CBM1 combinatorial library, Letiö *et al.* (2000) selected two variants (CBDPPA1:5 and CBDPPA1:4), both of which are incapable of binding cellulose but able to bind to a protein, the porcine  $\alpha$ -amylase (PPA). The CBDPPA1:4 mutations included I11M, S14P, P16E, T17L, V18R, N29H, L28S, V27K, P30G, Y31L, and Y32V, which likely modified the structure of the domain. These variants are capable of



**Figure 3**

Structural modifications of CBM4-2 variants from the *Rhodothermus marinus* xylanase selected by phage display. (A) Structural comparison of CBM4-2 (yellow) from the *Rhodothermus marinus* xylanase (PDB ID 1K45) and the XG-34 variant (cyan) (PDB ID 3JXS) with a more restricted binding site. (B) Structural comparison of the same CBM4-2 (yellow) and the X-2 variant (orange) (PDB ID 2Y6J). Leucine 110 in the X-2 mutant is responsible for the change in the substrate specificity to xylan. (C) Structural comparison of the same X-2 variant (orange) and the X-2 L110F mutant (dark gray) (PDB ID 2Y6H). The residues in positions 69 and 110 are important for binding. The visualizations were generated with the UCSF Chimera package (<https://www.cgl.ucsf.edu/chimera>).

competing with binding to the amylase inhibitor acarbose, which suggests that the selected CBD variants bind to the active site of the enzyme PPA.<sup>110</sup>

The same CBM1 combinatorial library was used for the selection of enzymes for Ni<sup>2+</sup> binding. Eight analyses of potential Ni<sup>2+</sup>-binding CBMs revealed a striking preference for histidine residues at the randomized positions. These proteins were displayed on the *Staphylococcus carnosus* surface using the cell wall-anchoring region of *Staphylococcus aureus* protein A (SpA) for efficient anchoring to the cell wall. Researchers have demonstrated that recombinant *S. carnosus* cells carrying surface-displayed CBM variants showed a greater (150%) Ni<sup>2+</sup>-binding capacity than wild-type *S. carnosus* cells.<sup>111</sup>

The most widely studied domain is CBM4-2 from the *R. marinus* xylanase. This domain shows a  $\beta$ -sandwich structure formed by 11 strands, with a prominent cleft in the binding site that allows binding to a range of substrates, including xylans,  $\beta$ -glucans and xyloglucans.<sup>27</sup>

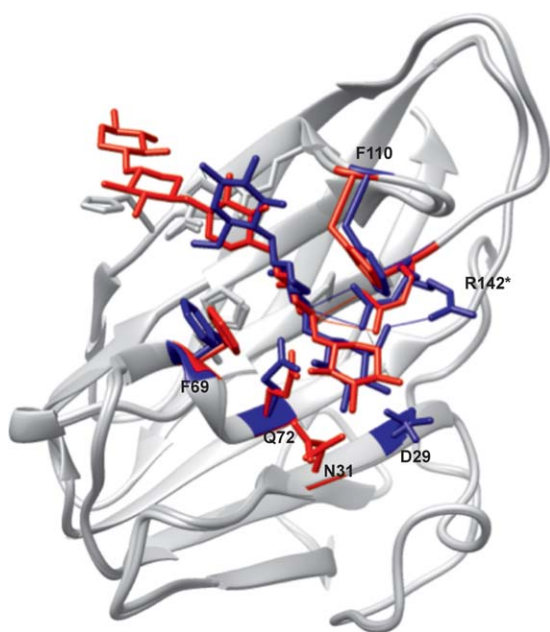
Gunnarsson *et al.* (2004) reported different selection strategies using a CBM4-2 combinatorial library displayed on the M13 phage. For ligand presentation, these researchers used a suspension of insoluble polysaccharides, such as birchwood xylan, Avicel (composed of ~50% amorphous and ~50% crystalline cellulose), ivory nut mannan, and Dynabeads coated with human

monoclonal IgG4,<sup>112</sup> and obtained a broad range of CBM4-2 variants with very interesting biochemical and functional qualities.

One of the selected CBMs was X-2, which was isolated from the insoluble fraction of the birchwood xylan as the selection ligand. X-2 conserved the classical  $\beta$ -sandwich fold and its specificity toward xylan but could not recognize glucose-configured polysaccharides, such as  $\beta$ -glucan, glucomannan, and xyloglucan.<sup>114</sup>

Although the X-2 variant is mutated in several positions (W69F, F110L, E72Q, F76L, Q111D, E118H, D79N, and W91R), its structural mechanism for xylan recognition is basically explained by the F110L mutation. Replication of this single mutation in the wild-type CBM4-2 yielded the same adsorption properties of the X-2 domain, that is, converting the cross-reactive module to a specific xylan binder. Conversely, the mutation of leucine 110 to phenylalanine in variant X-2 (X-2 L110F domain) reversed the adsorption capabilities of the X-2 L110F to the levels observed in the wild-type CBM4-2.<sup>127</sup>

As with other CBMs, hydrophobic stacking plays an important role in ligand attachment, that is, the F69 residue is situated opposite the L/F110 residue, creating the stacking zone (Fig. 3). However, the role of each amino acid is different: F69 provides the main  $\pi$ -surface in both



**Figure 4**

Relevance of the hydrogen bonds in the selectivity of the X-2L110F variant toward xylopentaose and cellopentaose. The main amino acids involved in hydrogen bonding are shown as sticks with respect to the oligosaccharides to which they bind. The visualization was generated with the UCSF Chimera package (<https://www.cgl.ucsf.edu/chimera>). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

modules and is apparently sufficient to bind xylopentaose and xylan, even if any of the aromatic amino acids (W, Y) or histidine are present at this site.<sup>114</sup> However, the presence of a phenylalanine (X-2L110F)<sup>127</sup> or a tyrosine (CBM4-2 F110Y)<sup>114</sup> in position 110 provides a second  $\pi$ -surface that helps to stabilize the recognition of gluco-configured polysaccharides, such as  $\beta$ -glucan and xyloglucan, diversifying the binding properties of CBMs.

The importance of hydrogen bonds for facilitating the CBM-substrate interaction was demonstrated in the X-2 and X-2 L110F modules because the same amino acids are interacting through hydrogen bonds, that is, the X-2 and X-2 L110F mutants establish seven hydrogen bonds with xylopentaose. The replacement of either Q72 or H146 with alanine completely abrogates the binding, indicating that these polar amino acids are crucial for xylan binding.<sup>127</sup>

Interestingly, the X-2 L110F mutant provides structural information for establishing new insights into CBM-carbohydrate interactions. The structure of X-2 associated with xylopentaose and cellopentaose shows the number of hydrogen bonds in such interactions: the interaction of X-2 L110F with cellopentaose involves only one bond from R142, whereas its interaction with xylopentaose involves the establishment of two hydrogen bonds with the same residue (Fig. 4). These results suggest the

relevance for these interactions on the substrate specificity of CBMs.<sup>128</sup>

Recently, the neutron structure of the same mutant, specifically X-2 L110F derived from CBM4-2 from the *R. marinus* xylanase, was reported,<sup>129</sup> providing the first CBM structure define using this technique and making it possible to demonstrate the importance of water molecules in hydrogen bonding in CBM-ligand interactions, a previously questioned issue. It is known that both the parental CBM X-2 and its X-2 L110F mutant contain ordered water molecules at the binding site and that at least 10 of these molecules are displaced by the union with a carbohydrate, although three of these molecules remain at the binding site. The precise orientation of these molecules cannot be identified through X-ray crystallography, but neutron diffraction allows the identification of a hydrogen-bonded network in which a molecule of water is positioned to accept a hydrogen bond from BGC1171 while forming hydrogen bonds with -OG1 of T104 and -OE2 of E112.

This technique allows the correct positioning of the amino acid side chains through networks linked by hydrogen bonds, demonstrating the relevance of water molecules in the preparation of the binding site, and demonstrating how the hydrogen bonds formed by water stabilize the interaction with both the ligand and the protein.<sup>129</sup>

In contrast, the same CBM4-2 library was selected for the identification of proteins that bind Dynabeads carrying immobilized xyloglucan (XG). In this study, stringency was modulated by the presence or absence of soluble xylan during the selection process for a protein with increasing affinity for xyloglucan. Among the large number of variants found,<sup>130</sup> the XG-34 mutant was selected. The engineered module XG-34 retains the wild-type folding and binding site. The main mutations in this variant (W69Y, F110H, E72H, F76L, Q111D, E112D, and Y149F) cause a binding cleft in XG-34 that is considerably more restricted than that in CBM4-2 (Fig. 3).<sup>131</sup> XG-34 lacks the wild-type affinity for xylan and  $\beta$ -glucan but displays specific binding affinity toward non-fucosylated xyloglucan.<sup>130</sup>

Additionally, Schantz *et al.* (2009) used the XG-34 variant to derive more mutants. The XG-34 domain library was again searched using Dynabeads carrying immobilized xyloglucan (XG), but in this study, the stringency was progressively modulated by decreasing the amount of ligand in the presence or absence of soluble xylan, with the aim of selecting only variants with greater substrate affinity.<sup>117</sup>

Two interesting variants were found: XG-34/1-X and XG-34/2-VI. The new XGBMs have higher affinity for galactose-decorated xyloglucan (XLLG; on the order of  $10^5 \text{ M}^{-1}$ ) than the wild-type XG-34 (on the order of  $10^3 \text{ M}^{-1}$ ). Interestingly, these variants have only two mutations compared with XG-34, and the most important of

**Table II**  
CBMs as Protein Scaffold

Family of engineered-CBM	Engineering strategies	Results	Advantage	Disadvantage	Reference
CBM21 CBM1 CBM4-2	Oligonucleotide-directed randomization and selection by yeast display	Novel variants that change their specificity from carbohydrate to proteins. In some cases, the specificity was improved toward carbohydrates as xylan and xyloglucan	Library size is large; allows control of the mutations; high efficiency	Structure-function relations must be clear; it is difficult to define the precise site of mutagenesis and their contribution in the selection mutants	113 109–111 112,114,115
XG-34 (variant of CBM4-2) CBM36	Random mutagenesis and phage display selection	Variants with a major affinity toward xyloglucan	Mimics natural evolution and does not require prior knowledge of structural data	Random; mutation library is large; low efficiency	117 132
CBM20 CBM2	Site-directed mutagenesis	Novel variants that improve or change their specificity to starch or polyethylene terephthalate (PET) fiber	Directed; simplifies methods for mutagenesis	Library size is small; structure–function relationships of the target enzyme should be clear; it is not possible to evaluate the contribution of different mutations at the same time	106 107

these is D112E. The glutamic acid residue at position 112 that is observed in the XG-34/1-X and XG-34/2-VI variants is responsible for the increased affinity to galactose-decorated xyloglucan because the D112E mutation in XG-34 shows increased affinity for galactose-decorated xyloglucan compared with that of the XG-34/2-VI mutant. Similarly, the addition of the E112D mutation to the XG-34/2-VI variant reverted its affinity for galactose-decorated xyloglucan to a value similar to that observed with the XG-34 mutant.<sup>117</sup>

Furthermore, the CBM4-2 library was selected using a human monoclonal antibody (IgG4) as the ligand. As expected, the selected variants recognized IgG4 and lost their ability to bind xylan and other carbohydrates. Interestingly, the authors reported the introduction of mutations that were not intended by the library design, that is, Q108R, D136G, or E138A/G. The relevance of those mutations in CBM4-2 from the *R. marinus* xylanase was evaluated residue by residue. However, none of these mutations modify the interaction with xylan or allow the binding to IgG4, showing the importance of the ensemble of mutations on the change in specificity.<sup>115</sup>

One of the selected variants (G-4) has an alanine at position 138, whereas CBM4-2 from the *R. marinus* xylanase has a glutamate at that position. The reversion of this mutation caused a complete loss of binding specificity for IgG4, without recovering the xylan-binding ability.<sup>115</sup> Thus, the authors suggest that mutations such as the E138A observed in the G-4 variant are important for the modified ability of this module to recognize human IgG4 but are not critically involved in xylan binding.

In domains such as CBM4-2 from the *R. marinus* xylanase, which is composed of 168 amino acids organized in a  $\beta$ -sandwich with 11  $\beta$ -strands, the modification of up to 12 amino acids allowed isolation of proteins that maintained and improved their specificity or affinity without changing the domain structure. However, modifications in this same domain led to the synthesis of a protein that ignores carbohydrates and recognizes proteins.

Every modification of CBM4-2 reviewed to date provides a framework for understanding how different residues are able to mediate the specificity and affinity toward different ligands through specific changes, both inside and outside of the carbohydrate-binding cleft, demonstrating the competence of this domain as a scaffold.

However, it appears unlikely that these results can be extrapolated to all CBMs or can be used to predict the modifications in specificity or affinity due to mutations in the domains. The engineering of CBMs is a promising approach for understanding the structural requirements of CBMs for achieving an optimal and diverse binding capacity to carbohydrates or different molecules.

Table II summarizes the principal results obtained in efforts to evolve different CBMs. Although few domains have been explored, the results clearly show that the size and arrangement of the domain have a direct impact on the transformation that the protein undergoes.

## CONCLUSIONS AND FUTURE PERSPECTIVES

CBMs are protein domains that are usually part of carbohydrate-active enzymes, with a discrete fold and



carbohydrate-binding activity. The recognition process can be understood through conformational studies, which have demonstrated the importance of aromatic residues in the hydrophobic interactions between carbohydrates and CBMs. In addition, these studies show how these interactions can be stabilized by the hydrogen bonds formed by polar residues. Although the importance of these residues at the binding cleft has been demonstrated, the determinants of selectivity are not clearly understood.

Modification of one single hydrogen bond can change the recognition of a linear carbohydrate to a branched one or, in other cases, improve affinity. However, the currently available knowledge and tools do not allow the rational selection of the residue(s) that should be modified to obtain the desired change in binding specificity or affinity. Molecular engineering is a powerful tool that is used to face these challenges by simplifying the analysis of several mutations at the same time.

Nevertheless, reports on the successful evolution of CBMs toward binding different ligands, such as proteins, constitute promising steps toward the use of CBMs as scaffolds or platforms for the design of new, interesting proteins. The engineered CBMs might have the potential to interact with different molecular targets, making them conducive to unexplored uses in different fields of interest, including research and industry.

As more CBMs are being incorporated into the databases, more possibilities for diversification exist. Therefore, given the plasticity for substrate recognition observed in CBMs, it is expected that many more CBMs of practical use will be created in the near future.

## ACKNOWLEDGMENTS

Silvia Armenta is a Ph.D. student of Biochemical Sciences at UNAM. The authors thank María Elena Munguía and Daniel Guillén for critically reading the manuscript.

## REFERENCES

- Abbott DW, van Bueren AL. Using structure to inform carbohydrate binding module function. *Curr Opin Struct Biol* 2014;28:32–40.
- Guillén D, Santiago M, Linares L, Pérez R, Morlon J, Ruiz B, Sánchez S, Rodríguez-Sanoja R. Alpha-amylase starch binding domains: cooperative effects of binding to starch granules of multiple tandemly arranged domains. *Appl Environ Microbiol* 2007;73:3833–3837.
- Gilbert HJ, Knox JP, Boraston AB. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr Opin Struct Biol* 2013;23:669–677.
- Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical J* 2004;382:769–781.
- Hashimoto H. Recent structural studies of carbohydrate-binding modules. *Cell Mol Life Sci* 2006;63:2954–2967.
- Xie H, Bolam DN, Nagy T, Szabó L, Cooper A, Simpson PJ, Lakey JH, Williamson MP, Gilbert HJ. Role of hydrogen bonding in the interaction between a xylan binding module and xylan. *Biochemistry* 2001;40:5700–5707.
- Campos BM, Liberato MV, Alvarez TM, Zanthorlin LM, Ematsu GC, Barud H, Polikarpov I, Ruller R, Gilbert HJ, Zeri AC, Squina FM. A novel Carbohydrate-binding module from sugar cane soil metagenome featuring unique structural and carbohydrate affinity properties. *J Biol Chem* 2016;291:23734–23743.
- Boraston AB. The interaction of carbohydrate-binding modules with insoluble non-crystalline cellulose is enthalpically driven. *Biochem J* 2005;385:479–484.
- Georgelis N, Yennawar NH, Cosgrove DJ. Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. *Proc Natl Acad Sci* 2012;109:14830–14835.
- Bolam DN, Xie H, White P, Simpson PJ, Hancock SM, Williamson MP, Gilbert HJ. Evidence for synergy between family 2b carbohydrate binding modules in *Cellulomonas fimi* Xylanase 11A. *Biochemistry* 2001;40:2468–2477.
- Boraston AB, Healey M, Klassen J, Ficko-Blean E, Van Bueren AL, Law V. A structural and functional analysis of  $\alpha$ -glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *J Biol Chem* 2006;281:587–598.
- Charnock SJ, Bolam DN, Nurizzo D, Szabó L, McKie VA, Gilbert HJ, Davis GJ. Promiscuity in ligand-binding: the three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. *Proc Natl Acad Sci USA* 2002;99:14077–14082.
- van Bueren AL, Boraston AB. The structural basis of  $\alpha$ -glucan recognition by a family 41 carbohydrate-binding module from *Thermotoga maritima*. *J Mol Biol* 2007;365:555–560.
- Bae B, Ohene-Adjei S, Kocherginskaya S, Mackie RI, Spies MA, Cann IKO, Nair SK. Molecular basis for the selectivity and specificity of ligand recognition by the family 16 carbohydrate-binding modules from *Thermoanaerobacterium polysaccharolyticum* ManA. *J Biol Chem* 2008;283:12415–12425.
- Luis AS, Venditto I, Temple MJ, Rogowski A, Baslé A, Xue J, Knox P, Prates JAM, Ferreira LMA, Fontes CMGA, Najmudin S, Gilbert HJ. Understanding how noncatalytic carbohydrate binding modules can display specificity for xyloglucan. *J Biol Chem* 2013;288:4799–4809.
- Boraston AB, Creagh AL, Alam MM, Kormos JM, Tomme P, Haynes CA, Warren RAJ, Kilburn DG. Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. *Biochemistry* 2001;40:6240–6247.
- Mattinen ML, Linder M, Drakenberg T, Annala A. Solution structure of the cellulose-binding domain of endoglucanase I from *Trichoderma reesei* and its interaction with cello-oligosaccharides. *Eur J Biochem* 1998;256:279–286.
- Nakamura T, Mine S, Hagihara Y, Ishikawa K, Ikegami T, Uegaki K. Tertiary structure and carbohydrate recognition by the chitin-binding domain of a hyperthermophilic chitinase from *Pyrococcus furiosus*. *J Mol Biol* 2008;381:670–680.
- Xu GY, Ong E, Gilkes NR, Kilburn DG, Muhandiram DR, Harris-Brandts M, Carver JP, Kay LE, Harvey TS. Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry* 1995;34:6993–7009.
- Mandelman D, Belaich A, Belaich JP, Aghajari N, Driguez H, Haser R. X-ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. *J Bacteriol* 2003;185:4127–4135.
- Sakon J, Irwin D, Wilson DB, Karplus A. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature* 1997;4:810–818.
- van Aalten DM, Synstad B, Brurberg MB, Hough E, Riise BW, Eijsink VG, Wierenga RK. Structure of a two-domain chitotriosidase

- from *Serratia marcescens* at 1.9 Å resolution. Proc Natl Acad Sci USA 2000;97:5842–5847.
23. Schiefner A, Angelov A, Liebl W, Skerra A. Structural basis for cellulose binding by the type A carbohydrate-binding module 64 of *Spirochaeta thermophila*. Proteins Struct Funct Bioinforma 2016;84: 855–858.
  24. Venditto I, Luis AS, Rydahl M, Schüchel J, Fernandes VO, Vidal-Melgosa S, Bule P, Goyal A, Pires VMR, Dourado CG, Ferreira LMA, Coutinho PM, Henrissat B, Knox JP, Baslé A, Najmudin S, Gilbert HJ, Willats WGT, Fontes C. Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. Proc Natl Acad Sci USA 2016;113:7136–7141.
  25. Simpson PJ, Xie H, Bolam DN, Gilbert HJ, Williamson MP. The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. J Biol Chem 2000;275:41137–41142.
  26. Boraston AB, Nurizzo D, Notenboom V, Ducros V, Rose DR, Kilburn DG, Davies GJ. Differential oligosaccharide recognition by evolutionarily-related  $\beta$ -1,4 and  $\beta$ -1,3 glucan-binding modules. J Mol Biol 2002;319:1143–1156.
  27. Simpson PJ, Jamieson SJ, Abou-hachem M, Karlsson EN, Gilbert HJ, Holst O, Williamson MP. The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase. Biochemistry 2002;5712–5719.
  28. Van Bueren AL, Morland C, Gilbert HJ, Boraston AB. Family 6 carbohydrate binding modules recognize the non-reducing end of  $\beta$ -1,3-linked glucans by presenting a unique ligand binding surface. J Biol Chem 2005;280:530–537.
  29. Van Bueren AL, Boraston AB. Binding sub-site dissection of a carbohydrate-binding module reveals the contribution of entropy to oligosaccharide recognition at “non- primary” binding subsites. J Mol Biol 2004;340:869–879.
  30. Correia MAS, Mazumder K, Brás JLA, Firbank SJ, Zhu Y, Lewis RJ, York WS, Fontes CMGA, Gilbert HJ. Structure and function of an arabinoxyylan-specific xylanase. J Biol Chem 2011;286:22510–22520.
  31. Carvalho AL, Goyal A, Prates JAM, Bolam DN, Gilbert HJ, Pires VMR, Ferreira LMA, Planas A, Romão MJ, Fontes CMGA. The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates  $\beta$ -1,4- and  $\beta$ -1,3–1,4-mixed linked glucans at a single binding site. J Biol Chem 2004;279:34785–34793.
  32. Szabó L, Jamal S, Xie H, Charnock SJ, Bolam DN, Gilbert HJ, Davies GJ. Structure of a family 15 carbohydrate-binding module in complex with xylopentaose: evidence that xylan binds in an approximate 3-fold helical conformation. J Biol Chem 2001;276:49061–49065.
  33. Notenboom V, Boraston AB, Chiu P, Freelove AC, Kilburn DG, Rose DR. Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. J Mol Biol 2001;314:797–806.
  34. Mikami B, Adachi M, Kage T, Sarikaya E, Nanmori T, Shinke R, Utsumi S. Structure of raw starch-digesting *Bacillus cereus*  $\beta$ -amylase complexed with maltose. Biochemistry 1999;38:7050–7061.
  35. Liu Y, Lai Y, Chou W, Chang MD, Lyu P. Solution structure of family 21 carbohydrate-binding module from *Rhizopus oryzae* glucoamylase. Biochem J 2007;403:21–30.
  36. Sainz-Polo MA, González B, Menéndez M, Pastor FJJ, Sanz-Aparicio J. Exploring multimodularity in plant cell wall deconstruction: structural and functional analysis of Xyn10C containing the CBM22–1-CBM22–2 tandem. J Biol Chem 2015;290:17116–17130.
  37. Najmudin S, Pinheiro BA, Prates JAM, Gilbert HJ, Romão MJ, Fontes CMGA. Putting an N-terminal end to the *Clostridium thermocellum* xylanase Xyn10B story: crystal structure of the CBM22–1-GH10 modules complexed with xylohexaose. J Struct Biol 2010;172: 353–362.
  38. Roske Y, Sunna A, Pfeil W, Heinemann U. High-resolution crystal structures of Caldicellulosiruptor strain Rt8B.4 carbohydrate-binding module CBM27–1 and its complex with mannohexaose. J Mol Biol 2004;340:543–554.
  39. Boraston AB, Revett TJ, Boraston CM, Nurizzo D, Davies GJ. Structural and thermodynamic dissection of specific mannan recognition by a carbohydrate binding module, TmCBM27. Structure 2003;11: 665–675.
  40. Tsukimoto K, Takada R, Araki Y, Suzuki K, Karita S, Wakagi T, Shoun H, Watanabe T, Fushinobu S. Recognition of cellooligosaccharides by a family 28 carbohydrate-binding module. FEBS Lett 2010;584:1205–1211.
  41. Jamal S, Nurizzo D, Boraston AB, Davies GJ. X-ray crystal structure of a non-crystalline cellulose-specific carbohydrate-binding module: CBM28. J Mol Biol 2004;339:253–258.
  42. Hashimoto H, Tamai Y, Okazaki F, Tamaru Y, Shimizu T, Araki T, Sato M. The first crystal structure of a family 31 carbohydrate-binding module with affinity to  $\beta$ -1,3-xylan. FEBS Lett 2005;579: 4324–4328.
  43. Kamitori S, Abe A, Ohtaki A, Kaji A, Tonozuka T, Sakano Y. Crystal structures and structural comparison of *Thermoactinomyces vulgaris* R-47  $\alpha$ -amylase 1 (TVAI) at 1.6 Å resolution and  $\alpha$ -amylase 2 (TVAII) at 2.3 Å resolution. J Mol Biol 2002;318:443–453.
  44. Hondoh H, Kuriki T, Matsuura Y. Three-dimensional structure and substrate binding of *Bacillus stearothermophilus* neopullulanase. J Mol Biol 2003;326:177–188.
  45. Montanier C, van Bueren AL, Dumon C, Flint JE, Correia M, Prates J, Firbank SJ, Lewis RJ, Grondin GG, Ghinet MG, Gloster TM, Herve C, Knox JP, Talbot BG, Turkenburg JP, Kerovuo J, Brzezinski R, Fontes CMGA, Davies GJ, Boraston, Gilbert HJ. Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proc Natl Acad Sci USA 2009;106:3065–3070.
  46. Suzuki N, Fujimoto Z, Kim YM, Momma M, Kishine N, Suzuki R, Suzuki S, Kitamura S, Kobayashi M, Kimura A, Funane K. Structural elucidation of the cyclization mechanism of  $\alpha$ -1,6-glucan by *Bacillus circulans* T-3040 cycloisomaltooligosaccharide glucanotransferase. J Biol Chem 2014;289:12040–12051.
  47. Jamal-Talabani S, Boraston AB, Turkenburg JP, Tarbouriech N, Ducros VMA, Davies GJ. Ab initio structure determination and functional characterization of CBM36: a new family of calcium-dependent carbohydrate binding modules. Structure 2004;12:1177–1187.
  48. Lammerts Van Bueren A, Finn R, Ausió J, Boraston AB.  $\alpha$ -Glucan recognition by a new family of carbohydrate-binding modules found primarily in bacterial pathogens. Biochemistry 2004;43: 15633–15642.
  49. Mikami B, Iwamoto H, Malle D, Yoon HJ, Demirkan-Sarikaya E, Mezaki Y, Katsuya Y. Crystal structure of pullulanase: evidence for parallel binding of oligosaccharides in the active site. J Mol Biol 2006;359:690–707.
  50. Najmudin S, Guerreiro CIPD, Carvalho AL, Prates JAM, Correia MAS, Alves VD, Ferreira LMA, Romão MJ, Gilbert HJ, Bolam DN, Fontes CMGA. Xyloglucan is recognized by carbohydrate-binding modules that interact with  $\beta$ -glucan chains. J Biol Chem 2006;281: 8815–8828.
  51. Hayashi M, Suzuki R, Colleoni C, Ball SG, Fujita N, Suzuki E. Bound substrate in the structure of cyanobacterial branching enzyme supports a new mechanistic model. J Biol Chem 2017;292: 5465–5475.
  52. Turkenburg JP, Brzozowski AM, Svendsen A, Borchert TV, Davies GJ, Wilson KS. Structure of a pullulanase from *Bacillus acidopullulyticus*. Proteins Struct Funct Bioinforma 2009;76:516–519.
  53. Koropatkin NM, Smith TJ. SusG: a unique cell-membrane-associated  $\alpha$ -amylase from a prominent human gut symbiont targets complex starch molecules. Structure 2010;18:200–215.
  54. Cid M, Pedersen HL, Kaneko S, Coutinho PM, Henrissat B, Willats WGT, Boraston AB. Recognition of the helical structure of  $\beta$ -1,4-galactan by a new family of carbohydrate-binding modules. J Biol Chem 2010;285:35999–36009.

55. Suits MDL, Pluvinage B, Law A, Liu Y, Palma AS, Chai W, Feizi T, Boraston AB. Conformational analysis of the *Streptococcus pneumoniae* hyaluronate lyase and characterization of its hyaluronan-specific carbohydrate-binding module. *J Biol Chem* 2014;289:27264–27277.
56. Notenboom V, Boraston AB, Kilburn DG, Rose DR. Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry* 2001;40:6248–6256.
57. Hoque MM, Suzuki K, Tsunoda M, Jiang J, Zhang F, Takahashi A, Ohbayashi N, Zhang X, Tanaka H, Omura S, Takénaka A. Structural insights into the specific anti-HIV property of actinohivin: structure of its complex with the  $\alpha(1\text{--}2)$ mannobiose moiety of gp120. *Acta Crystallogr Sect D Biol Crystallogr* 2012;68:1671–1679.
58. Notenboom V, Boraston AB, Williams SJ, Kilburn DG, Rose DR. High-resolution crystal structures of the lectin-like xylan binding domain from *Streptomyces lividans* xylanase 10A with bound substrates reveal a novel mode of xylan binding. *Biochemistry* 2002;41:4246–4254.
59. Fadel F, Zhao Y, Cousido-Siah A, Ruiz FX, Mitschler A, Podjarny A. X-ray crystal structure of the full length human chitotriosidase (CHIT1) reveals features of its chitin binding domain. *PLoS ONE* 2016;11:1–15.
60. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: a  $\beta$ -propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* 1995;3:1197–1205.
61. Abbott DW, Hryniuk S, Boraston AB. Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterocolitica*. *J Mol Biol* 2007;367:1023–1033.
62. Ficko-Blean E, Stuart CP, Suits MD, Cid M, Tessier M, Woods RJ, Boraston AB. Carbohydrate recognition by an architecturally complex  $\alpha$ -n-acetylglucosaminidase from *Clostridium perfringens*. *PLoS ONE* 2012;7:
63. Miyanaga A, Koseki T, Matsuzawa H, Wakagi T, Shoun H, Fushinobu S. Crystal structure of a family 54  $\alpha$ -L-arabinofuranosidase reveals a novel carbohydrate-binding module that can bind arabinose. *J Biol Chem* 2004;279:44907–44914.
64. Montanier CY, Correia MAS, Flint JE, Zhu Y, Baslé A, McKee LS, Prates JAM, Polizzi SJ, Coutinho PM, Lewis RJ, Henrissat B, Fontes CMGA, Gilbert HJ. A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization. *J Biol Chem* 2011;286:22499–22509.
65. Cuskin F, Flint JE, Gloster TM, Morland C, Baslé A, Henrissat B, Coutinho PM, Strazzulli A, Solovyova AS, Davies GJ, Gilbert HJ. How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc Natl Acad Sci USA* 2012;109:20889–20894.
66. Singh AK, Pluvinage B, Higgins MA, Dalia AB, Woodiga SA, Flynn M, Lloyd AR, Weiser JN, Stubbs KA, Boraston AB, King SJ. Unravelling the multiple functions of the architecturally intricate *Streptococcus pneumoniae*  $\beta$ -galactosidase, BgaA. *PLoS Pathog* 2014;10:e1004364.
67. Rodríguez-Sanoja R, Oviedo N, Escalante L, Ruiz B, Sánchez S. A single residue mutation abolishes attachment of the CBM26 starch-binding domain from *Lactobacillus amylovorus*  $\alpha$ -amylase. *J Ind Microbiol Biotechnol* 2009;36:341–346.
68. Bharali S, Purama R, Majumder A, Fontes C, Goyal A. Functional characterization and mutation analysis of family 11, carbohydrate-binding module (CtCBM11) of cellulosomal bifunctional cellulase from *Clostridium thermocellum*. *Indian J Microbiol* 2007;47:109–118.
69. Goto M, Semimaru T, Furukawa K, Hayashida S. Analysis of the raw starch-binding domain by mutation of a glucoamylase from *Aspergillus awamori* var. kawachi expressed in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 1994;60:3926–3930.
70. Gregg KJ, Finn R, Abbott DW, Boraston AB. Divergent modes of glycan recognition by a new family of carbohydrates-binding modules. *J Biol Chem* 2008;283:12604–12613.
71. Guillén D, Sánchez S, Rodríguez-Sanoja R. Carbohydrate-binding domains : multiplicity of biological roles. *Appl Microbiol Biotechnol* 2010;85:1241–1249.
72. Grondin JM, Chitayat S, Ficko-Blean E, Houliston S, Arrowsmith CH, Boraston AB, Smith SP. An unusual mode of galactose recognition by a family 32 carbohydrate-binding module. *J Mol Biol* 2014;426:869–880.
73. Gaulin E, Drame N, Lafitte C, Torto-Alalibo T, Martinez Y, Ameline-Torregrosa C, Khatib M, Mazarguil H, Villalba-Mateos F, Kamoun S, Mazars C, Dumas B, Bottin A, Esquerre-Tugaye MT, Rickauer M. Cellulose binding domains of a phytophthora cell wall protein are novel pathogen-associated molecular patterns. *Plant Cell* 2006;18:1766–1777.
74. Blake AW, McCartney L, Flint JE, Bolam DN, Boraston AB, Gilbert HJ, Knox P. Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *J Biol Chem* 2006;281:29321–29329.
75. Kim JS, Awano T, Yoshinaga A, Takabe K. Immunolocalization and structural variations in differentiating earlywood tracheid cell walls of *Cryptomeria japonica*. *Planta* 2010;232:817–824.
76. Walker JA, Takasuka TE, Deng K, Bianchetti CM, Udell HS, Prom BM, Kim H, Adams PD, Northen TR, Fox B. Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules. *Biotechnol Biofuels* 2015;8:220.
77. Mangala SL, Kittur FS, Nishimoto M, Sakka K, Ohmiya K, Kitaoka M, Hayashi K. Fusion of family VI cellulose binding domains to *Bacillus halodurans* xylanase increases its catalytic activity and substrate-binding capacity to insoluble xylan. *J Mol Catal B: Enzym* 2003;21:221–230.
78. Reyes-ortiz V, Heins RA, Cheng G, Kim EY, Vernon BC, Elandt RB, Adams PD, Sale KL, Hadi MZ, Simmons BA, Kent MS, Tullman-Ercek D. Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates. *Biotechnol Biofuels* 2013;6:93.
79. Mello BL, Polikarpov I. Family 1 carbohydrate binding-modules enhance saccharification rates. *AMB Express* 2014;4:36.
80. Furukawa T, Sawaguchi C, Watanabe A, Takahashi M, Nigorikawa M, Furukawa K, Limura Y, Kajita S, Oguchi T, Ito Y, Sonoki T. Application of fungal laccase fused with cellulose-binding domain to develop low-lignin rice plants. *J Biosci Bioeng* 2013;116:616–619.
81. Grisolia MJ, Peralta DA, Valdez HA, Barchiesi J, Gomez-casati DF, Busi MV. The targeting of starch binding domains from starch synthase III to the cell wall alters cell wall composition and properties. *Plant Mol Biol* 2016;93:121–135.
82. Shi X, Zheng F, Pan R, Wang J, Ding S. Engineering and comparative characteristics of double carbohydrate binding modules as a strength additive for papermaking applications. *BioResources* 2014;9:3117–3131.
83. Rivkin A, Abitbol T, Nevo Y, Verker R, Lapidot S, Komarov A, Veldhuis SC, Ziberman G, Reches M, Cranston ED, Shoseyov O. Bionanocomposite films from resilin-CBD bound to cellulose nanocrystals. *Ind Biotechnol* 2015;11:44–58.
84. Espigares I, Elvira C, Mano JF, Vázquez B, San Román J, Reis RL. New partially degradable and bioactive acrylic bone cements based on starch blends and ceramic fillers. *Biomaterials* 2002;23:1883–1895.
85. Michailova V, Titeva S, Kotsilkova R, Krusteva E, Minkov E. Influence of hydrogel structure on the processes of water penetration and drug release from mixed hydroxypropylmethyl cellulose/thermally pregelatinized waxy maize starch hydrophilic matrices. *Int J Pharm* 2001;222:7–17.
86. Levy I, Paldi T, Shoseyov O. Engineering a bifunctional starch-cellulose cross-bridge protein. *Biomaterials* 2004;25:1841–1849.



87. Andrade FK, Costa R, Domingues L, Soares R, Gama M. Improving bacterial cellulose for blood vessel replacement : functionalization with a chimeric protein containing a cellulose-binding module and an adhesion peptide. *Acta Biomater* 2010;6:4034–4041.
88. Ramos R, Domingues L, Gama M. *Escherichia coli* expression and purification of LL37 fused to a family III carbohydrate-binding module from *Clostridium thermocellum*. *Protein Expr Purif* 2010;71:1–7.
89. Sugimoto N, Igarashi K, Samejima M. Cellulose affinity purification of fusion proteins tagged with fungal family 1 cellulose-binding domain. *Protein Expr Purif* 2012;82:290–296.
90. Lin S, Lin I, Chou W, Hsieh C, Liu S, Huang R, Sheu C, Chang M. CBM21 starch-binding domain: a new purification tag for recombinant protein engineering. *Protein Expr Purif* 2009;65:261–266.
91. Guillén D, Moreno-Mendieta S, Aguilera P, Sánchez S, Farres A, Rodríguez-Sanoja R. The starch-binding domain as a tool for recombinant protein purification. *Appl Microbiol Biotechnol* 2013;97:4141–4148.
92. Lu L, Xu S, Zhao R, Zhang D, Li Z, Li Y, Xiao M. Synthesis of galactooligosaccharides by CBD fusion  $\beta$ -galactosidase immobilized on cellulose. *Bioresour Technol* 2012;116:327–333.
93. Torres DPM, Goncalves PF, Teixeira JA, Rodrigues L. Galacto-oligosaccharides: production, properties, applications, and significance as prebiotics. *Compr Rev Food Sci Food Saf* 2010;9:438–454.
94. Moreno-mendieta SA, Guillén D, Espitia C, Hernández-pando R, Sanchez S, Rodríguez-sanoja R. A novel antigen-carrier system: the *Mycobacterium tuberculosis* Acr protein carried by raw starch micro-particles. *Int J Pharma* 2014;474:241–248.
95. Guillén D, Moreno-mendieta S, Pérez R, Espitia C, Sánchez S, Rodríguez-sanoja R. Starch granules as a vehicle for the oral administration of immobilized antigens. *Carbohydr Polym* 2014;112:210–215.
96. McCartney L, Gilbert HJ, Bolam DN, Boraston AB, Knox JP. Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. *Anal Biochem* 2004;326:49–54.
97. Skerra A. Engineered protein scaffolds for molecular recognition. *J Mol Recognit* 2000;13:167–187.
98. Nygren PÅ, Skerra A. Binding proteins from alternative scaffolds. *J Immunol Method* 2004;290:3–28.
99. Skerra A. Alternative non-antibody scaffolds for molecular recognition. *Curr Opin Biotechnol* 2007;18:295–304.
100. Wiederstein M, Sippl MJ. Protein sequence randomization: efficient estimation of protein stability using knowledge-based potentials. *J Mol Biol* 2005;345:1199–1212.
101. Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH. Evolving strategies for enzyme engineering. *Curr Opin Struct Biol* 2005;15:447–452.
102. Lo W-C, Lyu P-C. CPSARST: an efficient circular permutation search tool applied to the detection of novel protein structural relationships. *Genome Biol* 2008;9:R11.
103. Stephen P, Tseng K-L, Liu Y-N, Lyu P-C. Circular permutation of the starch-binding domain: inversion of ligand selectivity with increased affinity. *Chem Commun* 2012;48:2612–2614.
104. Yang H, Li J, Shin H-D, Du G, Liu L, Chen J. Molecular engineering of industrial enzymes: recent advances and future prospects. *Appl Microbiol Biotechnol* 2014;98:23–29.
105. Linder M, Lindeberg G, Reinikainen T, Teeri TT, Pettersson G. The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. *FEBS Lett* 1995;372:96–98.
106. Chang H, Irwin PM, Nikolov ZL. Effects of mutations in the starch-binding domain of *Bacillus macerans* cyclodextrin glycosyl-transferase. *J Biotechnol* 1998;65:191–202.
107. Zhang Y, Wang L, Chen J, Wu J. Enhanced activity toward PET by site-directed mutagenesis of *Thermobifida fusca* cutinase-CBM fusion protein. *Carbohydr Polym* 2013;97:124–129.
108. Arunachalam TS, Wichert C, Appel B, Müller S. Mixed oligonucleotides for random mutagenesis: best way of making them. *Org Biomol Chem* 2012;10:4641.
109. Smith GP, Patel SU, Windass JD, Thornton JM, Winter G, Griffiths D. Small binding proteins selected from a combinatorial repertoire of knottins displayed on phage. *J Mol Biol* 1998;277:317–332.
110. Lehtiö J, Teeri TT, Nygren P. a. Alpha-amylase inhibitors selected from a combinatorial library of a cellulose binding domain scaffold. *Proteins* 2000;41:316–322.
111. Wernéus H, Lehtiö J, Teeri T, Nygren PÅ, Ståhl S. Generation of metal-binding staphylococci through surface display of combinatorially engineered cellulose-binding domains. *Appl Environ Microbiol* 2001;67:4678–4684.
112. Gunnarsson LC, Karlsson EN, Albrekt AS, Andersson M, Holst O, Ohlin M. A carbohydrate binding module as a diversity-carrying scaffold. *Protein Eng Des Sel* 2004;17:213–221.
113. Shiraga S, Kawakami M, Ueda M. Construction of combinatorial library of starch-binding domain of *Rhizopus oryzae* glucoamylase and screening of clones with enhanced activity by yeast display method. *J Mol Catal B: Enzym* 2004;28:229–234.
114. Gunnarsson LC, Montanier C, Tunnicliffe RB, Williamson MP, Gilbert HJ, Karlsson EN, Ohlin M. Novel xylan-binding properties of an engineered family 4 carbohydrate-binding module. *Biochem J* 2007;214:209–214.
115. Gunnarsson LC, Dexlin L, Nordberg E, Holst O, Ohlin M. Evolution of a carbohydrate binding module into a protein-specific binder. *Biomol Eng* 2006;23:111–117.
116. McCullum EO, Williams B, Zhang J, Chaput JC. Random mutagenesis by error-prone PCR. In: Braman L, editor. *In vitro mutagenesis protocols*, 3 ed. *Methods in Molecular Biology*; 2010. pp 103–110.
117. Schantz LV, Gullfot F, Scheer S, Filonova L, Gunnarsson LC, Flint JE, Daniel G, Karlsson EN, Brumer H, Ohlin M. Affinity maturation generates greatly improved xyloglucan-specific carbohydrate binding modules. *BMC Biotechnol* 2009;9:92.
118. Mattheakis LC, Bhatt RR, Dower WJ. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Biochemistry* 1994;91:9022–9026.
119. Hanes J, Plückthun A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Aca Sci USA* 1997;94:4937–4942.
120. Nemoto N, Miyamoto-Sato E, Husimi Y, Yanagawa H. In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett* 1997;414:405–408.
121. Roberts RW, Szostak JW. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Biochemistry* 1997;94:12297–12302.
122. Francisco JA, Campbell R, Iversont BL, Georgiou G. Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. *PNAS* 1993;90:10444–10448.
123. Boder ET, Wittrup KD. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* 1997;15:553–557.
124. Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 1985;228:1315–1317.
125. Wentz AE, Shusta EV. A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl Environ Microbiol* 2007;73:1189–1198.
126. Mattinen ML, Kontteli M, Kerovuo J, Linder M, Annala A, Lindeberg G, Reinikainen T, Drakenberg T. Three-dimensional structures of three engineered cellulose-binding domains of cellobiohydrolase I from *Trichoderma reesei*. *Protein Sci* 1997;6:294–303.
127. Schantz LV, Håkansson M, Logan DT, Walse B, Österlin J, Nordberg-karlsson E, Ohlin M. Structural basis for carbohydrate-



- binding specificity—a comparative assessment of two engineered carbohydrate-binding modules. *Glycobiology* 2012;22:948–961.
128. Schantz LV, Hakansson M, Logan DT, Nordberg-karlsson E, Ohlin M. Carbohydrate binding module recognition of xyloglucan defined by polar contacts with branching xyloses and CH- $\pi$  interactions. *Proteins Struct Funct Bioinforma* 2014;82:3466–3475.
  129. Fisher SZ, Von Schantz L, Hakansson M, Logan DT, Ohlin M. Neutron crystallographic studies reveal hydrogen bond and water-mediated interactions between a carbohydrate-binding module and its bound carbohydrate ligand. *Biochemistry* 2015;54:6435–6438.
  130. Gunnarsson LC, Zhou Q, Montanier C, Karlsson EN, Brumer H, Ohlin M. Engineered xyloglucan specificity in a carbohydrate-binding module. *Glycobiology* 2006;16:1171–1180.
  131. Gullfot F, Tan T, Schantz LV, Karlsson EN, Ohlin M, Brumer H, Divne C. The crystal structure of XG-34, an evolved xyloglucan-specific carbohydrate-binding module. *Struct Note* 2009;785–789.
  132. Sakata T, Takakura J, Miyakubo H, Osada Y, Wada R, Takahashi H, Yatsunami R, Fukui T, Nakamura S. Improvement of binding activity of xylan-binding domain by amino acid substitution. *Nucleic Acids Symp Ser* 2004;253–254.