

Caracterización bioinformática y funcional de una enzima bacteriana aislada del tracto digestivo de larvas de *Diatraea saccharalis* (Lepidoptera: Crambidae) para la degradación de la celulosa.

Harold Hernando Galvis Pérez

Trabajo de Grado para optar al Título de Biólogo

Director

Edinson Andrés Puentes Cala

Dr. rer. nat.

Codirector

Jorge Hernández Torres

Doctorat ès Sciences

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Biología

Bucaramanga

2025

Dedicatoria

Este trabajo lo dedico al verdadero amor de mi vida. A ese por el que hoy vivo y alcanzo este logro, el amor que me ha movido desde que nací y que me acompañará hasta mi partida. Este logro es para ti, madre, que me has enseñado el valor del trabajo y la dedicación, que aún en la adversidad le haces honor a tu nombre y me das a mí también un poco de ella, Esperanza. Todos mis logros serán siempre por ti y para ti, madre. Te amo.

Este trabajo es para ti, viejito lindo. Sin ti no me habría decidido a adentrarme en este camino científico tan espectacular. Me alentaste y me respaldaste en mi decisión de ser un biólogo loco, uno que se cuestionara y fascinara por todo, que buscara la razón de las cosas y compartiera los hallazgos con los demás. Gracias por abrirme las puertas de la ciencia y cultivarme el inmenso amor por el conocimiento que ahora tengo, cosas que me acompañarán toda la vida, y que estoy seguro me han abierto la mente para continuar estudiando la belleza de la vida.

Y finalmente, todo esto es para ti, Mercurio, que te levantaste una y otra vez, y lo intentaste hasta más allá del límite. Solo nosotros sabemos lo que nos ha costado, y hoy lo estamos logrando. Hemos culminado, y lo hemos hecho bien. ¡Felicidades!

Agradecimientos

Al iBGA y a la CIC por confiar en mí y en mis habilidades para llevar a cabo con éxito este proyecto.

Al Dr. Edinson, su guía, retroalimentación, comprensión y carisma hicieron de este un camino enriquecedor y de mucho aprendizaje. Gracias por creer en mí y acompañarme en esta etapa de mi vida académica.

A mi hermosa familia, por siempre respaldarme y animarme a culminar este viaje de la mejor forma. Sin su apoyo y compañía no hubiese tenido las herramientas para llegar a la meta.

A todas mis tostiarepitas, gracias por tan maravillosos momentos, los llevaré siempre en mi mente y mi corazón. ¡Que aburrida hubiese sido esta carrera sin ustedes!

Al profesor Jorge por su acompañamiento y guía.

A Wendy, mi linda compañera, sin ella nuestros trabajos de grado no tendrían el mérito que tienen ahora. ¡Gracias por todo, cosas increíbles nos esperan!

A todo el personal del comedor estudiantil y bienestar universitario, la gran labor que hacen fue indispensable para poder sacar adelante esta investigación.

A Julieth y Valentina, por sus valiosos aportes a la ejecución de este trabajo.

A todas las personas que de alguna u otra forma hicieron esto posible, muchas gracias, mi cariño siempre será para ustedes.

Tabla de Contenido

	Pág.
Introducción _____	10
1. Marco teórico _____	12
1.1 Las celulasas y la degradación de la celulosa _____	12
1.2 Caracterización de celulasas _____	14
1.3 Algunos aspectos relevantes en la producción de celulasas _____	17
2. Objetivos _____	19
2.1 Objetivo General _____	19
2.2 Objetivos Específicos _____	19
3. Metodología _____	20
3.1 Identificación de la actividad celulasa en la cepa nativa _____	20
3.2 Determinación cualitativa de la actividad celulasa <i>in vitro</i> _____	20
3.3 Métodos analíticos _____	21
3.4 Optimización del ensayo y caracterización de las celulasas _____	23
3.5 Análisis bioinformático de secuencias nucleotídicas y aminoacídicas _____	24
3.5.1 Análisis bioinformático de secuencias nucleotídicas _____	24
3.5.2 Análisis bioinformático de secuencias aminoacídicas _____	26
4. Resultados _____	27
4.1 Identificación de la actividad celulasa en la cepa nativa _____	27
4.2 Comprobación cualitativa y cuantificación de la actividad celulasa _____	28
4.3 Optimización de catálisis y caracterización fisicoquímica de las celulasas _____	33
4.4 Análisis bioinformático de secuencias nucleotídicas y aminoacídicas _____	39

4.4.1 Análisis bioinformático de secuencias nucleotídicas _____	39
4.4.2 Análisis bioinformático de secuencias aminoacídicas _____	48
5. Discusión _____	56
6. Conclusiones _____	64
Referencias Bibliográficas _____	66

Lista de Tablas

	Pág.
Tabla 1. _____	32
Tabla 2. _____	40
Tabla 3. _____	41
Tabla 4. _____	42
Tabla 5. _____	44
Tabla 6. _____	46
Tabla 7. _____	47
Tabla 8. _____	49
Tabla 9. _____	51
Tabla 10. _____	52

Lista de Figuras

	Pág.
Figura 1. _____	14
Figura 2. _____	28
Figura 3. _____	30
Figura 4. _____	31
Figura 5. _____	34
Figura 6. _____	35
Figura 7. _____	36
Figura 8. _____	37
Figura 9. _____	38
Figura 10. _____	39
Figura 11. _____	45
Figura 12. _____	55

Resumen

Título: Caracterización bioinformática y funcional de una enzima bacteriana aislada del tracto digestivo de larvas de *Diatraea saccharalis* (Lepidoptera: Crambidae) para la degradación de la celulosa.*

Autor: Galvis-Pérez, Harold Hernando. **

Palabras Clave: Celulasa, Caracterización *in silico*, Caracterización funcional, Extractos crudos, Celulosa, Glucosa.

La celulosa es un polímero abundante y altamente resistente a la degradación. Su degradación a carbohidratos pequeños con alto valor comercial ha sido un reto complejo de abordar y las enzimas se presentan como una solución. Las celulasas son enzimas que catalizan la degradación de celulosa a glucosa, haciendo de esta cualidad hidrolítica una de particular demanda. La búsqueda de biocatalizadores aplicables al contexto científico e industrial ha conducido al hallazgo de cepas bacterianas nativas con potencial biotecnológico aprovechable. En este trabajo se llevó a cabo una caracterización bioinformática y funcional de dos celulasas aisladas de la cepa C1 de *Microbacterium*, recuperada del intestino del barrenador de la caña. Los métodos incluyeron el aislamiento y cultivo de la cepa en medio estándar, la obtención de extractos crudos del cultivo, así como de la proteína junto con su cinética enzimática, la identificación de los potenciales genes codificantes de la enzima, la caracterización *in silico* de sus propiedades fisicoquímicas y el modelamiento 3D de su estructura terciaria. Dos probables endoglucanasas de las familias GH5 y GH9 fueron identificadas en el genoma de la cepa C1, cuyo análisis de secuencias *in silico* corroboró la calidad de celulasa termófila y levemente acidófila de la enzima caracterizada experimentalmente. Características adicionales como el rápido alcance de su pico de actividad y su alta afinidad a la carboximetilcelulosa la hacen atractiva biotecnológicamente. La predicción *in silico* de su estructura terciaria permitió aportar evidencia a la asignación de estas probables enzimas a las familias de endoglucanasas GH5 y GH9. Estos resultados demuestran el potencial biotecnológico de la celulasa aislada de C1 y evidencian la ventaja de implementar aproximaciones experimentales e *in silico* para la caracterización de biocatalizadores derivados de bioprospección con el potencial de ser soluciones innovadoras en Colombia.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Biología. Biología. Director: Edinson Andrés Puentes Cala. Dr. rer. nat. Codirector: Jorge Hernández Torres. Doctorat ès Sciences.

Abstract

Title: Bioinformatic and functional characterization of a bacterial enzyme isolated from *Diatraea saccharalis* (Lepidoptera: Crambidae) digestive tract for cellulose degradation. *

Author(s): Galvis-Pérez, Harold Hernando. **

Key words: Cellulase, *In silico* characterization, Functional characterization, Crude extracts, Cellulose, Glucose.

Cellulose is an abundant and highly degradation-resistant polymer. Its degradation into smaller carbohydrates with high commercial value has been a complex challenge to address, and enzymes provide a solution. Cellulases are enzymes that catalyze the degradation of cellulose to glucose, thus representing a hydrolytic attribute with a high demand. The search for biocatalysts applicable to scientific and industrial areas has led to the discovery of new native bacterial strains with promising biotechnological potential. In this study, the bioinformatic and functional characterization of two cellulases isolated from *Microbacterium* C1 strain, recovered from the gut of the sugarcane borer, was performed. Methods included strain isolation and culture on standard medium, crude extract obtention, as well as protein characterization along with its enzyme kinetics, identification of potential enzyme-coding genes, *in silico* characterization of physicochemical properties and 3D modeling of tertiary structures. Two putative GH5 and GH9 endoglucanases were identified within the genome of C1 strain, which *in silico* sequence analysis corroborated thermophilic and mild acidophilic attributes of the enzyme determined by wet lab. Additional features, such as its rapid activity peak and high affinity for carboxymethylcellulose, make it a biotechnologically attractive enzyme. *In silico* prediction of tertiary structure allowed for the assignment of GH5 y GH9 endoglucanase families to these putative enzymes. These results demonstrate the biotechnological potential of C1-isolated cellulase and provide evidence of the advantages of applying wet lab and *in silico* approaches for bioprospection-derived biocatalysts with potential to offer innovative solutions in Colombia.

* Degree Work

** Faculty of Sciences. Biology School. Biology. Director: Edinson Andrés Puentes Cala. Dr. rer. nat. Codirector: Jorge Hernández Torres. Doctorat ès Sciences.

Introducción

Las enzimas son catalizadores biológicos (biocatalizadores) que reducen la energía de activación requerida para que una reacción bioquímica se lleve a cabo. Estos biocatalizadores son de naturaleza proteica y se caracterizan por mantener una alta especificidad por sus respectivos sustratos. Dada su naturaleza catalítica, las enzimas aceleran reacciones bioquímicas sin consumirse en el proceso y son requeridas en pequeñas cantidades. Las enzimas son cruciales para el metabolismo y la adaptación de los organismos. Al ser agentes promotores de diversidad biológica, su presencia o ausencia puede influir en el *fitness*, haciéndolas fundamentales para las futuras adaptaciones que garanticen la vida (Peracchi, 2018; Robinson, 2015). Las enzimas pueden ser extraídas de las células para ser aplicadas en una vasta variedad de procesos industriales y comerciales, incluyendo su aplicación en el campo farmacéutico y ambiental (Robinson, 2015). La aplicación de enzimas desempeña un papel crucial en la optimización y facilitación de numerosos procesos industriales contemporáneos. Dentro de los tipos de enzimas, las hidrolasas juegan un papel crucial al ser líticas de compuestos, y dentro de ellas, las glicosil hidrolasas destacan por descomponer grandes polisacáridos en oligómeros glucídicos. De particular interés, las celulasas facilitan la hidrólisis de la celulosa desde distintos flancos en su acción catalítica. La celulosa, uno de los compuestos orgánicos más abundantes y resistentes en la naturaleza, al ser degradada enzimáticamente, libera moléculas de glucosa de gran valor para diversas aplicaciones industriales (Bhat, 2000; Singhanía et al., 2021; Xing et al., 2012). La transformación de esta materia prima mediante el uso de celulasas se ha consolidado como una estrategia biotecnológica eficaz para acelerar este proceso. Sin embargo, el aislamiento y la caracterización de estas enzimas representan desafíos significativos, demandando una inversión de tiempo y recursos considerable.

Para la aplicación industrial de enzimas se requiere una caracterización bioquímica rigurosa que asegure su rendimiento óptimo. Con el descubrimiento de enzimas celulolíticas bacterianas surge el interés de caracterizar sus propiedades fisicoquímicas, estructurales y funcionales. La caracterización empieza con la identificación del gen responsable de su expresión. El uso de herramientas bioinformáticas permite la identificación de funciones de genes y proteínas desconocidos (Teng et al., 2020). Este paso es crucial para estudios posteriores de actividad enzimática, y si es el caso, mejorar las propiedades de la enzima, optimizar su funcionamiento, producir enzimas fusionadas o quiméricas (Chandel et al., 2012), entre otras posibilidades.

En el proceso de búsqueda de nuevas enzimas, se detectan isoformas responsables de la misma actividad enzimática y la selección del gen con el mejor desempeño requiere de evaluaciones bioquímicas meticulosas. Aunque las predicciones bioinformáticas pueden acelerar el proceso de selección de enzimas candidatas, estas simulaciones deben ser corroboradas *in vitro*. Algunas pruebas bioquímicas que se realizan son la temperatura y pH óptimos, búsqueda de caracteres extremófilos y otras características relevantes como las actividades enzimática y específica (Cerdeja-Mejía, 2016). Estas evaluaciones permiten estimar el potencial de la enzima como biocatalizador en aplicaciones industriales específicas.

La Corporación para la Investigación de la Corrosión (CIC) y el Grupo de Investigación en Biotecnología Industrial y Biología Molecular (CINBIN) (Atehortúa-Bueno, 2022), lograron aislar genes candidatos a productores de celulasas de microorganismos presentes en el intestino del gusano barrenador de la caña de azúcar, *Diatraea saccharalis* (Puentes-Cala et al., 2023). El objetivo de este trabajo es identificar y someter a una caracterización bioquímica preliminar una celulasa aislada de la cepa C1. Los métodos implementados incluyen el análisis bioinformático y la caracterización bioquímica de la enzima en extractos crudos.

1. Marco teórico

1.1 Las celulasas y la degradación de la celulosa

Las celulasas son enzimas hidrolíticas que hacen parte de la familia de las glicosil hidrolasas, las cuales degradan oligosacáridos y polisacáridos en monosacáridos. Se consideran enzimas celulasas a las endoglucanasas (EC 3.2.1.4), exoglucanasas o celobiohidrolasas (EC 3.2.1.91) y celobiasas o β -glucosidasas (EC 3.2.1.21), las cuales actúan en sinergia para transformar la celulosa en monómeros de glucosa (Gray et al., 2006; Kuhad et al., 2011). Las endoglucanasas hidrolizan los enlaces β -1,4-glucosídicos internos o de regiones amorfas de la celulosa, produciendo unidades de polisacáridos de celulosa. Las exocelulasas digieren las unidades de polisacáridos en disacáridos de celobiosa en sus extremos, y, finalmente, las celobiasas digieren celooligosacáridos como la celobiosa en unidades monoméricas de glucosa (Chandel et al., 2012; Dantur et al., 2015; Li, C. et al., 2016; Singhanian et al., 2021). La producción biológica de estas enzimas se encuentra ampliamente distribuida en microorganismos como bacterias y hongos (Ogura et al., 2006), aunque también pueden ser producidas por rumiantes y plantas (Park, Y. W. et al., 2003; Teunissen et al., 1992).

Estas enzimas presentan una amplia diversidad, incluyendo la de la conformación estructural. Dentro de las endoglucanasas, por ejemplo, se han registrado distintas formas del sitio activo de la proteína, siendo algunas de ellas el plegamiento de barril (α/β)₈, barril (β/α) modificado, *β -jelly roll* y barril (α/α)₆ (Alzari et al., 1996; Davies & Henrissat, 1995; Okano et al., 2015; Sakon et al., 1997; Yennamalli et al., 2011). Estas variaciones adaptadas a la forma fibrilar de la celulosa pueden provocar diferencias en propiedades funcionales de interés como la procesividad, que hace referencia a cuanto sustrato puede hidrolizar la enzima sin desprenderse de él (Liu et al., 2021). Sin embargo, esta familia de proteínas cuenta con sitios conservados como su

región catalítica y la geometría alrededor del enlace glicosídico (Park, S. et al., 2017) siendo esto potencialmente útil en la identificación de regiones importantes en las secuencias nucleotídicas y aminoacídicas.

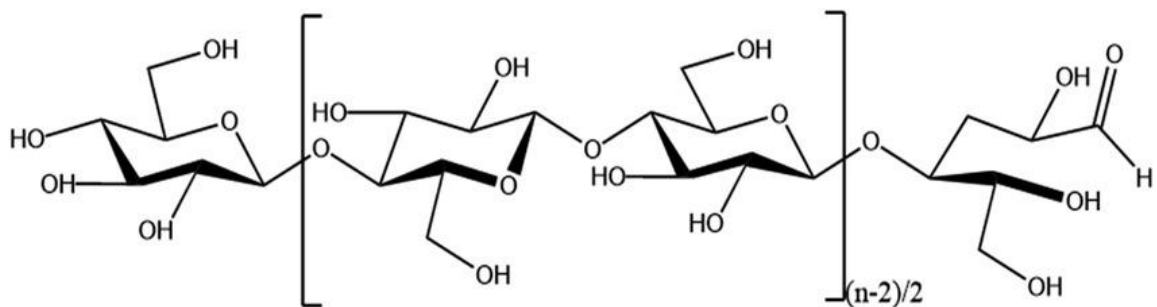
El sustrato de estas enzimas, la celulosa, es un polisacárido conformado por monómeros de glucosa unidos mediante enlaces β -1,4 glicosídicos (Fig. 1) que hacen de este compuesto uno altamente cristalino, compacto y resistente a la degradación (Gray et al., 2006). Este biopolímero es abundante en la naturaleza y su despolimerización en monómeros de glucosa tiene amplios usos industriales, entre ellos el de la fermentación alcohólica para la producción de bioetanol, tratamiento de textiles, industria de alimentos, entre otros. Por su alto impacto industrial y comercial, la digestión enzimática de fibras de celulosa se ha venido investigando en las últimas décadas. Se ha encontrado que este compuesto cristalino presenta irregularidades en su superficie a través de las cuales, moléculas de agua permean las fibras haciéndolas crecer en tamaño cuando se hidratan. Con esto, los microporos y cavidades circundantes a las fibras permiten el paso de grandes moléculas, incluyendo enzimas (Sukumaran et al., 2005). Este comportamiento se ha aprovechado para alcanzar una mayor superficie de contacto entre las celulasas y su sustrato, logrando reducir el tiempo de tratamiento.

No obstante, las celulasas pueden no alcanzar siempre su máximo rendimiento debido a la regulación genética y bioquímica, que incluye la inducción, inhibición del producto final y represión de catabolitos (Kumar et al., 2008). Este sistema de regulación presenta el desafío de los inductores, lo que implica que su respuesta catalítica fluctúa en función de la cantidad de sustrato y producto disponible. Es decir, ante una gran cantidad de celulosa, la liberación de glucosa incrementa, sin embargo, el exceso de glucosa se unirá a enzimas libres inhibiendo su actividad (Liu et al., 2021) y consecuentemente reduciendo la digestión de celulosa. Factores como este

sistema de regulación, las condiciones a las que la enzima esté adaptada y las circunstancias al momento de su expresión y producción, pueden influir en la calidad degradativa de las celulasas y se convierten en elementos de interés a caracterizar.

Figura 1.

Representación estructural de los homodímeros que conforman las fibras de celulosa



Nota. Adaptado de “Preparation of nanocellulose and application of nanocellulose polyurethane composites”, por Y. Mo, X. Huang, M. Yue, L. Hu y C. Hu, 2024, RSC Advances, 14, 18247–18257. © The Royal Society of Chemistry.

1.2 Caracterización de celulasas

La determinación de parámetros clave como el pH y la temperatura óptimos de actividad, así como constantes cinéticas fundamentales (K_m , V_{max} , entre otros), constituye una etapa esencial en la investigación de las celulasas y en la evaluación de su potencial para aplicaciones industriales. Estos análisis pueden abordarse desde dos perspectivas complementarias. En un enfoque bioinformático se predicen las propiedades bioquímicas y estructurales a partir de la secuencia, mientras que, en un enfoque bioquímico, se determina experimentalmente su comportamiento catalítico.

La caracterización *in silico* inicia con el análisis de las secuencias nucleotídicas y aminoacídicas obtenidas con herramientas como BLAST (Altschul et al., 1990) y portales de anotación funcional como ExPASy (Gasteiger, E. et al., 2003). La predicción de estructuras secundarias en celulasas se ha realizado con SOPMA (Geourjon & Deléage, 1995) y PSIPRED (McGuffin et al., 2000), mientras que para el modelado tridimensional se han utilizado herramientas como SWISS-MODEL (Waterhouse et al., 2018), RaptorX-Property (Wang et al., 2016), LOMETS (Wu, S. & Zhang, 2007) e I-TASSER (Zhang, Y., 2008). Los recientes avances en inteligencia artificial han incrementado la precisión de las predicciones estructurales, destacando la herramienta AlphaFold (Jumper et al., 2021). Refinamientos adicionales pueden realizarse con ModRefiner (Xu, D. & Zhang, 2011) o HyperChem (Froimowitz, 1993) para optimizar las geometrías moleculares.

La predicción de parámetros fisicoquímicos mediante herramientas como ProtParam (Gasteiger, Elisabeth et al., 2005) permite estimar propiedades como el índice alifático (AI), el punto isoeléctrico (pI), el coeficiente de extinción (EC), la distribución de cargas, entre otros (Hoda et al., 2021). En etapas posteriores, simulaciones de acoplamiento molecular con AutoDock (Morris et al., 2009) pueden evaluar la afinidad por sustratos, cofactores e inhibidores potenciales (Palma, 2012).

Desde la bioquímica, la caracterización experimental se inicia con la obtención de extractos enzimáticos, ya sea a partir de cultivos nativos o mediante expresión heteróloga en sistemas como *Escherichia coli* o *Pichia pastoris*. Estos extractos, también denominados extractos crudos o celulares, son obtenidos directamente del cultivo bacteriano mediante un protocolo estándar que generalmente se compone del cultivo, lisis, centrifugación, y separación de fracciones, haciendo que en sus fases resultantes se hallen el catalizador, así como sales, aminoácidos, cofactores, ADN

o ARN, entre otros componentes, sin la presencia de células funcionales (*cell-free*) (Kwon & Jewett, 2015).

Los ensayos de actividad suelen emplear sustratos modelo como carboximetilcelulosa (CMC) para endoglucanasas, papel filtro para exoglucanasas y *p*-nitrofenil- β -D-glucopiranosido para β -glucosidasas. La liberación de azúcares reductores se cuantifica mediante el ensayo DNS (ácido 3,5-dinitrosalicílico) (Miller, 1959), basado en la reducción del DNS por los grupos carbonilo de los azúcares liberados de la hidrólisis de celulosa, formando 3-amino-5-nitrosalicilato. Este compuesto aromático presenta coloración naranja con pico de absorbancia a 540 nm (Başkan et al., 2016). Alternativamente, la actividad celulasa puede ser detectada mediante ensayos cualitativos como la tinción con Rojo Congo.

La determinación de pH y temperatura óptimos se realiza incubando la enzima con el sustrato, en rangos amplios de, por ejemplo, pH 3 –11 y de 20 – 90 °C, para establecer la estabilidad catalítica. Algunos puntos de pH óptimo que se han hallado para distintas celulasas de distintos organismos han sido: 4.0 para *Synechococcus* sp. (Kim, Soo-Jin et al., 2008), 7.0 para *Cellulomonas* sp. (Yin et al., 2010) y 10.0 para *Aspergillus flavus* (Bano et al., 2019), siendo esta última una celulasa alcalófila. Para el caso de la temperatura, se han reportado celulasas con óptimos de 40 °C para cepas de *Aspergillus niger* (Sulyman et al., 2020), 60 °C para *Bacillus subtilis* (Malik & Javed, 2021) y *Aspergillus flavus* (Bano et al., 2019), considerándose estas dos últimas como celulasas termófilas.

Asimismo, se evalúa la influencia de iones metálicos u otras moléculas que pueden actuar como cofactores. Cationes como Ca^{2+} , Ni^{2+} , Mn^{2+} , y Pb^{2+} han demostrado incrementar la actividad celulasa, mientras que otros como Na^+ , K^+ y Zn^{2+} han mostrado efecto inhibitorio (Saroj et al., 2022).

1.3 Algunos aspectos relevantes en la producción de celulasas

Las celulasas se han consolidado como una alternativa atractiva para el reemplazo de diversas técnicas industriales que requieren digestión de celulosa, constituyendo alrededor del 20% del mercado mundial de biocatalizadores (Chandel et al., 2012).

Sus altos costos de producción, estimados en USD \$10.14/kg (Klein-Marcuschamer et al., 2012), han impulsado estrategias de optimización como la producción *in situ*, el uso de materias primas de bajo costo, la reducción del tiempo de fermentación y la integración de procesos como la sacarificación y fermentación simultánea (SSF, por sus siglas en inglés), que permite reciclar enzimas y disminuir hasta un 44% los costos (Zhang, J. et al., 2017).

La demanda creciente ha intensificado la búsqueda y caracterización de nuevas isoformas con cinéticas y propiedades fisicoquímicas relevantes para la industria biotecnológica. El conocimiento acerca de las propiedades funcionales, fisicoquímicas y su análisis bioinformático, ha permitido conocer las condiciones de optimalidad para el uso de las celulasas y sus limitaciones, así como el diseño y mejoramiento de algunas de estas con funciones especiales, adicionales o sobresalientes. La caracterización enzimática se ha convertido en una herramienta indispensable en la investigación de estas proteínas y en un paso obligatorio previo al reconocimiento de una nueva proteína, ya sea en el mercado o en la comunidad científica.

En Colombia, la digestión enzimática de las fibras de celulosa es estratégica para diversos sectores productivos. En el país se generan aproximadamente 9.76 Mton de biomasa residual al año solo en la industria de alimentos (Gobierno de la República de Colombia, 2019). El aprovechamiento de dicho material, aunado a la bioprospección de nuevas enzimas celulolíticas aisladas de organismos constituyentes de la biodiversidad nacional, representan una ventaja competitiva para el desarrollo de productos con valor agregado. Por ello, la caracterización de

celulasas nativas en Colombia debe apuntar a la búsqueda de mejores y novedosas características enzimáticas; propiedades que permitan reducir el costo de producción de estas proteínas al evitar los costos de empaçado, transporte e importación. En este trabajo de grado se adopta un enfoque multidisciplinario orientado a analizar las propiedades de una celulasa aislada del microbioma intestinal del gusano barrenador de la caña. El objetivo es describir su estructura, su comportamiento fisicoquímico y su función biológica. A largo plazo, trabajos como este contribuirán a reducir la dependencia de material biotecnológico importado, impulsando el aprovechamiento de recursos enzimáticos obtenidos a partir de la biodiversidad nacional.

2. Objetivos

2.1 Objetivo General

Caracterizar las propiedades estructurales, fisicoquímicas y funcionales de una celulasa aislada del bacterioma intestinal del gusano barrenador la caña *Diatraea saccharalis* (Lepidoptera: Crambidae).

2.2 Objetivos Específicos

- Identificar la actividad celulasa *in vivo* de la cepa nativa seleccionada.
- Comprobar la actividad celulasa *in vitro* utilizando extractos crudos obtenidos de la cepa nativa.
- Caracterizar las propiedades fisicoquímicas de las celulasas presentes en los extractos crudos.
- Identificar genes candidatos responsables de la actividad celulasa observada *in vitro* mediante herramientas bioinformáticas.
- Determinar las propiedades fisicoquímicas y estructurales de las proteínas codificadas por los genes candidatos mediante análisis *in silico*.

3. Metodología

3.1 Identificación de la actividad celulasa en la cepa nativa

En un trabajo previo, la CIC aisló un conjunto de cepas bacterianas con actividad celulolítica a partir del bacterioma intestinal del barrenador de la caña de azúcar *D. saccharalis* (Atehortúa-Bueno, 2022). Debido a la alta actividad catalítica aparente observada, la cepa C1 fue seleccionada para ser caracterizada en mayor detalle. En el presente trabajo, el microorganismo fue recuperado del cepario CIC a partir de *criostocks* conservados a -80 °C en 20% glicerol. La cepa fue reactivada en medio Luria-Bertani (LB) sólido estándar (agar 15 g/L) e incubada durante 48 h a temperatura ambiente (25 – 30 °C). A partir de este cultivo se seleccionaron colonias aisladas para ser subcultivadas en medio LB sólido suplementado con carboximetilcelulosa (CMC, 2 g/L). Para identificar *in vivo* la existencia de actividad celulolítica, los cultivos sólidos se inundaron con 15 mL de Rojo Congo 0.1% durante 10 min y posteriormente fueron lavados dos veces con 15 mL de NaCl 2M durante 10 min. Para mejorar el contraste a un tono violeta-azul oscuro y facilitar la identificación de la actividad enzimática, ocasionalmente se agregaron 10 mL de ácido acético al 5% a las placas reveladas durante 10 min.

3.2 Determinación cualitativa de la actividad celulasa *in vitro*

Con el objeto de detectar cualitativamente la actividad celulasa *in vitro*, se diseñó un ensayo de difusión en medio sólido. Para ello, se prepararon cajas Petri con una solución de agar (15 g/L) y CMC (20 g/L). Una vez solidificadas, sobre la superficie de los agares se perforaron pozos de 3 mm de profundidad, en los que se depositaron 20 µL de diferentes extractos de cultivos bacterianos que posteriormente fueron incubados *overnight* a 42 °C. El revelado de la actividad se realizó siguiendo los pasos de tinción con Rojo Congo descritos previamente.

En total se realizaron 4 ensayos de difusión. En un primer ensayo se aplicaron directamente en los pozos 20 μ L de medios de cultivo líquidos de la cepa C1 y la cepa C10, otro microorganismo aislado por la CIC. Como controles negativos se utilizaron un volumen equivalente de un cultivo de *E. coli* EPI-300 y otro de un control no biológico con *buffer* tris-HCl 50 mM pH 8.0. En un segundo ensayo se utilizaron sobrenadantes de cultivos líquidos de la cepa C1 incubados a 24 y 29 h, a razón de dos muestras por cada período de incubación. El tercer ensayo se llevó a cabo con: dos muestras del sobrenadante de cultivos de la cepa C1, dos alícuotas de extractos proteicos del sobrenadante (previamente concentrados con sulfato de amonio) y, finalmente, 50 mM acetato de sodio pH 5.0 como control negativo. El cuarto ensayo utilizó cuatro diluciones de proteína extraída del sobrenadante de cultivo de la cepa C1, dos de ellas inactivadas previamente por temperatura a 100 °C durante 10 min., y, como control negativo, 50 mM acetato de sodio pH 5.0.

3.3 Métodos analíticos

Para los ensayos enzimáticos se prepararon cultivos de 200 mL de medio líquido LB a una concentración 0.5X con respecto a la concentración estándar, suplementados con CMC 10 g/L como fuente mayoritaria de carbono. La incubación se realizó durante 72 h a temperatura ambiente (25 – 33 °C) y con agitación constante a 230 rpm en una incubadora MAXQ 4450 (Thermo Fisher Scientific, EE. UU.). El cultivo se centrifugó a 4500 rpm durante 45 min en una centrífuga Sorvall ST 16 (Thermo Fisher Scientific, EE. UU.) para separar la biomasa del sobrenadante. El sobrenadante se almacenó a -80 °C y el pellet de biomasa se resuspendió en 50 mM acetato de sodio pH 5.0 en una relación 3:1 con respecto a su peso húmedo. Para el fraccionamiento celular se empleó un procesador ultrasónico de 130 W modelo CPX130PB (Cole-Parmer Instruments, EE. UU.) a una amplitud del 65% del total de la potencia, con pulsos de 30 s dentro de las muestras, seguido por incubación en hielo durante 1 min, repitiendo el ciclo durante 10 min. El lisado celular

se centrifugó a 15.000 x g en una centrífuga Sorvall ST 16 (Thermo Fisher Scientific, EE. UU.) durante 10 min para separar la fracción soluble de la fracción insoluble. Esta última se resuspendió en 500 μ L de 50 mM acetato de sodio pH 5.0. Con esto, las fracciones evaluadas correspondieron al sobrenadante del cultivo y a las fracciones soluble e insoluble del lisado celular.

Cada fracción del cultivo fue sometida a análisis espectrofotométrico por triplicado haciendo uso del método DNS (Miller, 1959) modificado. Por su parte, la concentración de proteína de las muestras fue determinada con el kit Qubit Protein Broad Range (BR) Assay y un fluorómetro Qubit 4 (Thermofisher Scientific, EE. UU.), siguiendo el protocolo del fabricante.

Las proteínas fueron concentradas mediante precipitación con sulfato de amonio saturado en una relación 1:1 con respecto al volumen del sobrenadante. Dicha precipitación se incubó *overnight* a 4 °C. El precipitado fue centrifugado a 4500 rpm durante 40 min en una centrífuga Sorvall ST 16 (Thermo Fisher Scientific, EE. UU.) y el *pellet* fue resuspendido en 1/10 de su volumen inicial con 50 mM acetato de sodio pH 5.0.

El análisis cuantitativo de la actividad celulasa se realizó mediante la detección de los azúcares reductores liberados, utilizando el método DNS (Miller, 1959) modificado. La preparación del reactivo DNS comprendió la disolución de 1 g de ácido 3,5-dinitrosalicílico (DNS) y 30 g de tartrato de sodio y potasio en 20 mL de 2N NaOH. Posteriormente, se adicionaron 0.83 g de metabisulfito de sodio y se aforó a 100 mL con agua destilada.

La reacción enzimática se llevó a cabo en un volumen de 500 μ L, de los cuales, 250 μ L correspondían al extracto bacteriano y 250 μ L a 1% CMC como sustrato. Como control negativo, el extracto bacteriano fue reemplazado por 250 μ L de agua destilada. La incubación de las reacciones se realizó a 45 °C durante 30 min. Posteriormente, el revelado consistió en la adición de 500 μ L de reactivo DNS 1%, seguido de una incubación a 100 °C durante 10 min. Para

estabilizar el color y mejorar la lectura de la absorbancia, las reacciones fueron dejadas en reposo a temperatura ambiente (Brodersen & Ricketts, 1949; Miller, 1959). Posteriormente, fueron medidas a 540 nm con un NanoPhotometer NP80 (IMPLEN, Alemania).

La curva de calibración para la detección de azúcares reductores fue preparada con estándares de glucosa. Las soluciones estándar tuvieron concentraciones de 0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5 y 2.75 mM de glucosa. Cada solución estándar se preparó por triplicado a partir de una solución *stock* de 100 mM glucosa en 50 mM acetato de sodio pH 5.0. Con los datos de absorbancia a 540 nm, se calculó mediante regresión lineal la proporcionalidad entre este parámetro y la concentración equivalente de glucosa.

3.4 Optimización del ensayo y caracterización de las celulasas

La optimización inició determinando la cantidad óptima de proteína para agregar a cada reacción enzimática (500 μ L). Para esto, se incubaron reacciones con 0, 25, 50, 100 y 125 μ g de proteína del sobrenadante de cultivos bacterianos.

Para la optimización de pH, las reacciones tuvieron 50 μ g de proteína en *buffers* 100 mM de: acetato de sodio a pH 3, 5 y 6; tris-maleato a pH 6, 7 y 8; y glicina a pH 8 y 10.

Para la optimización de temperatura se incubaron reacciones con 50 μ g de proteína en *buffer* 100 mM tris-maleato pH 6.0 durante 30 min. Para incubar las reacciones a 10, 20 y 30 °C se empleó un termociclador MiniAmp Plus (Thermofisher Scientific, EE. UU.) en su función de incubación VeriFlex; para las muestras a 40, 45, 50 y 80 °C se usó un baño seco con calentamiento y agitación (Thermofisher Scientific, EE. UU.), y, para las de 60 °C, se usó una incubadora digital IN55 (Mettler, Alemania).

En la optimización de tiempo de incubación se prepararon reacciones con 50 μ g de proteína en *buffer* 60 mM tris-maleato pH 6.0, a 50 °C durante 0, 10, 20, 30, 45, 60 y 120 min.

Finalmente, para la optimización de la cantidad de sustrato por reacción, se incubaron reacciones con 50 μg de proteína concentrada del sobrenadante en *buffer* 60 mM tris-maleato pH 6.0 a 50 $^{\circ}\text{C}$ durante 10 min, con CMC como sustrato a 0.05, 0.25, 0.5, 1.25, 2.5, 3.75, 5 y 10 g/L.

Para todos los casos de optimización, se organizaron los datos y se obtuvieron sus gráficos de dispersión como actividad específica en función de cada una de las variables anteriormente mencionadas, exceptuando la optimización del tiempo de incubación, en la que, debido a la variable a evaluar, no era viable graficar la actividad específica, por lo que se graficó la cantidad de nmol de glucosa producidos en función del tiempo de incubación de la reacción. De igual forma, en la optimización de la cantidad de proteína, debido a la variable a evaluar, no se graficó la actividad específica, en su lugar, se usaron los valores de actividad enzimática.

Para calcular la K_m y la V_{max} (teóricas) de la enzima responsable de la actividad celulasa se empleó la linealización de la curva de Michaelis-Menten mediante el método de Eadie-Hofstee (Eadie, 1942; Hofstee, 1959). Este método de reordenamiento de la ecuación de Michaelis-Menten permite hallar K_m y V_{max} en la ecuación de la curva de regresión lineal, en la que el intercepto en el eje y corresponde al valor de V_{max} , y la pendiente de la curva a K_m .

3.5 Análisis bioinformático de secuencias nucleotídicas y aminoacídicas

3.5.1 Análisis bioinformático de secuencias nucleotídicas

Los datos de la secuenciación génica de la cepa C1 previamente obtenidos por la CIC con la plataforma Illumina (NGS, *Next Generation Sequencing*, por sus siglas en inglés) fueron analizados mediante herramientas disponibles en la plataforma Galaxy Europe (<https://usegalaxy.eu/>) (The Galaxy Community, 2024). Para evaluar la calidad de lecturas obtenidas se utilizó FastQC (Andrews, s.f.). El ensamble *de novo* fue construido con Unicycler

(Wick et al., 2017), seguido de su ajuste y optimización con la herramienta Pilon (Walker et al., 2014). La calidad del ensamble fue evaluada con QUAST (Mikheenko et al., 2018).

El ensamble obtenido con lecturas cortas (*short-reads*) de C1 fue analizado con la plataforma *Type (Strain) Genome Server* (TYGS, por sus siglas en inglés (<https://tygs.dsmz.de>)) (Meier-Kolthoff et al., 2022; Meier-Kolthoff & Göker, 2019) para obtener una aproximación taxonómica. Las cepas tipo más cercanas se determinaron comparando el ensamble de C1 con los genomas disponibles en la base de datos del TYGS bajo el algoritmo MASH y sus secuencias 16S ADNr. Posteriormente se seleccionaron los genomas de las 10 cepas genéticamente más cercanas. Los valores de hibridación digital ADN-ADN (dDDH, *digital DNA-DNA Hybridization*, por sus siglas en inglés) e intervalos de confianza (I.C.) se calcularon con la configuración recomendada de GGDC 4.0 (Meier-Kolthoff et al., 2022, 2013). De las 10 cepas tipo seleccionadas, las tres más cercanas a C1 fueron empleadas para construir una visualización comparativa con *BLAST Ring Image Generator* (Alikhan et al., 2011).

Posteriormente, el genoma fue anotado con Prokka (Seemann, 2014). Dentro de los genes anotados, se realizó la búsqueda y curado manual de secuencias codificantes que potencialmente fueran las responsables de la actividad celulolítica. Para ello se llevó a cabo una búsqueda de enzimas relacionadas con la degradación de compuestos carbonados, especificando los nombres de las enzimas objetivo o características similares que pudieran llevar a identificarlas, como su sustrato, su actividad, su pertenencia a alguna familia de proteínas, entre otras. Así mismo se hizo uso de la plataforma dbCAN3 (Zheng et al., 2023), servidor web especializado en la anotación de genes codificantes para enzimas carbohidrato-activas (*CAZymes*, por su acrónimo en inglés), dentro de las cuales se hallan las celulasas.

Las secuencias anotadas como codificantes para *CAZymes*, provenientes tanto de Prokka como de dbCAN3, se contrastaron con las bases de datos UniProt (<https://www.uniprot.org/>) e InterPro (<https://www.ebi.ac.uk/interpro/>) en busca de similitudes con familias de proteínas y dominios, así como de información complementaria acerca de dichas enzimas. Para ello se hizo uso de la traducción de la secuencia genómica a secuencia proteica obtenida del ensamble anotado con Prokka.

Con base en la información previamente obtenida acerca de los genes anotados, se determinaron dos secuencias de posibles genes como candidatos a ser los codificantes de las enzimas responsables de la actividad celulasa de la cepa C1, y, con ellas, se continuó el proceso de caracterización fisicoquímica y estructural *in silico*.

3.5.2 Análisis bioinformático de secuencias aminoacídicas

Una vez estuvieron determinados los genes candidatos, se amplió la ventana informativa de cada una de las secuencias traducidas con anotaciones predictivas hechas a dichas secuencias de aminoácidos con las bases de datos InterPro y UniProt, incluyendo en tales anotaciones aspectos como posibles componentes estructurales de las proteínas, señalizaciones o etiquetas propias de enzimas de su tipo.

Las secuencias aminoacídicas fueron sometidas al cálculo de propiedades fisicoquímicas con la herramienta ProtParam del portal web ExPASy. Se obtuvieron los valores teóricos del punto isoeléctrico (pI), peso molecular (MW), índice de inestabilidad (II) e índice alifático (AI).

Una vez sus características fisicoquímicas fueron determinadas *in silico*, las secuencias de aminoácidos fueron depositadas en el servidor de la herramienta PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) con el fin de obtener la estructura secundaria de cada una de las enzimas.

Como paso final, un modelamiento 3D de la estructura terciaria de las proteínas fue realizado con el servidor de AlphaFold (<https://alphafoldserver.com/welcome>). La visualización de las predicciones *in silico* también se llevó a cabo en la interfaz de este servidor.

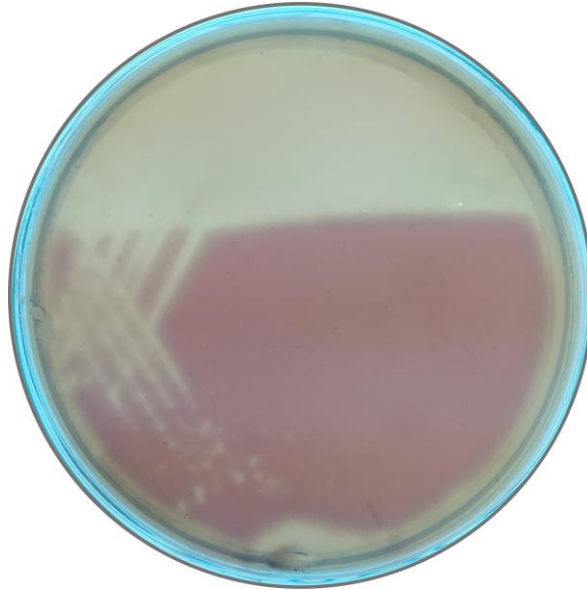
4. Resultados

4.1 Identificación de la actividad celulasa en la cepa nativa

La cepa C1 fue reactivada de la ultracongelación en medio LB sólido. Luego de comprobar su pureza, fue subcultivada en medio LB sólido suplementado con 2 g/L CMC. El revelado con Rojo Congo exhibió los halos de degradación del polímero CMC (Fig. 2). Dichos halos de degradación enzimática confirmaron la actividad celulasa de la cepa C1 y brindaron indicios preliminares del carácter extracelular de la enzima.

Figura 2.

Revelación de la actividad celulasa de la cepa C1 usando el tincionante Rojo Congo



Nota. En amarillo tenue se diferencian las zonas en las que el polímero sintético CMC fue hidrolizado, en rojo las zonas sin influencia enzimática.

4.2 Comprobación cualitativa y cuantificación de la actividad celulasa

La actividad celulasa *in vitro* fue comprobada mediante la observación de halos de degradación en el medio sólido Agar-CMC, producto de los ensayos de difusión que fueron teñidos con Rojo Congo. Las muestras del primer ensayo de difusión revelaron la actividad celulolítica de la cepa C1, así como la falta de esta misma en las cepas C10 y EPI-300 que fungieron como controles negativos (Fig. 3A).

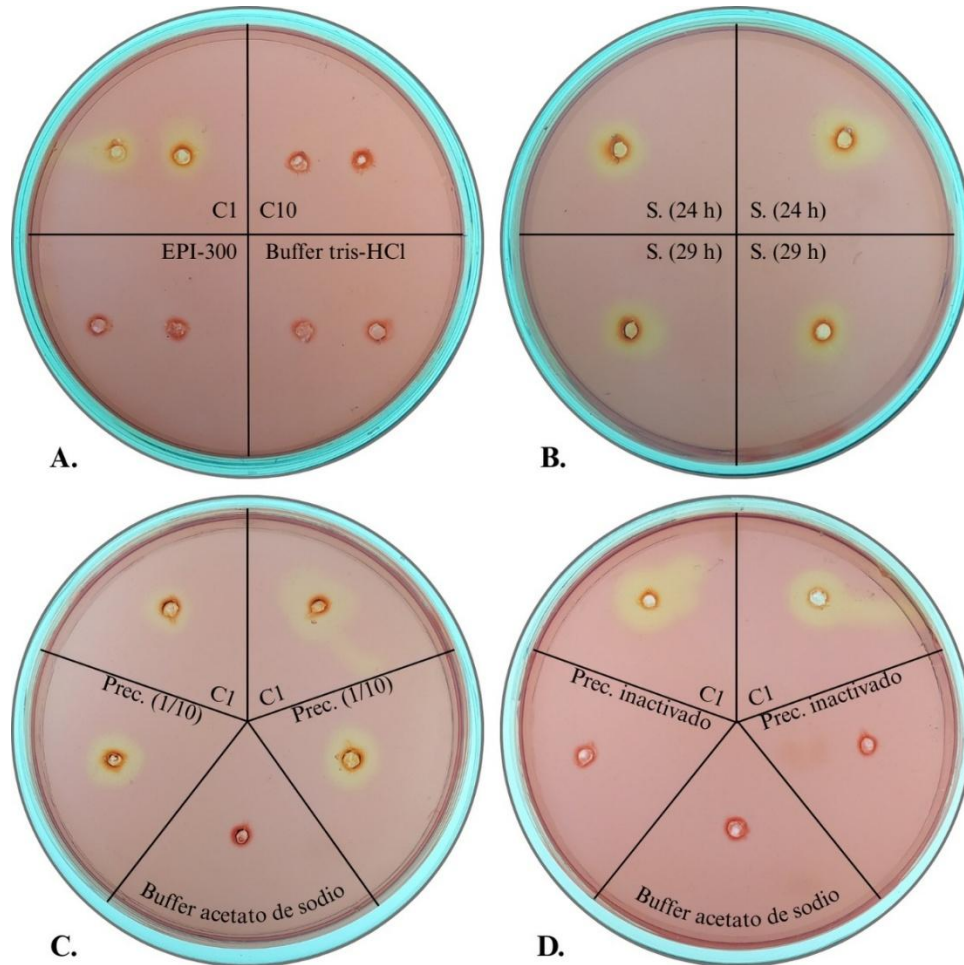
Para el caso del segundo ensayo, las 4 muestras de sobrenadante obtenidas de cultivo líquido fresco de la cepa C1, a distintos tiempos de recolección (24 y 29 h), demostraron actividad celulasa independientemente de las diferencias entre sus tiempos de incubación (Fig. 3B).

Al identificar que la actividad se concentraba en el sobrenadante del medio de cultivo, en un tercer ensayo, se realizó una precipitación y concentración de las proteínas en esta fracción del cultivo y se evaluó su actividad sobre las placas de difusión. Las proteínas resuspendidas en *buffer* acetato de sodio presentaron actividad catalítica (Fig. 3C). Como control positivo se utilizó sobrenadante del cultivo de la cepa C1 no precipitado, y como control negativo *buffer* acetato de sodio.

En el cuarto ensayo, se adicionaron muestras de la proteína concentrada previamente e inactivadas por calor. Estas últimas no presentaron halos de degradación de CMC (Fig. 3D), lo que refuerza la hipótesis de la naturaleza proteica de la actividad catalítica.

Figura 3.

Ensayos de difusión enzimática en cajas Petri de Agar-CMC



C1: Cepa nativa *Microbacterium* sp.
C10: Cepa nativa *Klebsiella* sp.
EPI-300: Cepa estándar de *E. coli*

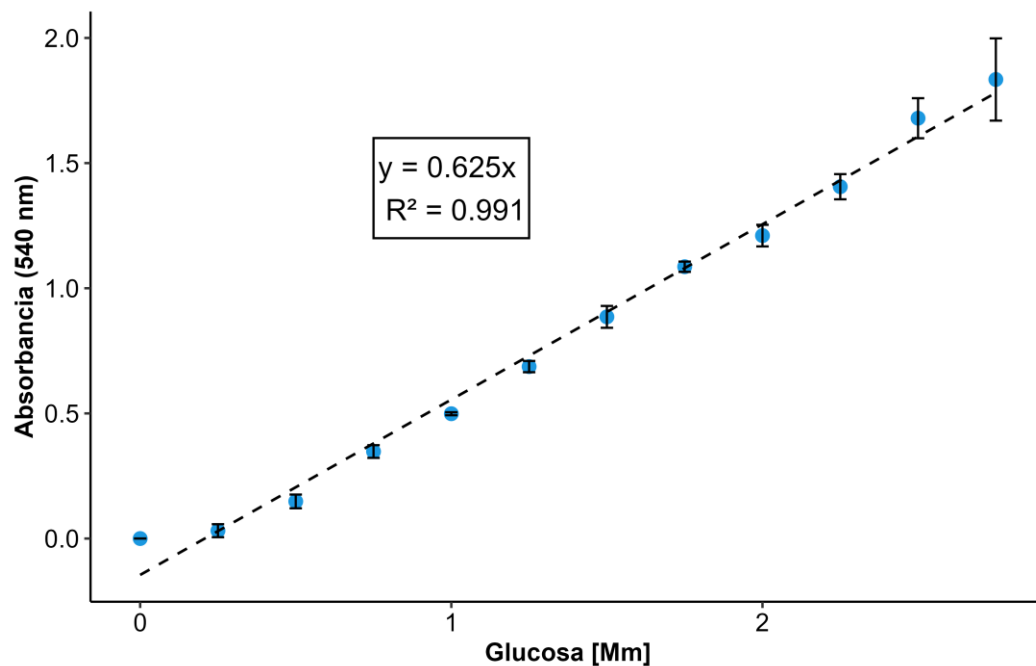
S.: Sobrenadante
Prec.: Precipitado

Nota. Comprobaciones *in vitro* de **A:** actividad celulasa de la cepa C1, **B:** actividad celulasa en el sobrenadante, **C:** actividad celulasa en la precipitación proteica, **D:** inactivación por calor de la enzima. En amarillo tenue se diferencian las zonas en las que el polímero sintético CMC fue hidrolizado.

La cuantificación de producto y de la actividad enzimática se realizó mediante el método espectrofotométrico DNS. Para realizar la cuantificación se preparó una curva estándar de glucosa en el rango en el que el incremento de la reducción del DNS a 3-amino-5-nitrosalicilato es proporcional al aumento de glucosa en cada muestra, detectado como un incremento en la señal de absorbancia (Fig. 4).

Figura 4.

Curva de calibración de glucosa para la implementación del ensayo enzimático DNS



Nota. Todas las mediciones se realizaron por triplicado, los puntos en el gráfico corresponden a sus medias y las barras de error a sus desviaciones estándar. Los valores de la pendiente y el coeficiente de determinación fueron aproximados a 3 decimales.

La actividad específica obtenida del sobrenadante alcanzó 5.24 nkat/mg, una cifra 2.7 veces superior a los 1.76 y 1.92 nkat/mg que se obtuvieron con su control negativo e inactivación (Tabla 1). Por otro lado, la fracción soluble exhibió una actividad específica de 0.24 nkat/mg, un valor 20 veces inferior al del sobrenadante del medio de cultivo. Lo anterior, nuevamente sugiere el carácter extracelular de la actividad catalítica. Esta situación fue similar con los controles e inactivaciones de esta fracción, que presentaron 0.07 y 0.06 nkat/mg, respectivamente. Por su parte, en la fracción insoluble, tanto esta como su control e inactivación presentaron 0.15 nkat/mg de actividad específica.

Tabla 1.

Resultados de actividad enzimática en las fracciones celulares evaluadas

Muestra	Volumen (mL)	Proteína total (mg)	Actividad total (nkat)	Actividad específica (nkat/mg)
S.	0.25	0.064	0.335 (0.012)	5.239 (0.184)
Control S.	0.25	0.064	0.113 (0.009)	1.763 (0.145)
S. inactivado	0.25	0.064	0.123 (0.006)	1.923 (0.095)
F.S.	0.25	1.085	0.275 (0.044)	0.253 (0.040)
Control F.S.	0.25	1.085	0.077 (0.008)	0.071 (0.007)
F.S. inactivada	0.25	1.085	0.069 (0.025)	0.063 (0.023)
F.I.	0.25	8.050	1.201 (0)	0.149 (0)
Control F.I.	0.25	8.050	1.201 (0)	0.149 (0)
F.I. inactivada	0.25	8.050	1.195 (0.011)	0.148 (0.001)

Nota. **S.:** Sobrenadante, **F.S.:** Fracción Soluble, **F.I.:** Fracción Insoluble. Todas las mediciones se realizaron por triplicado. Las inactivaciones se realizaron a 100 °C. En la tabla se muestran los promedios y entre paréntesis sus desviaciones estándar.

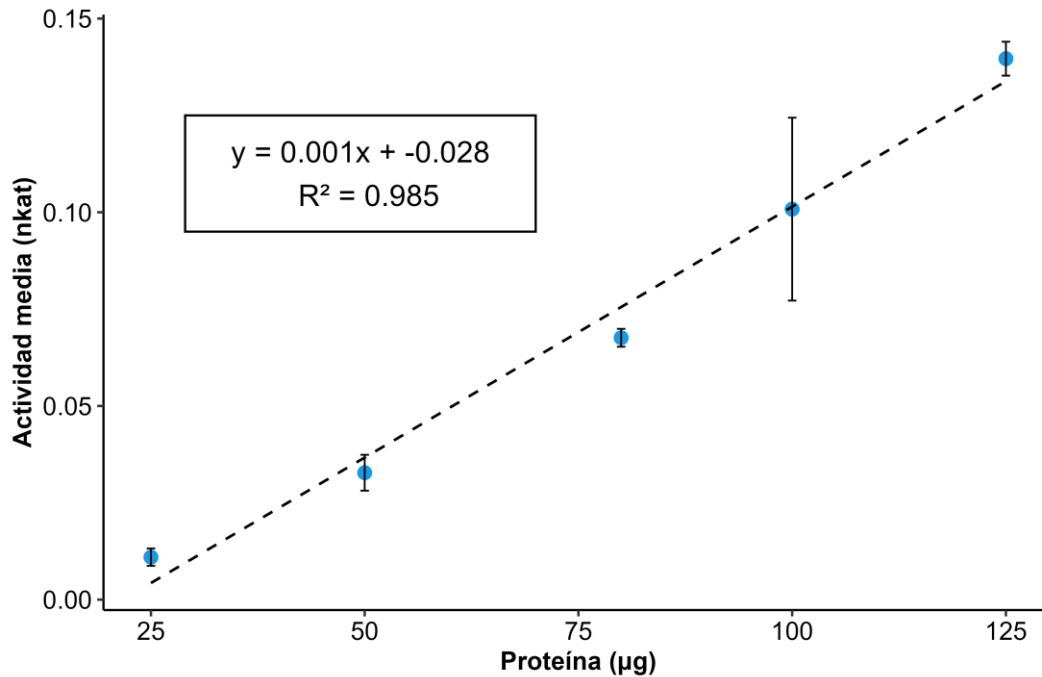
Los anteriores resultados evidencian que la fracción con mayor actividad específica es el sobrenadante del medio de cultivo. En consecuencia, esta fue la fracción seleccionada para continuar con la caracterización bioquímica de la actividad celulasa.

4.3 Optimización de catálisis y caracterización fisicoquímica de las celulasas

La optimización de concentración de proteína por reacción dio como resultado una curva de regresión lineal en la que, teóricamente, podría usarse cualquiera de las concentraciones evaluadas y la actividad seguiría comportándose de forma proporcional con respecto a la producción de glucosa (Fig. 5). El coeficiente de determinación en este caso, con un valor de 0.99, permitió darles confiabilidad a los datos obtenidos, por lo cual, la cantidad de proteína seleccionada para el ensayo enzimático in vitro fue de 50 μg .

Figura 5.

Resultados de la optimización de concentración de proteína por reacción



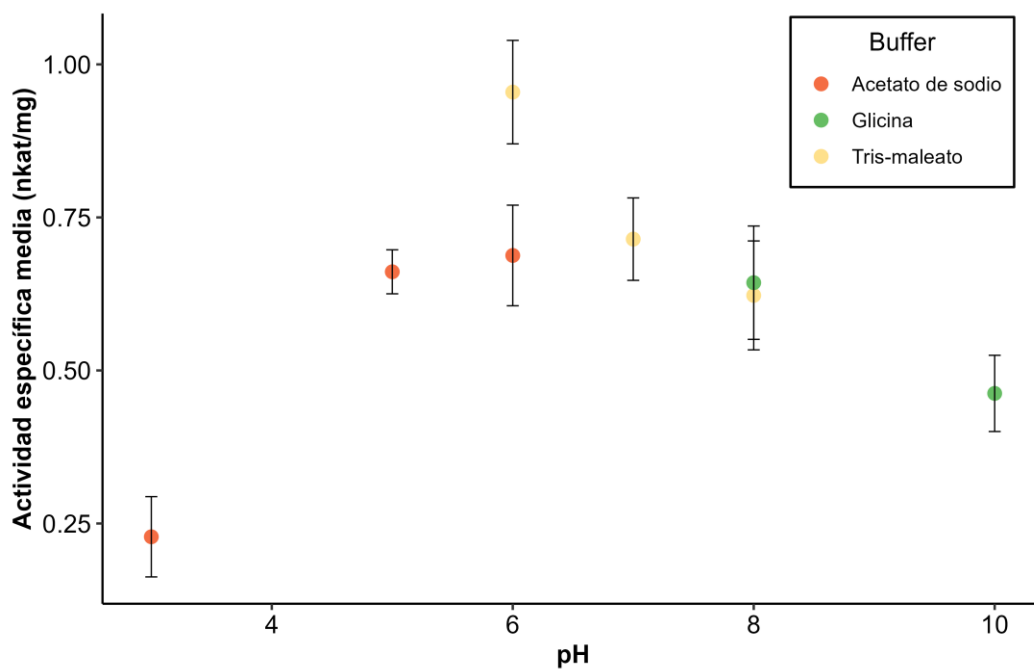
Nota. Los datos mostrados corresponden a mediciones promedio de triplicados y las barras de error a sus desviaciones estándar.

Para la optimización de pH, tres tipos de *buffer* fueron testeados; acetato de sodio, tris-maleato y glicina. En general, el *buffer* que mejor desempeño presentó en la amortiguación del pH de la reacción fue el de tris-maleato, provocando que las muestras en las que este fue agregado se detectara mayor actividad con respecto a los *buffers* acetato de sodio y glicina (Fig. 6). La actividad específica más alta se reportó haciendo uso del *buffer* 100 mM tris-maleato pH 6.0, con 0.96 nkat/mg. Los puntos de pH en los que se solapan los tipos de *buffer* ayudaron a determinar si existían diferencias relevantes entre el uso de uno u otro al evaluar la producción de glucosa, siendo evidente la mejoría al usar el de tris-maleato pH 6.0. Con lo anterior, se puede indicar que la enzima

de interés es levemente acidófila, pues sus picos de actividad se encuentran entre pH 5 y 7, sugiriendo las condiciones a las que la enzima, y por ende la cepa, se han adaptado. Dados los resultados anteriores, el *buffer* seleccionado para continuar con la optimización fue el de tris-maleato pH 6.0.

Figura 6.

Resultados de la optimización de pH de la reacción



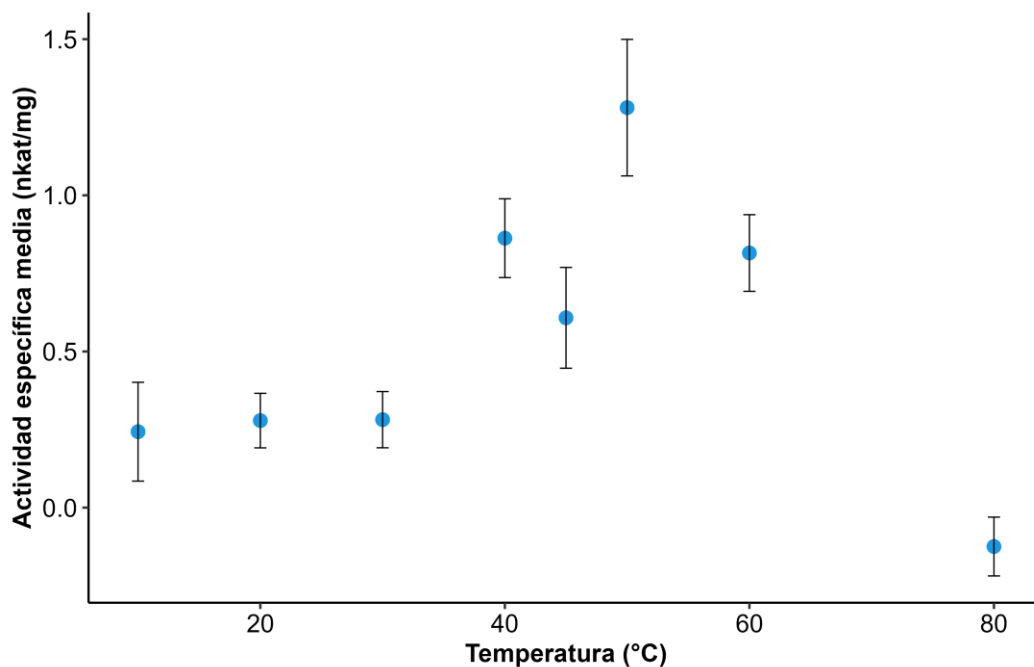
Nota. Los datos mostrados corresponden a mediciones promedio de triplicados y las barras de error a sus desviaciones estándar.

En el caso de la temperatura, la enzima tuvo su máxima actividad específica (1.28 nkat/mg) a 50 °C (Fig. 7). La baja producción de glucosa obtenida a temperaturas entre 10 y 30 °C deja entrever que la reacción catalítica no es óptima a temperaturas bajas. Por otro lado, la disminución

de la actividad específica a temperaturas superiores a 50 °C sugiere susceptibilidad de la enzima a condiciones extremas de temperatura. Dado esto, la temperatura seleccionada fue de 50 °C.

Figura 7.

Resultados de la optimización de temperatura de la reacción



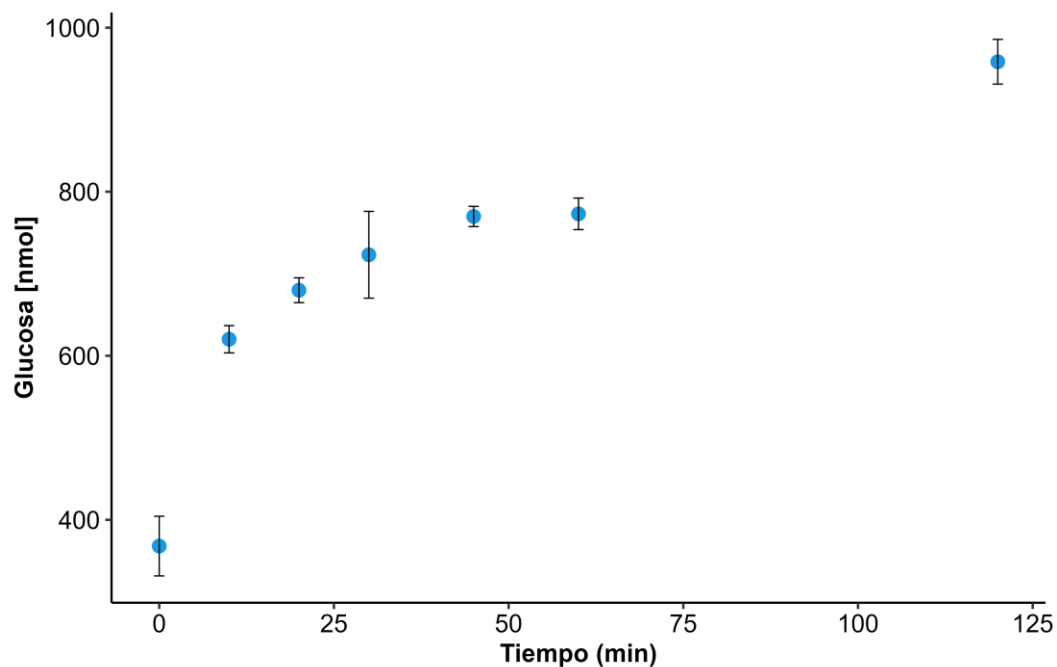
Nota. Los datos mostrados corresponden a mediciones promedio de triplicados y las barras de error a sus desviaciones estándar.

En la evaluación del tiempo de incubación se evidenció estabilización de la producción de glucosa en función del tiempo. El valor más alto de glucosa se obtuvo a los 120 min, siendo de 958.48 nmol (Fig. 8). No obstante, se observó que la fase de transición entre la fase exponencial y la fase estacionaria de la curva se alcanza hacia los 30 minutos de incubación, y en cuyo caso presentó una actividad específica de 8.04 nkat/mg. A pesar de ello, la actividad más alta se obtuvo

a los 0 minutos de incubación, seguida de la incubación de 10 min. Con esto, el tiempo de incubación seleccionado fue de 10 min.

Figura 8.

Resultados de la optimización del tiempo de la reacción

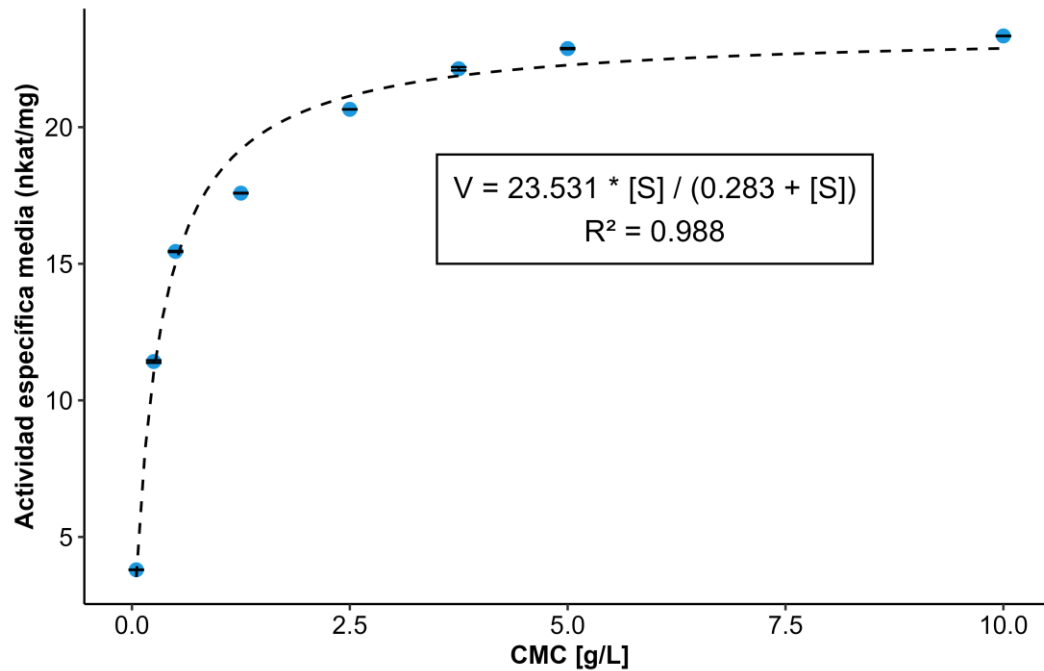


Nota. Los datos mostrados corresponden a mediciones promedio de triplicados y las barras de error a sus desviaciones estándar.

Una vez determinadas las condiciones óptimas del ensayo, en términos de pH, temperatura, tiempos de incubación y concentración de enzima, se procedió con la determinación de la cinética del extracto enzimático concentrado del sobrenadante del cultivo a partir de su comportamiento a diferentes concentraciones de CMC. Los datos mostraron una hipérbola rectangular, sugiriendo que la enzima es de naturaleza michaeliana (Fig. 9).

Figura 9.

Cinética enzimática de la enzima evaluada

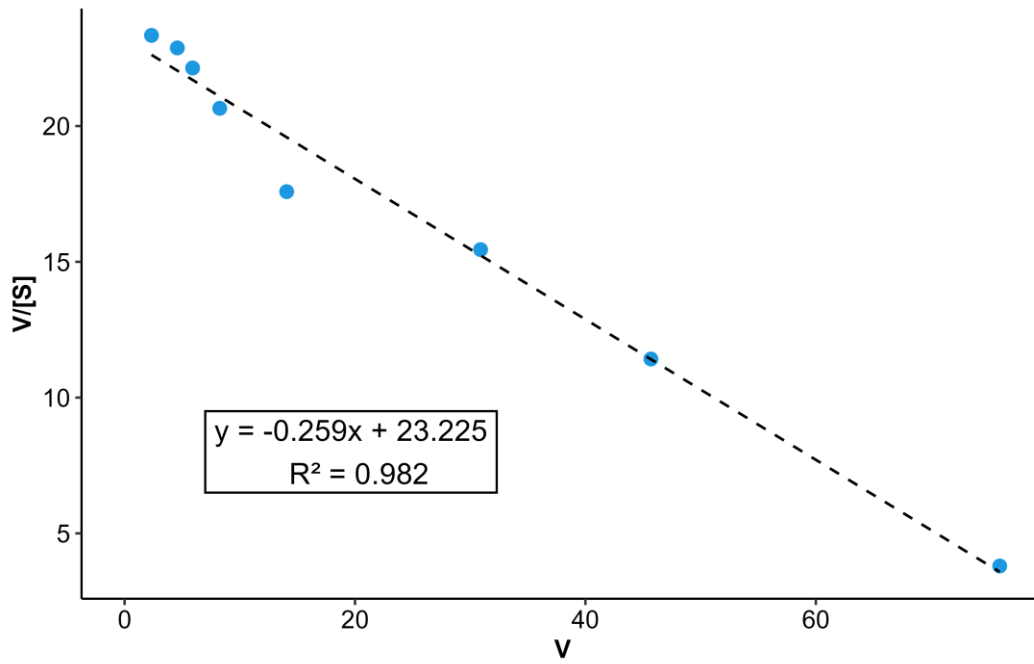


Nota. Los datos mostrados corresponden a mediciones promedio de triplicados y las barras de error a sus desviaciones estándar. La fórmula corresponde a la ecuación de Michaelis-Menten.

A partir de la curva de Michaelis-Menten obtenida se aplicó el modelo de linealización de Eadie-Hofstee. Al aplicar esta regresión lineal, la K_m , correspondiente al valor de la pendiente, y que hace referencia a la afinidad de la enzima por el sustrato y con qué cantidad de este se alcanza $\frac{1}{2}$ de la V_{max} , fue de 0.26 g/L de CMC; y la V_{max} , correspondiente al valor del intercepto en el eje y (Fig. 10), y que hace referencia a la velocidad máxima de catálisis bajo condiciones óptimas, fue de 23.23 nkat/mg.

Figura 10.

Linealización de la curva de Michaelis-Menten bajo el método de Eadie-Hofstee



Nota. V = Velocidad, $[S]$ = Concentración de sustrato. En la ecuación de la regresión: $m = K_m$, y $b = V_{max}$.

4.4 Análisis bioinformático de secuencias nucleotídicas y aminoacídicas

4.4.1 Análisis bioinformático de secuencias nucleotídicas

A partir de las *short-reads* obtenidas mediante un ejercicio previo de secuenciación en la plataforma Illumina, se construyó el ensamble *de novo* del genoma bacteriano de la cepa C1. Como primera medida, con la herramienta FastQC se evaluó la calidad de las secuencias pareadas obtenidas (Tabla 2). En ambos casos, tanto para la secuencia *forward* como para la *reverse*, los porcentajes de GC estuvieron alrededor de 70. Ninguna lectura tuvo bases no identificadas, no se

hallaron secuencias sobrerrepresentadas ni contaminación de los adaptadores usados para la secuenciación; sin embargo, se reportaron niveles moderados de duplicación de secuencias.

Tabla 2.

Resultados de la evaluación de calidad de secuencias NGS con FastQC

Secuencia pareada	Total secuencias	Total Mpb	% GC	Longitud media por secuencia (pb)	Phred score medio (Q)	% Persistencia pos-duplicación
Forward	5,118,792	767.8	70	150	36	62.66
Reverse	5,118,792	767.8	70	150	36	62.17

Estos porcentajes de duplicación, que estuvieron entre 37 – 38%, fueron desestimados debido a su condición de “moderados”, pues entre el 62 – 63% de las secuencias se mantendrían en el *dataset* aún después de llevar a cabo una deduplicación.

Para la creación del ensamble *de novo* se utilizó Unicycler bajo los tres modelos de *bridging* habilitados por la herramienta (*conservative, normal* y *bold*). De estos, se obtienen como resultados un número mayor de *contigs* (secuencias continuas) pero con menos errores de ensamblaje; *contigs* y errores de ensamblaje en una tasa balanceada; y un menor número de *contigs*, pero con una tasa de errores de ensamblaje más alta, respectivamente.

Haciendo uso de la herramienta QUAST se evaluó la calidad de los tres ensambles obtenidos anteriormente (Tabla 3).

Tabla 3.

Resultados de los tres ensambles sometidos a evaluación de calidad con QUASt

<i>Bridging</i>	Nº <i>contigs</i>	<i>Contigs</i> ≥ 1000 pb	<i>Contig</i> más largo (pb)	Tamaño total	N50 (pb)	N90 (pb)	L50	L90
<i>Conservative</i>	56	31	434,855	3,262,726	269,639	63,279	5	14
<i>Normal</i>	47	28	704,654	3,263,370	279,061	63,279	4	13
<i>Bold</i>	41	24	704,654	3,264,607	358,984	67,592	4	11

Para el caso del número de *contigs* de cada ensamble, se evidenció como del modo *conservative* al modo *bold* estos disminuían, mostrando la menor fragmentación del ensamble en el modo *bold*. De igual forma, en los otros ítems a evaluar como el N50 y N90, se puede observar el aumento en el número de pares de bases conforme se hace menos estricto el modelo de *bridging*, así como la disminución del número de *contigs* en L50 y L90 si se toma en este mismo sentido. Valores como los de L50, L90, N50 y N90 permiten inferir que la diferencia en el número de *contigs* entre los ensambles está representada por fragmentos pequeños que no abarcan más del 10% del total de sus longitudes totales. Un ejemplo de esto puede darse con el L90 y el tamaño total de los ensambles; mientras que bajo el modelo *bold* es posible abarcar el 90% del ensamble con 11 *contigs*, bajo el modelo *conservative* fueron necesarios 14. Si comparamos esta métrica con el número de *contigs* totales, se puede observar que el 10% del ensamble bajo el modelo *bold* estaría fragmentado en 30 *contigs*, mientras que bajo el modelo *conservative* el 10% estaría repartido entre 42 *contigs*, siendo este 10% alrededor de 320 kpb en ambos ensambles. Adicional a esto, el número de *contigs* mayores a 1000 pb mostró que una buena parte del número de ellos superan esta cifra.

Antes de continuar con el análisis *in silico*, los ensamblajes fueron sometidos a un mejoramiento o *polishing* con ayuda de Pilon, una herramienta bioinformática que permite mejorar la continuidad y calidad de los ensamblajes generados *de novo*. Los ensamblajes se cargaron junto con archivos BAM (*Binary Alignment Map*, por sus siglas en inglés) de las secuencias pareadas originales alineadas a cada ensamblaje; de esta forma, el algoritmo puede identificar posibles errores de ensamblaje para corregirlos teniendo como molde los datos NGS y, así, mejorar las métricas de cada uno de los ensamblajes. Una vez el *polishing* fue realizado, se evaluaron los ensamblajes con QUAST (Tabla 4). En cuanto al número de *contigs*, L50, L90, N50, N90, la longitud del ensamblaje y las métricas clásicas de QUAST, no hubo variaciones entre los ensamblajes con y sin *polishing*. Adicional a esto, algunos otros aspectos como los porcentajes de mapeo, de emparejamiento apropiado, de *singletons* y pares mal unidos también aportaron información sobre la calidad del ensamblaje.

Tabla 4.

Resultados de la evaluación de calidad con QUAST a los tres ensamblajes pos-polishing

<i>Bridging</i>	Total mapeado	% Mapeo	% Emparejamiento apropiado	% <i>Singletons</i>	% Pares mal unidos
<i>Conservative</i>	10,182,396	99.46	98.85	0.15	0.34
<i>Normal</i>	10,187,259	99.51	98.91	0.13	0.33
<i>Bold</i>	10,188,626	99.52	98.96	0.13	0.31

En términos del porcentaje de mapeo, en los tres ensamblajes fue superior al 99.4%; el emparejamiento apropiado estuvo por encima del 98.8%, y los porcentajes de *singletons* y pares mal unidos fueron inferiores al 0.16 y 0.35%, respectivamente. Estos valores muestran que, desde el comienzo, los ensamblajes tuvieron buena calidad, lo que se evidencia en los altos porcentajes de mapeo y emparejamiento apropiado, así como en las métricas de sus *contigs*, demostrando que la mayoría del genoma secuenciado fue correctamente ensamblado. Por otro lado, los bajos porcentajes de *singletons* indican que fueron pocos los pares de bases que no lograron ser emparejados, y de igual forma, el bajo porcentaje de pares mal unidos refleja la baja tasa de error que el ensamblador tuvo desde el inicio, lo que refuerza aún más el hecho de que los ensamblajes tuvieron una alta calidad incluso antes del *polishing*.

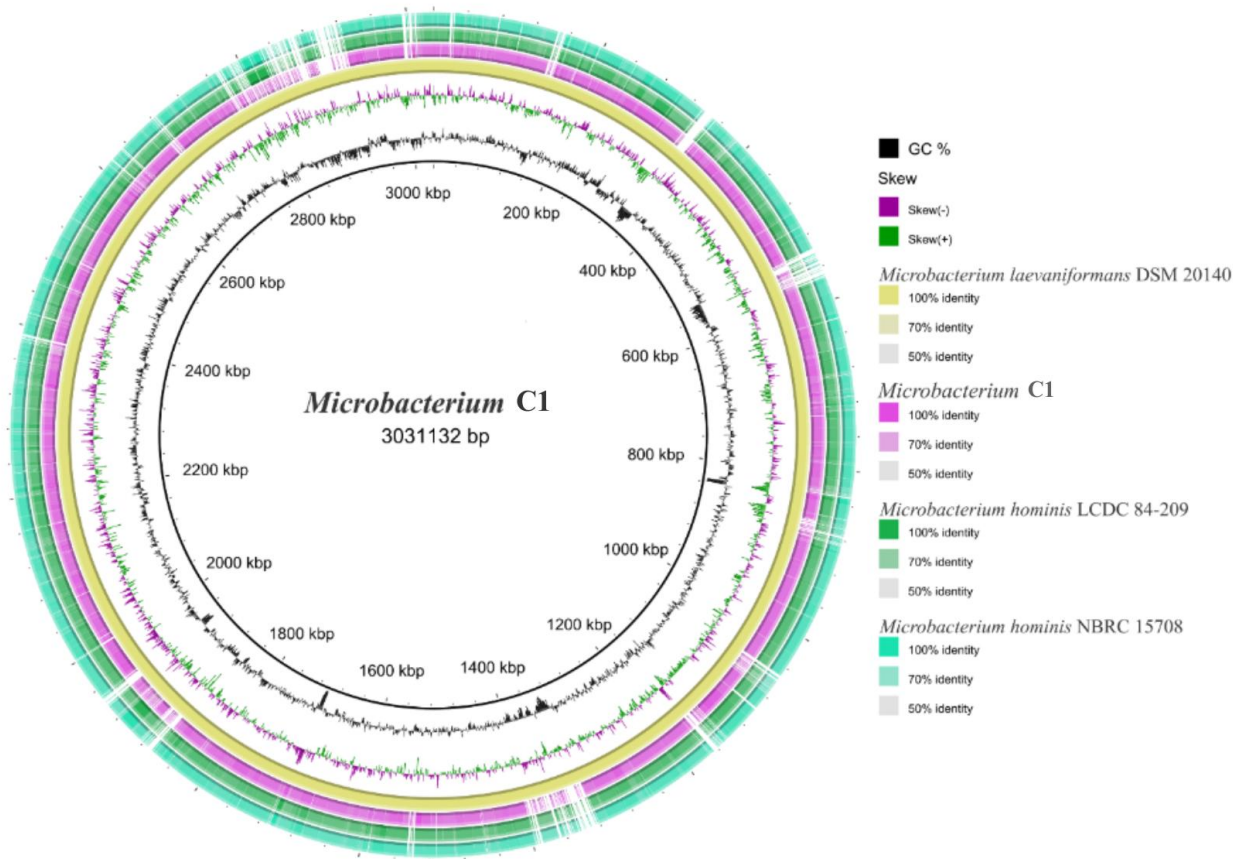
Con el objeto de aproximarse a la identidad taxonómica de C1 se contrastó su genoma contra el de cepas tipo disponibles en la TYGS. El análisis dDDH mostró que la cepa más cercana a C1 sería *Microbacterium laevaniformans* DSM 20140 con un valor d_4 de 32.1% y 49.6% para d_6 (Tabla 5). Dada la evidente cercanía entre cepas tipo del género *Microbacterium* y C1, esta última fue asignada a este taxón al nivel de género para su análisis posterior. La comparación por medio de un *BLAST Ring* con *M. laevaniformans* DSM 20140 como cepa de referencia evidenció las diferencias entre los genomas más cercanos a la cepa C1 (Fig. 11), destacando decenas de regiones genómicas en las que la resolución es baja y requieren de mayor esfuerzo de mapeo.

Tabla 5.*Hibridación digital ADN-ADN (dDDH) entre cepas tipo de la TYGS y CI*

Cepa tipo (TYGS)	dDDH			
	d ₄	I.C.	d ₆	I.C.
<i>Microbacterium laevaniformans</i> DSM 20140	32.1	[29.7 – 34.6]	49.6	[46.6 – 52.7]
<i>Microbacterium hominis</i> NBRC 15708	28.2	[25.9 – 30.7]	47.6	[44.5 – 50.6]
<i>Microbacterium hominis</i> LCDC 84-209	28.2	[25.8 – 30.7]	46.7	[43.6 – 49.7]
<i>Microbacterium aurum</i> DSM 8600	24.0	[21.7 – 26.5]	26.5	[23.6 – 29.6]
<i>Microbacterium paulum</i> 2C	23.2	[20.9 - 25.6]	26.0	[23.1 - 29.1]
<i>Microbacterium lacticum</i> JCM 1379	23.1	[20.8 - 25.5]	25.7	[22.8 - 28.8]
<i>Microbacterium lacticum</i> DSM 20427	23.1	[20.8 - 25.5]	25.2	[22.3 - 28.3]
<i>Microbacterium lacticum</i> NBRC 14135	23.1	[20.8 - 25.6]	25.6	[22.7 - 28.7]
<i>Microbacterium dextranolyticum</i> DSM 8607	22.8	[20.5 - 25.2]	27.3	[24.4 - 30.4]
<i>Microbacterium telephonicum</i> S2T63	21.9	[19.6 - 24.3]	24.9	[22.0 - 28.0]

Figura 11.

Representación BLAST Ring entre el ensamble de C1 y sus tres cepas tipo más cercanas



Nota. El genoma de referencia corresponde al de la cepa tipo *M. laevaniformans* DSM 20140. En lila brillante se observa el anillo de C1 correspondiente a su *draft genome*.

Debido a lo anterior, se decidió someter los tres ensamblados resultantes a la anotación genómica utilizando Prokka. Luego de un curado manual de los resultados, se halló que, para los tres ensamblados, Prokka encontró y anotó cuatro potenciales genes relacionados con la función de degradación enzimática de celulosa, dos de ellos correspondientes a probables codificantes para endoglucanasas (identificadas como C y E1), y los otros dos a β -glucosidasas (una de ellas

identificada como B) (Tabla 6). Las secuencias correspondientes a estos hallazgos fueron completamente idénticas en los tres ensamblajes sometidos a anotación, por lo que se determinó continuar los análisis con el ensamblaje obtenido bajo el modelo *conservative* y seguidamente se continuó la búsqueda de características con las secuencias aminoacídicas anotadas. Dichas secuencias de aminoácidos anotadas también fueron proporcionadas por Prokka.

Tabla 6.

Resultados de la anotación de genes realizada con Prokka

Identificador ensamblaje	Nombre de la proteína	Longitud (nucleótidos)	Longitud (aa)
AKHMMEEA_02300	Endoglucanasa C	3,388	1,128
AKHMMEEA_02341	Endoglucanasa E1	1,646	547
AKHMMEEA_01537	β -glucosidasa	1,253	416
AKHMMEEA_01663	β -glucosidasa B	1,534	510

Para obtener más información al respecto sobre posibles secuencias codificantes, el ensamblaje se cargó al sitio web de la herramienta dbCAN3 (<https://bcb.unl.edu/dbCAN2/>) para realizar una búsqueda y anotación de genes codificantes para *CAZymes* o proteínas con actividad enzimática asociada a carbohidratos, basándose en la información reportada en bases de datos y herramientas como dbCAN, dbCAN-sub (Zheng et al., 2023), dbCAN-PUL (Ausland et al., 2021), eCAMI (Xu, J. et al., 2020), y la CAZy *database* (Drula et al., 2022). Los resultados de la anotación arrojaron decenas de enzimas que probablemente estarían relacionadas a la degradación de carbohidratos como la celulosa o compuestos similares.

Al someter a curado manual estos datos bioinformáticos, se lograron identificar 11 secuencias preliminares de genes que tenían actividad enzimática relacionada con la degradación de celulosa, a la vez que cuatro de estas resultaron ser las mismas secuencias anotadas previamente por Prokka como codificantes de enzimas celulasas (Tabla 7). Ya que los indicios previos muestran una proteína extracelular, de las cuatro secuencias anotadas, las dos correspondientes a probables enzimas extracelulares (endoglucanasas C y E1) fueron determinadas como las candidatas principales para ser las codificantes de las enzimas con la actividad celulasa de interés.

Tabla 7.

Resultados de la búsqueda y anotación de potenciales genes asociados a la degradación enzimática de celulosa con dbCAN3

Identificador ensamble	Herramienta predictiva	Familia	EC asociados	Péptido señal	Sustrato (dbCAN_sub)
AKHMMEEA_02341	HMMER	GH5 *	3.2.1.4	1-34	β -glucano, celulosa, quitina, xilano
AKHMMEEA_02300	HMMER	GH9 + CBM4 *	3.2.1.4 3.2.1.91 3.2.1.- 3.2.1.73 3.2.1.8	1-35	β -glucano, celulosa, xilano
AKHMMEEA_01537	HMMER	GH1 *	3.2.1.21	-	β -glucano
AKHMMEEA_01663	HMMER	GH1 *	3.2.1.21 3.2.1.- 3.2.1.38 3.2.1.23 3.2.1.74	-	Polifenol, β -glucano, β -galactano, β -fucósidos
AKHMMEEA_01380	HMMER	GH3	3.2.1.21 3.2.1.38 3.2.1.-	-	β -glucano

Tabla 7 (continuación).

Identificador ensamble	Herramienta predictiva	Familia	EC asociados	Péptido señal	Sustrato (dbCAN_sub)
AKHMMEEA_00283	HMMER	GH3	3.2.1.21 3.2.1.38 3.2.1.-	-	β -glucano
AKHMMEEA_00021	HMMER	GH6	-	1-32	-
AKHMMEEA_00152	HMMER	GH3	3.2.1.37	-	Xilano
AKHMMEEA_02946	HMMER	GH1	3.2.1.86 3.2.1.21 3.2.1.37	-	β -glucano, xilano
AKHMMEEA_02869	HMMER	GH3	3.2.1.21	-	β -glucano
AKHMMEEA_02841	HMMER	GH3	3.2.1.21	-	β -glucano

Nota. Los resultados corresponden a las 11 secuencias identificadas preliminarmente luego de someter los datos a curado manual. El asterisco en la columna de “Familia” señala las secuencias que también fueron anotadas por Prokka.

4.4.2 Análisis bioinformático de secuencias aminoacídicas

Las secuencias primarias fueron sometidas a corroboración y ampliación de la información haciendo búsquedas de similitud en la base de datos de UniProt y realizando anotaciones de las secuencias con InterPro (Blum et al., 2025). Con estas búsquedas se confirmó la potencial calidad de enzimas celulasas de los genes identificados tras la anotación y se halló información adicional, como péptidos señal o algunos componentes que permitirían desvelar un poco más a fondo su rol en los procesos bioquímicos de la cepa (Tabla 8).

Tabla 8.

Resultados de la búsqueda de caracteres asociados a las secuencias aminoacídicas con InterPro

Enzima candidata	Familia	Atributos	Localización (aa)
Endoglucanasa C	-	Dominio de unión a carbohidratos	39-161
		Celulasa N	354-439
		Endo-1,4- β -glucanasa	445-921
		Glicosil hidrolasa familia 9	445-919
		Adhesina de superficie CshA	933-1048
		Ancla a pared celular LPXTG	1077-1109
		Péptido señal Grampositivo	1-34
		Dominio no citoplasmático	35-1086
Endoglucanasa E1	-	Principal endoglucanasa extracelular	38-405
		CBM2	436-539
		Dominio no citoplasmático	34-539
β-glucosidasa	GH1, GH familia 1 β -glucosidasa	Glicosidasa	2-479
		6-fosfo- β -glucosidasa	5-480
β-glucosidasa B	GH1, GH familia 1 β -glucosidasa	Glicosidasa	5-406

Las endoglucanasas tuvieron en común péptidos señal propios del grupo de las bacterias grampositivas, así como dominios de unión a carbohidratos y especificaciones de que sus dominios más grandes no correspondían a componentes citoplasmáticos. De forma individual, para la endoglucanasa C se encontraron dominios correspondientes a componentes relacionados con la membrana, como lo son las adhesinas de superficie, así como un dominio de anclaje a pared

celular. Adicional, una región de su secuencia fue identificada como perteneciente a la familia GH9.

En cuanto a las β -glucosidasas, para ambos casos se determinó que contaban con regiones atribuidas a enzimas pertenecientes a la familia GH1 como algunas firmas o *signatures* de esta familia de proteínas. En el caso específico de la β -glucosidasa B, se halló una región determinada como perteneciente a glucosidasas.

Dados estos resultados, los siguientes pasos de caracterización *in silico* se llevaron a cabo solo con las dos secuencias correspondientes a las dos endoglucanasas. Para dar más detalles de la posible funcionalidad de estas enzimas, una anotación de términos GO (*Gene Ontology*, por sus siglas en inglés) también fue obtenida en la búsqueda a través de InterPro (Tabla 9). Para las dos enzimas, el proceso biológico en el que estarían involucradas sería el metabolismo de carbohidratos y la función molecular que comparten es la hidrólisis de compuestos O-glicosilados. Especificando en cada una de las proteínas, la endoglucanasa C fue anotada con actividad celulasa, pero no mostró otro tipo de funciones como las de unión a carbohidratos y polisacáridos, característica que sí presentó la endoglucanasa E1, así como su inclusión en el proceso biológico de catabolismo de polisacáridos.

Otras estructuras identificadas son los módulos de unión a carbohidratos o CBMs (*Carbohydrate Binding Module*, por sus siglas en inglés), en específico los CBM2 y CBM4, que fueron identificados junto a las probables endoglucanasas de las familias GH5 y GH9, respectivamente, y cuya función reside en unirse a las moléculas de sustrato para lograr acercarlas al dominio catalítico que realizará la hidrólisis.

Tabla 9.*Resultados de la anotación de términos GO con InterPro*

Identificador ensamble	Enzima	Término GO	Código GO
AKHMMEEA_02300	Endoglucanasa C	Metabolismo de carbohidratos	GO:0005975
		Hidrólisis de compuestos O- glicosílicos	GO:0004553
		Actividad celulasa	GO:0008810
AKHMMEEA_02341	Endoglucanasa E1	Metabolismo de carbohidratos	GO:0005975
		Catabolismo de polisacáridos	GO:0000272
		Unión a carbohidratos	GO:0030246
		Hidrólisis de compuestos O- glicosílicos	GO:0004553
		Unión a polisacáridos	GO:0030247

De este modo, hacia el exterior celular se hallarían las endoglucanasas junto a sus estructuras de anclaje a la célula y sus módulos de unión a carbohidratos, con los que se unirían a fibras de celulosa para hidrolizarlas; y al interior celular estarían las β -glucosidasas, esperando el ingreso de los disacáridos obtenidos de la degradación parcial ocurrida en el exterior, y que usarían como sustrato para llevar a cabo la última parte de la catálisis, la obtención de monómeros de glucosa.

Posteriormente, se llevó a cabo la predicción de algunas propiedades (Tabla 10) y estructuras a partir de su composición aminoacídica.

Tabla 10.*Resultados de la predicción de propiedades fisicoquímicas con ProtParam*

Identificador ensamble	Enzima	pI	AI	II	kDa
AKHMMEEA_02300	Endoglucanasa C	4.57	78.34	28.94	115.9
AKHMMEEA_02341	Endoglucanasa E1	7.04	73.47	36.14	56.4

Las secuencias aminoacídicas de estos dos genes fueron procesadas con la herramienta ProtParam del portal web ExPASy; con esto se obtuvieron valores teóricos de algunas propiedades fisicoquímicas. Con referencia al punto isoelectrico (pI), la endoglucanasa C reportó un valor de 4.57 y la E1 7.04. Estos resultados predictivos pueden ser útiles en la medida que se requiera manipular la carga neta de la proteína para lograr su purificación y determinación experimental de MW, entre otros. Aunque no estuviesen dentro del rango óptimo, estos valores teóricos de isoelectricidad permiten apoyar la idea de que las enzimas involucradas en el proceso degradativo de celulosa de la cepa C1 puedan estar adaptadas a un rango de pH desde levemente ácido a un poco más neutro. En cuanto al índice alifático (AI), en la endoglucanasa C fue de 78.34 y en la E1 de 73.47. Este índice, en ambos casos > 70, predice termoestabilidad moderada para las enzimas.

Con respecto al índice de inestabilidad (II), las secuencias de aminoácidos arrojaron valores < 40, indicando un rango considerable de estabilidad para ensayos con condiciones estándar.

Los resultados del número de aminoácidos y el peso molecular de las endoglucanasas C y E1, con 1,110 aa y 115.9 kDa, y 539 aa y 56.4 kDa, respectivamente, están relacionados con información obtenida previamente, como la presencia de CBMs o estructuras con funciones de sujeción similares. Las endoglucanasas, que presentan dominios no catalíticos como estos módulos

de unión a carbohidratos, tienen secuencias más largas y, por ende, pueden presentar pesos moleculares más grandes.

Para predecir probables regiones, motivos o residuos catalíticos se usó la herramienta online ScanProsite (de Castro et al., 2006), identificando regiones de interés a través de búsquedas y alineamientos contra la base de datos de perfiles y patrones de Prosite (Sigrist et al., 2013). Con este escaneo se identificaron motivos conservados de las familias GH5 y GH9. La endoglucanasa E1 presentó el motivo IGADLHNEPH, ajustado al patrón consenso [LIV]-[LIVMFYWGA](2)-[DNEQG]-[LIVMGST]-{SENR}-N-E-[PV]-[RHDNSTLIVFY], en el que el E193 conservado sería el residuo activo. En la endoglucanasa C fue posible identificar tres patrones consenso de la familia GH9. El patrón consenso I es [LVS]-x-[GK]-G-[WFYLM]-[YHF]-D-[ACGS]-G-[DSN]-x(2)-[KMR]-[FAILY]-x-[FWYLQTV]-[APTNS]-[MLGAQS], con el que fue identificado el motivo VSGGWYDAGDHGKYVVNG, siendo el residuo conservado D del extremo C-terminal (D520) el catalizador ácido/base, mientras que el residuo conservado D del extremo N-terminal (D517) estaría determinado como el nucleófilo. El patrón consenso II está definido así: [HLY]-[AILMV]-[FIL]-G-x-[NSTW]-x(2,4)-[SCTV]-[FY]-[LIVMFY]-[SITV]-G-x(1,5)-[GSY]-x(2)-[AFPSTY]-[FLPSV]-x(2)-[AILPQVM]-[HV]-[DHLS]-[KRS], en el cual, una H (histidina) sería el residuo activo, sin embargo, esta no hace parte del motivo identificado YLFGRNALAQSIVTGYGTKYSQNQHSR. El patrón consenso III [FYW]-x-D-x(4)-[FYW]-x(3)-E-x-[STA]-x(3)-N-[STA] presenta los residuos conservados D898 y E907 en el motivo YVDDIGAWSVNEITVNWNS identificado.

La predicción de estructuras secundarias con PSIPRED permitió inferir preliminarmente la composición de dominios, mostrando probables agrupaciones de láminas β , hélices α , y láminas β

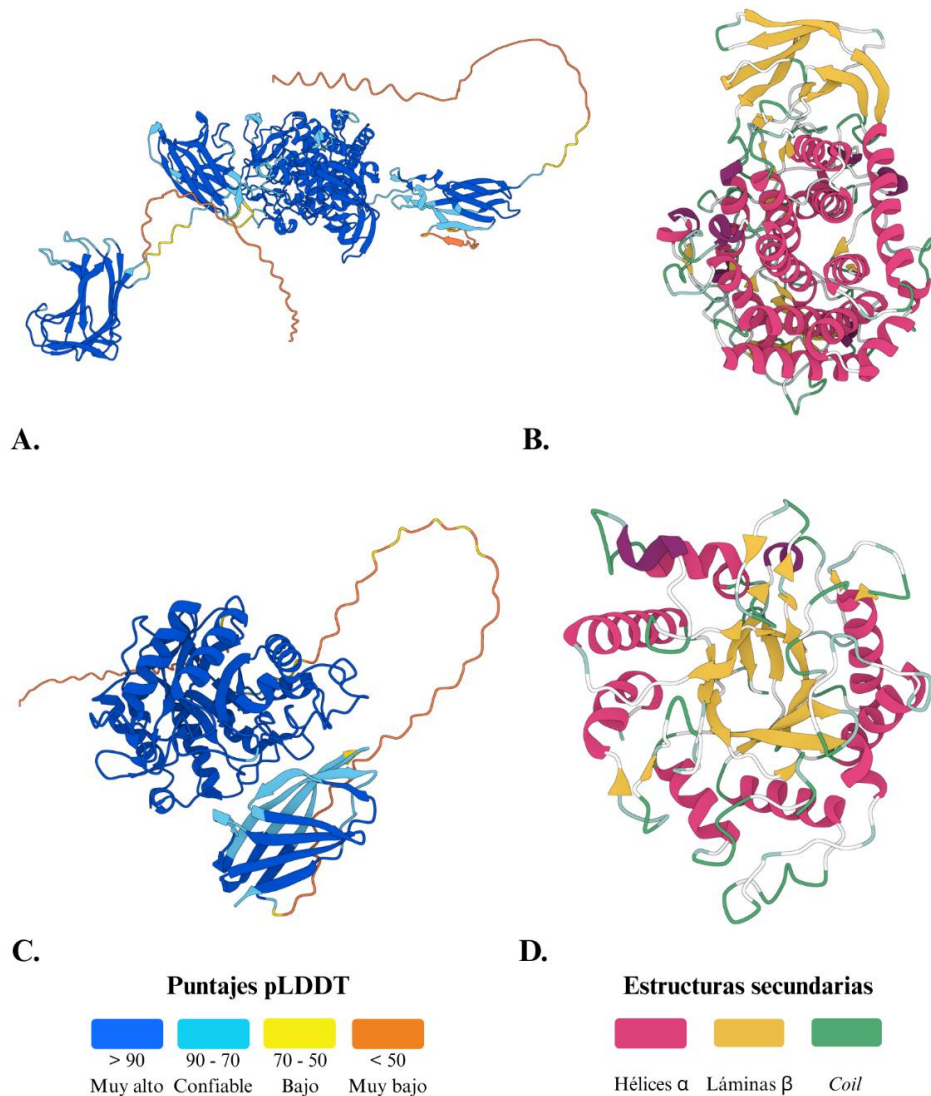
intercaladas con hélices α , además de las regiones *coil*. No obstante, una predicción más fiable requiere modelamiento 3D.

Seguidamente, se continuó el análisis de las secuencias aminoacídicas con la predicción *in silico* de la estructura terciaria de las enzimas. Para lograr dicha predicción se depositaron las secuencias de aminoácidos obtenidas con Prokka en el servidor de AlphaFold (<https://alphafoldserver.com>). Posteriormente, la estructura tridimensional obtenida de cada proteína fue visualizada en la misma plataforma con el fin de analizar su sitio activo y determinar la confiabilidad de la predicción.

La endoglucanasa C mostró un plegamiento tridimensional en forma de barril (α/α)₆, es decir, un túnel interno de 6 hélices α rodeado por 6 hélices α externas unidas de forma alternada por regiones *coil* (Fig. 12B), y tres dominios no catalíticos adyacentes con forma de bloque de láminas β (Fig. 12A); caso diferente al de la endoglucanasa E1, que mostró una conformación de barril (β/α)₈, es decir, un túnel de 8 láminas β internas rodeadas hacia el exterior por 8 hélices α y a su vez unidas alternadamente a través de regiones *coil* (Fig. 12D), junto a un solo dominio no catalítico, también un bloque de láminas β . (Fig. 12C).

Figura 12.

Modelamiento tridimensional de las dos endoglucanasas de C1 con AlphaFold



Nota. Visualización de los modelos predichos *in silico* de las endoglucanasas C y E1 de la cepa C1. **A:** puntaje pLDDT de la estructura terciaria de la endoglucanasa C, **B:** composición de hélices α y láminas β del sitio activo de la endoglucanasa C, **C:** puntaje pLDDT de la estructura terciaria de la endoglucanasa E1, **D:** composición de hélices α y láminas β del sitio activo de la endoglucanasa E1.

AlphaFold implementa una prueba de distancia para determinar en una escala de 0 – 100 la confiabilidad del modelo, sin embargo, dicha prueba no es global, sino local, razón por la cual es denominada pLDDT (*predicted-Local Distance Difference Test*, por sus siglas en inglés). Con el fin de visualizar la confiabilidad de los plegamientos se utilizó el puntaje pLDDT, que asigna un código de colores según la clasificación de confiabilidad asignada a cada residuo aminoacídico.

En el caso de la Endoglucanasa C, se observó una composición diversa en términos de los puntajes pLDDT. Tanto el extremo C-terminal como el N-terminal presentaron regiones internas de baja confiabilidad y regiones *coil* de muy baja confiabilidad. No obstante, de forma general, tanto el dominio catalítico como los no catalíticos fueron clasificados con una confiabilidad predominantemente alta (Fig. 12A). Para la endoglucanasa E1 prácticamente todo su dominio catalítico fue predicho de una forma altamente confiable, y su dominio no catalítico mostró confiabilidad alta y muy alta, a diferencia de las regiones *coil* de los extremos y que unen ambos dominios proteicos, con una confiabilidad muy baja (Fig. 12C).

5. Discusión

La caracterización de la cepa C1 aislada del intestino de *D. saccharalis* confirma de manera consistente su capacidad para producir enzimas celulasas, respondiendo directamente al objetivo general de este trabajo. La observación de halos de degradación en placas suplementadas con CMC y teñidas con Rojo Congo constituyen evidencia cualitativa robusta de la actividad celulasa *in vivo*, además de ser un indicador de la secreción de celulasas extracelulares (Hossain et al., 2020; Meddeb-Mouelhi et al., 2014), yendo en línea con hallazgos de cepas de *Microbacterium* que

presentan genes relacionados con esta actividad enzimática o que tienen actividad celulasa confirmada (An et al., 2024).

El pico de producción celulasa de C1 a 24 h de cultivo presenta una ventaja explotable biotecnológicamente. Cepas de *Microbacterium* se han seleccionado bajo este criterio para su caracterización e implementación industrial (Islam & Roy, 2018; Ugras et al., 2024), demostrando la importancia de encontrar una enzima con catálisis rápida.

La corroboración de la actividad celulasa *in vitro* con C1, al igual que el estudio de da Silva et al. (2021), mostró su predominancia en el sobrenadante con respecto a las fracciones soluble e insoluble de los extractos crudos, al tiempo que la pérdida de dicha actividad tras la inactivación térmica de los precipitados proteicos reafirmó la naturaleza enzimática de la catálisis y soporta la elección del sobrenadante para la caracterización enzimática.

La cuantificación espectrofotométrica de la actividad enzimática, realizada mediante el método DNS, reveló que la fracción con mayor actividad específica corresponde al sobrenadante (5.24 nkat/mg), alcanzando valores aproximadamente 20 veces superiores a los de las fracciones insoluble (0.15 nkat/mg) y soluble (0.25 nkat/mg). Esta diferencia confirma el carácter extracelular de la enzima e indica la fracción a la que se deben dirigir los esfuerzos de caracterización. Aunque celulasas de *Microbacterium* han mostrado 10 nkat/mg (Li, Y. et al., 2019), enzimas GH optimizadas y purificadas han mostrado hasta 97.2 nkat/mg (Asha et al., 2012; Kim, Sang Jin et al., 2023), sugiriendo que la celulasa de C1 pueda incrementar considerablemente su actividad después de someterse a estos procedimientos. La actividad enzimática medida en el sobrenadante también superó a la observada en los controles negativos e inactivaciones de las otras fracciones, donde los valores se mantuvieron alrededor de 0.06 – 0.15 nkat/mg, por lo que artefactos asociados a la metodología de cuantificación se descartan.

En la literatura, los valores de actividad en bacterias celulolíticas varían según el sistema de medición y el sustrato empleado, incluyendo cepas de *Paenibacillus* con 1.5 – 3 nkat/mg (Doan et al., 2024; Islam & Roy, 2018) y de otros géneros bacterianos con cifras de actividad entre los 0.009 – 0.7 nkat/mg (Menshaway et al., 2022; Yin et al., 2010). La actividad específica registrada para la cepa C1 (5.24 nkat/mg) corresponde a su extracto crudo y no a una fracción optimizada o purificada, como es el caso de algunas de las enzimas comparadas, por lo que esta, aun siendo baja, podría aumentar considerablemente al implementarse la purificación proteica parcial o total, sugiriendo un potencial biotecnológico relevante, aunque por debajo de los sistemas más eficientes reportados. Estos contrastes demuestran la necesidad de estandarizar la expresión de resultados en unidades comparables, idealmente actividad específica, para posibilitar evaluaciones más precisas entre especies y condiciones experimentales, sugiriendo la implementación de la unidad SI (Sistema Internacional), el katal (kat).

La disonancia entre las fracciones analizadas despertó interrogantes sobre estos resultados al observar que mientras la actividad enzimática se concentró principalmente en el sobrenadante, la mayor acumulación de glucosa libre se detectó en la fracción insoluble. Esta aparente contradicción puede responder a factores como la acumulación intracelular de glicógeno como reserva energética (Montero et al., 2009; Wilson et al., 2010; Zmasek & Godzik, 2014), cuya eventual hidrólisis liberaría glucosa detectable, al igual que con trazas de CMC retenidas en el *pellet* del lisado. Maquinaria enzimática para el metabolismo del glicógeno se ha encontrado en *Microbacterium* (Lenchi, 2023), por lo que el hallazgo de un gen de C1 que corresponde parcialmente a una enzima desramificadora del glicógeno (PDB: 7U39; Identidad 42%) acreditaría esta hipótesis. No obstante, fenómenos como la difusión lumínica en suspensiones turbias (Myers et al., 2013; Swanson et al., 1999) pueden presentar interferencias metodológicas considerables en

el ensayo DNS, dada la calidad heterogénea de la fracción insoluble que influiría en la absorbancia medida. La incorporación de controles adicionales como la clarificación de lisados, o técnicas avanzadas de cuantificación como el HPLC (*High Performance Liquid Chromatography*, por sus siglas en inglés) cobran gran importancia para discriminar entre artefactos analíticos y procesos biológicos genuinos.

Las condiciones óptimas de reacción registradas (pH 6; 50 °C) se acercan a lo esperado para enzimas de bacterias mesófilas. En cepas de *Bacillus* se han hallado enzimas termófilas de alta actividad (pH 7; 50 °C), termófilas neutras (pH 7; 50 °C) y termófilas levemente acidófilas (pH 6; 50 °C) (Irfan et al., 2017; Padilha et al., 2015; Ugras et al., 2024), mientras que enzimas de Actinomycetes se han caracterizado con actividades óptimas entre pH 5 – 8 y 50 – 55 °C, como pH 7 y 25 °C, siendo esta última una enzima criogénica (Das et al., 2022; Huang et al., 2015; Khadka et al., 2022; Kulkarni & Maurya, 2017; Li, Y. et al., 2019) . Con esto, la celulasa de C1 podría clasificarse como potencial celulasa termófila levemente acidófila. La rápida liberación de producto en los primeros minutos de incubación constituye una ventaja potencial en procesos que requieren tasas de hidrólisis inicial elevadas, sin embargo, la disminución posterior de la actividad señala la necesidad de evaluar con mayor detalle su termoestabilidad y persistencia catalítica en períodos prolongados. Así, se recomienda realizar estudios adicionales de estabilidad frente a variaciones de pH y temperatura para establecer la aplicabilidad de estas enzimas en entornos industriales, particularmente el nacional.

Los parámetros cinéticos determinados aportan información valiosa sobre la interacción de las enzimas con el sustrato. En *Microbacterium*, celulasas con K_m de 2.12 g/L (celulosa) y V_{max} de 255.5 nkat/mg; y K_m de 1.2 g/L (celobiosa) y V_{max} de 753.67 nakt/mg se han reportado, señalando afinidad por sus sustratos y alta velocidad de reacción (Lin et al., 2019; Purohit et al., 2023). El

valor de K_m de la celulasa de C1 (0.26 g/L) sugiere una afinidad elevada hacia la CMC, indicando su amplia capacidad de unión al sustrato aun en extractos crudos, mientras que su V_{max} (23.23 nkat/mg) podría elevarse con la implementación de la purificación enzimática. No obstante, la interpretación de estos parámetros depende fuertemente del tipo de sustrato empleado, la pureza enzimática y el método de ajuste utilizado. Teniendo en cuenta que los sustratos difieren en accesibilidad y mecanismos de reconocimiento, comparaciones entre sus enzimas hidrolíticas deben hacerse con cautela, por lo que el empleo de ajustes no lineales a curvas de velocidad inicial, junto a comparaciones de rendimiento con varios sustratos resultarían en estimaciones cinéticas más confiables y comparables.

Desde el punto de vista genómico, el ensamblaje de la cepa C1 mostró métricas de calidad aceptables para un *draft genome* basado en *short-reads* (N50 de 358 kb, Q medio \approx 36). Aunque el umbral Q de alta calidad está fijado en 30 (Andrews, s.f.) y la poda no estricta de pares de base con baja calidad ($Q < 30$) se recomienda (Del Fabbro et al., 2013; Mbandi et al., 2014; Paszkiewicz & Studholme, 2010), el Q36 medio de este caso permite omitir esta recomendación. Por otro lado, la elección del ensamble con base en la métrica N50, como lo recomiendan múltiples autores (Chen et al., 2015; Melnikova et al., 2021; Soueidan et al., 2013), permitió la contigüidad necesaria del ensamble para realizar una anotación funcional del genoma. A modo comparativo, el *draft genome* de C1 mostró ser \sim 125 kpb más pequeño que la media de un estudio con 10 cepas de *Microbacterium* (Corretto et al., 2015), sin embargo, con un L50 de 4 con respecto a la media de 7 de la comparación, el ensamble de C1 evidenció mejor contigüidad de la secuencia. Ensamblajes de muy alta calidad con genomas de más de 3 Mpb, pocos *contigs*, N50 altos y L50 bajos se han obtenido con cepas de *Microbacterium* (Lenchi, 2023), demostrando que esta combinación de parámetros es fiable para elegir ensambles de alta calidad.

El *polishing* con Pilon puede darle más contiguidad, integridad y precisión a los ensamblajes (Sigova et al., 2024; Walker et al., 2014), no obstante, la inclusión de un paso de *polishing* con Pilon por parte de Unicycler (Luan et al., 2024) permite confiar en los resultados de alta calidad del ensamblaje *de novo*, dadas las mejoras como la resolución de regiones repetitivas (RR), puenteo de contigs (*contig bridging*), optimización de selección de k-meros, entre otras que el ensamblador implementa, dándole confiabilidad para la obtención de *draft genomes* basados en *short-reads* de alta calidad (Wick et al., 2017). Aun así, el abordaje mediante un ensamblaje híbrido con *long-reads* representaría una ventaja dadas sus características mejoradas (Merda et al., 2024; Trisakul et al., 2024; Zhao et al., 2025), permitiendo identificar posibles regiones conservadas o clústeres de genes con mayor precisión, algo esencial para comprender la regulación y el potencial de producción enzimática. En este contexto, la integración de análisis transcriptómicos y proteómicos será clave para discriminar entre genes funcionalmente activos y pseudogenes, así como posibles rutas bioquímicas asociadas a esta actividad.

Los bajos porcentajes de dDDH impidieron asignar una especie a la cepa C1, sin embargo, gracias a su similitud con más de 10 cepas tipo de *Microbacterium*, su asignación taxonómica cobra validez suficiente.

La anotación funcional reveló la presencia de genes pertenecientes a familias GH1 (UniProt: A0A919KNQ2; A0A150HHU3), GH5 (UniProt: A0A150HEX2) y GH9 (UniProt: A0A9W6HIW8) (Bernier & Bernard, 2016; Wu, L. & Ma, 2019), constituyendo un soporte adicional a la actividad celulasa observada. La anotación de *CAZymes* con dbCAN ha conducido al hallazgo de genes de las familias GH1 y GH5 compartidos entre decenas de cepas de *Microbacterium* (An et al., 2024; Gupta et al., 2022), dándole robustez a la predicción de dbCAN para C1. Agregando, la asociación de genes codificantes de celulasas de C1 con términos GO

asignados a la degradación de carbohidratos corroboran su potencial como codificadores de biocatalizadores.

El alto AI de 78.34 y el bajo II de 28.94 predichos para la endoglucanasa C estarían indicando termoestabilidad de la proteína y estabilidad para ensayos *in vitro* (Gasteiger, Elisabeth et al., 2005; Mohamad Sobri et al., 2020), respectivamente, mientras que sus 115.92 kDa indicarían la presencia de varios dominios no catalíticos, dados otros reportes de enzimas GH9 (Li, F. et al., 2020; Phitsuwan et al., 2021), y su pI (4.57) daría indicios de sus condiciones de estabilidad isoelectrónica. La endoglucanasa E1, con su pI de 7.04 difiere de otras enzimas estudiadas con pI más altos o bajos, mientras que su tamaño de 56.4 kDa es similar al de dichas proteínas (Ma et al., 2020; Sanjaya et al., 2021). Su AI de 73.47 y su II de 36.14 indicarían termoestabilidad y estabilidad para ensayos *in vitro* (Gasteiger, E. et al., 2003; Mohamad Sobri et al., 2020; Sanjaya et al., 2021).

Estas características muestran una enzima de gran tamaño (C) y una común (E1), con distintos pI que podrían hacerlas diferir en sus condiciones de estabilidad. De esta forma, los índices alifáticos (> 70) y de inestabilidad (< 40) que hablan de termoestabilidad y estabilidad podrían relacionarse con el rango de máxima actividad entre 40 y 60 °C con pico en 50 °C, obtenido experimentalmente.

Con base en la estructura secundaria, agrupaciones de láminas β hacia los extremos de las secuencias de las endoglucanasas C y E1 predicen la presencia de CBMs, mientras que la alternancia entre hélices α , y hélices α y láminas β , respectivamente, en el *core* de la secuencia, harían referencia a sus sitios catalíticos. El patrón consenso del motivo catalítico de GH5 (<https://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS00659>) fue identificado en la endoglucanasa E1 de C1, en el cual, el glutamato conservado sería un residuo catalítico; a pesar

de ello, la ausencia del segundo residuo difiere de reportes previos que mencionan dos glutamatos catalíticos (Henrissat et al., 1995; Sanjaya et al., 2021). La secuencia de la endoglucanasa C presentó las tres regiones consenso de su familia (<https://prosite.expasy.org/cgi-bin/prosite/nicedoc.pl?PS60032>), reportadas también en otro trabajo (Guerriero et al., 2018), y de las cuáles, la primera posee el motivo conservado DAGD, y la tercera el glutamato catalítico junto a un aspartato que alinea correctamente las tres regiones; no obstante, la H (histidina) de la región II se vio reemplazada por una serina. La detección de sustituciones en regiones funcionales plantea interrogantes sobre su impacto en la afinidad y procesividad de la enzima, dado el cambio de residuos que actúan como soporte o sitio de unión con el sustrato. Estas predicciones requieren ser validadas experimentalmente mediante análisis proteómico del secretoma, zimogramas, y la expresión heteróloga de los genes candidatos.

La obtención del plegamiento tipo *TIM-barrel* (β/α)₈ (*Triosephosphate Isomerase-barrel*, por su acrónimo en inglés) en la predicción tridimensional de la endoglucanasa E1 concuerda con cristalizaciones previas de enzimas GH5 (Couturier et al., 2013; Garg & Manoj, 2023; Glasgow et al., 2020; Lafond et al., 2016). Dicho plegamiento comprende un túnel interno de láminas β unidas por *loops* a hélices α exteriores y sus dos residuos catalíticos se han localizado en las láminas β 4 y 7 del sitio activo (Couturier et al., 2013; Garg & Manoj, 2023; Glasgow et al., 2020; Jenkins et al., 1995). De igual forma, el barril tipo (α/α)₆ predicho por AlphaFold para la endoglucanasa C, y típico de la familia GH9 (Ketudat & Esen, 2010) tiene correspondencia con cristalizaciones hechas a enzimas de esta familia previamente (Honda et al., 2016; Petkun et al., 2015). El respaldo de la predicción, gracias a los altos valores de confiabilidad (pLDDT), se puede incrementar con las conformaciones catalíticas reportadas de las familias GH5 (<https://www.cazy.org/GH5.html>) y GH9 (<https://www.cazy.org/GH9.html>) en la rigurosamente curada CAZy *database* (Drula et al.,

2022; Lombard et al., 2014), aportando evidencia a la caracterización y clasificación de las celulasas extracelulares de C1, sin olvidar que la corroboración experimental con cristalografía es requerida.

En conjunto, este estudio constituye una aproximación integral que combina evidencia fenotípica, cinética, genómica e *in silico* para caracterizar dos de las celulasas de *Microbacterium* C1. Si bien los resultados respaldan el potencial de la cepa C1 como fuente novedosa de celulasas, la aplicación industrial de sus biocatalizadores aún enfrenta desafíos críticos. Entre ellos destacan la purificación y caracterización de isoenzimas individuales, la validación experimental de las anotaciones bioinformáticas, la estandarización de métricas de actividad y la evaluación del desempeño sobre sustratos lignocelulósicos complejos. Superar estas limitaciones no solo permitirá confirmar la aplicabilidad de estas enzimas en bioprocesos industriales, sino que también ampliará el conocimiento sobre el papel de actinobacterias intestinales en la degradación de polisacáridos, abriendo nuevas perspectivas y opciones en biotecnología microbiana en Colombia.

6. Conclusiones

La caracterización de las propiedades estructurales, fisicoquímicas y funcionales de una enzima celulasa aislada de la cepa C1 fue llevada a cabo mediante el uso de las aproximaciones bioquímica e *in silico*. La cepa C1 identificada previamente por su potencial calidad degradativa de celulosa evidenció esta actividad hidrolítica en la detección cualitativa con Rojo Congo, dando indicios preliminares sobre su función extracelular. Dado ello, la obtención de los extractos crudos y su posterior análisis de actividad corroboraron que el origen de la hidrólisis sería enzimático y que esta se concentraría mayormente en el sobrenadante, abriendo la puerta a la optimización de

la catálisis paso a paso, que desembocó en el hallazgo de una proteína con actividad máxima en los primeros 10 minutos de reacción a condiciones de acidez leve y temperaturas altas, pero susceptible a temperaturas extremas. La determinación de una elevada afinidad por la CMC y velocidad máxima relativamente baja por parte de la proteína respondieron a la obtención de la cinética enzimática, destacando que su purificación podría incrementar considerablemente la velocidad máxima de la enzima. Con base en el *draft genome* de C1 fue posible asignar esta cepa de manera preliminar al género *Microbacterium*, anotar funcionalmente su genoma e identificar los potenciales genes codificantes de dos endoglucanasas y dos β -glucosidasas. Enfatizando en las endoglucanasas, su caracterización *in silico* arrojó resultados similares con las características determinadas *in vitro* en el laboratorio, evidenciando la ventaja de usar herramientas bioinformáticas, sin descartar su requerida comprobación experimental. La identificación de la pertenencia de las enzimas a las familias GH5 y GH9 permitió contrastar la predicción *in silico* de la estructura tridimensional de las proteínas con reportes previos para estas familias, poniendo de manifiesto la utilidad y fiabilidad de los avances predictivos en el modelamiento de proteínas, no obstante, la corroboración con cristalografía de rayos X es vital para determinar con certeza la conformación espacial de estas celulasas.

De este modo, la caracterización bioinformática y bioquímica de dos enzimas aisladas del tracto digestivo del barrenador de la caña constituyen un avance significativo en la búsqueda de biocatalizadores provenientes de la biodiversidad nacional, con el potencial de consolidarse como referentes degradativos e innovadores en la industria una vez las etapas críticas de su estudio y mejoramiento sean superadas. En el mismo sentido, la ampliación de la información con respecto a las cualidades degradativas de estas enzimas se hace necesaria en cuanto este trabajo representa una primera aproximación a las características de estas celulasas.

Referencias Bibliográficas

- Alikhan, N., Petty, N. K., Ben Zakour, N. L., & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, *12*(1), 402–402. 10.1186/1471-2164-12-402
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. 10.1016/S0022-2836(05)80360-2
- Alzari, P. M., ne Souchon, H., & Dominguez, R. (1996). The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure*, *4*(3), 265–275. 10.1016/S0969-2126(96)00031-7
- An, P., Yang, C., Li, W., Zhao, D., & Xiang, H. (2024). The Isolation and Characterization of a Novel Psychrotolerant Cellulolytic Bacterium, *Microbacterium* sp. QXD-8T. *Microorganisms (Basel)*, *12*(2), 303. 10.3390/microorganisms12020303
- Andrews, S. (s.f.). *FastQC A Quality Control tool for High Throughput Sequence Data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Asha, B. M., Revathi, M., Yadav, A., & Sakthivel, N. (2012). Purification and Characterization of a Thermophilic Cellulase from a Novel Cellulolytic Strain, *Paenibacillus barcinonensis*. *Journal of Microbiology and Biotechnology*, *22*(11), 1501–1509. <http://dx.doi.org/10.4014/jmb.1202.02013>
- Atehortúa-Bueno, M. (2022). *Estructura y características del bacterioma intestinal del gusano barrenador de la caña Diatraea saccharalis (Lepidoptera: Crambidae)*. <https://noesis.uis.edu.co/items/f9f358d5-9b6c-4859-a19b-930c8850f1bf>

- Ausland, C., Zheng, J., Yi, H., Yang, B., Li, T., Feng, X., Zheng, B., & Yin, Y. (2021). dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Research*, *49*(D1), D523–D528. 10.1093/nar/gkaa742
- Bano, A., Chen, X., Prasongsuk, S., Akbar, A., Lotrakul, P., Punnapayak, H., Anwar, M., Sajid, S., & Ali, I. (2019). Purification and Characterization of Cellulase from Obligate Halophilic *Aspergillus flavus* (TISTR 3637) and Its Prospects for Bioethanol Production. *Applied Biochemistry and Biotechnology*, *189*(4), 1327–1337. 10.1007/s12010-019-03086-y
- Başkan, K. S., Tütem, E., Akyüz, E., Özen, S., & Apak, R. (2016). Spectrophotometric total reducing sugars assay based on cupric reduction. *Talanta*, *147*, 162–168. 10.1016/j.talanta.2015.09.049
- Bernier, A., & Bernard, K. (2016). Draft Genome Sequences of *Microbacterium hominis* LCDC-84-0209T Isolated from a Human Lung Aspirate and *Microbacterium laevaniformans* LCDC 91-0039 Isolated from a Human Blood Culture. *Genome Announcements (Washington, DC)*, *4*(5)10.1128/genomeA.00989-16
- Bhat, M. K. (2000). Cellulases and related enzymes in biotechnology. *Biotechnology Advances*, *18*(5), 355–383. 10.1016/S0734-9750(00)00041-0
- Blum, M., Andreeva, A., Florentino, L., Chuguransky, S., Grego, T., Hobbs, E., Pinto, B., Orr, A., Paysan-Lafosse, T., Ponamareva, I., Salazar, G., Bordin, N., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D., Letunic, I., Llinares-López, F., . . . Bateman, A. (2025). InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Research*, *53*(D1), D444–D456. 10.1093/nar/gkae1082

- Brodersen, R., & Ricketts, H. T. (1949). Evaluation of a modified Sumner's method (dinitrosalicylic acid) for determination of glucose in urine. *The Journal of Laboratory and Clinical Medicine*, 34(10), 1447–1456. <https://www.ncbi.nlm.nih.gov/pubmed/18143609>
- Cerda-Mejía, L. (2016). *Enzimas modificadoras de la pared celular vegetal. Celulasas de interés biotecnológico* papelero.
https://diposit.ub.edu/dspace/bitstream/2445/104248/1/LCM_TESIS.pdf
- Chandel, A. K., Chandrasekhar, G., Silva, M. B., & Silvério da Silva, S. (2012). The realm of cellulases in biorefinery development. *Critical Reviews in Biotechnology*, 32(3), 187–202. 10.3109/07388551.2011.595385
- Chen, T., Gan, R., Chang, Y., Liao, W., Wu, T. H., Lee, C., Huang, P., Lee, C., Chen, Y. M., Chiu, C., & Tang, P. (2015). Is the whole greater than the sum of its parts? De novo assembly strategies for bacterial genomes based on paired-end sequencing. *BMC Genomics*, 16(1), 648. 10.1186/s12864-015-1859-8
- Corretto, E., Antonielli, L., Sessitsch, A., Kidd, P., Weyens, N., & Brader, G. (2015). Draft Genome Sequences of 10 *Microbacterium* spp., with Emphasis on Heavy Metal-Contaminated Environments. *Genome Announcements* (Washington, DC), 3(3)10.1128/genomeA.00432-15
- Couturier, M., Roussel, A., Rosengren, A., Leone, P., Stålbrand, H., & Berrin, J. (2013). Structural and Biochemical Analyses of Glycoside Hydrolase Families 5 and 26 β -(1,4)-Mannanases from *Podospira anserina* Reveal Differences upon Manno-oligosaccharide Catalysis. *The Journal of Biological Chemistry*, 288(20), 14624–14635. 10.1074/jbc.M113.459438

- da Silva, R. N., Melo, L. F. d. A., & Luna Finkler, C. L. (2021). Optimization of the cultivation conditions of *Bacillus licheniformis* BCLLNf-01 for cellulase production. *Biotechnology Reports (Amsterdam, Netherlands)*, 29, e00599. 10.1016/j.btre.2021.e00599
- Dantur, K. I., Enrique, R., Welin, B., & Castagnaro, A. P. (2015). Isolation of cellulolytic bacteria from the intestine of *Diatraea saccharalis* larvae and evaluation of their capacity to degrade sugarcane biomass. *AMB Express*, 5(1), 15. 10.1186/s13568-015-0101-z
- Das, T., Ali, F., & Rahman, M. S. (2022). Cellulase activity of a novel bacterial strain *Arthrobacter woluwensis* TDS9: its application on bioconversion of paper mill sludge. *Journal of Genetic Engineering and Biotechnology*, 20(1), 87. 10.1186/s43141-022-00373-w
- Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, 3, 853–859.
<https://www.sciencedirect.com/science/article/pii/S0969212601002209?via%3Dihub>
- de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., & Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, 34(suppl-2), W362–W365. 10.1093/nar/gkl124
- Del Fabbro, C., Scalabrin, S., Morgante, M., & Giorgi, F. M. (2013). An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLoS ONE*, 8(12), e85024. 10.1371/journal.pone.0085024
- Doan, C. T., Tran, T. N., Pham, T. P., Tran, T. T. T., Truong, B. P., Nguyen, T. T., Nguyen, T. M., Bui, T. Q. H., Nguyen, A. D., & Wang, S. (2024). Production, Purification, and Characterization of a Cellulase from *Paenibacillus elgii*. *Polymers*, 16(14), 2037. 10.3390/polym16142037

- Drula, E., Garron, M., Dogan, S., Lombard, V., Henrissat, B., & Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research*, *50*(D1), D571–D577. 10.1093/nar/gkab1045
- Eadie, G. S. (1942). The Inhibition of Cholinesterase by Physostigmine and Prostigmine. *Journal of Biological Chemistry*, *146*(1), 85–93. 10.1016/S0021-9258(18)72452-6
- Froimowitz, M. (1993). HyperChem: a software package for computational chemistry and molecular modeling. *BioTechniques*, *14*(6), 1010–3.
- Garg, P., & Manoj, N. (2023). Structure of an iminosugar complex of a glycoside hydrolase family 5 lichenase provides insights into the active site. *Biochimie*, *204*, 69–77. 10.1016/j.biochi.2022.09.001
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R., & Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13), 3784–3788. 10.1093/nar/gkg563
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press. 10.1385/1-59259-890-0:571
- Geourjon, C., & Deléage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*, *11*(6), 681–684. 10.1093/bioinformatics/11.6.681
- Glasgow, E. M., Kemna, E. I., Bingman, C. A., Ing, N., Deng, K., Bianchetti, C. M., Takasuka, T. E., Northen, T. R., & Fox, B. G. (2020). A structural and kinetic survey of GH5_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for

biomass hydrolysis. *The Journal of Biological Chemistry*, 295(51), 17752. 10.1074/jbc.RA120.015328

Gobierno de la República de Colombia. (2019). *Estrategia nacional de economía circular. Cierre de ciclos de materiales, innovación tecnológica, colaboración y nuevos modelos de negocio*. Presidencia de la República; Ministerio de Ambiente y Desarrollo Sostenible; Ministerio de Comercio, Industria y Turismo.

Gray, K. A., Zhao, L., & Emptage, M. (2006). Bioethanol. *Current Opinion in Chemical Biology*, 10(2), 141–146. 10.1016/j.cbpa.2006.02.035

Guerriero, G., Sergeant, K., Legay, S., Hausman, J., Cauchie, H., Ahmad, I., & Siddiqui, K. S. (2018). Novel Insights from Comparative In Silico Analysis of Green Microalgal Cellulases. *International Journal of Molecular Sciences*, 19(6), 1782. 10.3390/ijms19061782

Gupta, S., Han, S., Kim, B., Lee, C., & Oh, T. (2022). Comparative analysis of genome-based CAZyme cassette in Antarctic *Microbacterium* sp. PAMC28756 with 31 other *Microbacterium* species. *Genes & Genomics*, 44(6), 733–746. 10.1007/s13258-022-01254-9

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J., & Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences*, 92(15), 7090–7094. 10.1073/pnas.92.15.7090

Hoda, A., Tafaj, M., & Sallaku, E. (2021). In silico Structural, Functional and Phylogenetic Analyses of cellulase from *Ruminococcus albus*. *Journal of Genetic Engineering and Biotechnology*, 19(1), 58. 10.1186/s43141-021-00162-x

- Hofstee, B. H. J. (1959). Non-Inverted Versus Inverted Plots in Enzyme Kinetics. *Nature*, *184*(4695), 1296–1298. 10.1038/1841296b0
- Honda, Y., Arai, S., Suzuki, K., Kitaoka, M., & Fushinobu, S. (2016). The crystal structure of an inverting glycoside hydrolase family 9 exo- β -D-glucosaminidase and the design of glycosynthase. *Biochemical Journal*, *473*(4), 463–472. 10.1042/BJ20150966
- Hossain, T. J., Chowdhury, S. I., Mozumder, H. A., Chowdhury, M. N. A., Ali, F., Rahman, N., & Dey, S. (2020). Hydrolytic Exoenzymes Produced by Bacteria Isolated and Identified from the Gastrointestinal Tract of Bombay Duck. *Frontiers in Microbiology*, *11*, 2097. 10.3389/fmicb.2020.02097
- Huang, S., Deng, G., Yang, Y., Wu, Z., & Wu, L. (2015). Optimization of Endoglucanase Production from a Novel Bacterial Isolate, *Arthrobacter* sp. HPG166 and Characterization of its Properties. *Brazilian Archives of Biology and Technology*, *58*(5), 692–701. 10.1590/S1516-89132015050256
- Irfan, M., Mushtaq, Q., Tabssum, F., Shakir, H. A., & Qazi, J. I. (2017). Carboxymethyl cellulase production optimization from newly isolated thermophilic *Bacillus subtilis* K-18 for saccharification using response surface methodology. *AMB Express*, *7*(1), 29. 10.1186/s13568-017-0331-3
- Islam, F., & Roy, N. (2018). Screening, purification and characterization of cellulase from cellulase producing bacteria in molasses. *BMC Research Notes*, *11*(1), 445. 10.1186/s13104-018-3558-4
- Jenkins, J., Lo Leggio, L., Harris, G., & Pickersgill, R. (1995). β -Glucosidase, β -galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold β/α architecture and with two conserved glutamates near the carboxy-

terminal ends of β -strands four and seven. *FEBS Letters*, 362(3), 281–285. 10.1016/0014-5793(95)00252-5

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. 10.1038/s41586-021-03819-2

Ketudat, J. R., & Esen, A. (2010). β -Glucosidases. *Cellular and Molecular Life Sciences: CMLS*, 67(20), 3389–3405. 10.1007/s00018-010-0399-2

Khadka, S., Khadka, D., Poudel, R. C., Bhandari, M., Baidya, P., Sijapati, J., & Maharjan, J. (2022). Production Optimization and Biochemical Characterization of Cellulase from *Geobacillus* sp. KP43 Isolated from Hot Spring Water of Nepal. *BioMed Research International*, 2022(1), 6840409. 10.1155/2022/6840409

Kim, S. J., Shin, K., Kim, D. W., Kim, Y., & Park, C. (2023). Cloning and Characterization of Cellulase from *Paenibacillus peoriae* MK1 Isolated from Soil. *Fermentation (Basel)*, 9(10), 873. 10.3390/fermentation9100873

Kim, S., Lee, C., Han, B., Kim, M., Yeo, Y., Yoon, S., Koo, B., & Jun, H. (2008). Characterization of a gene encoding cellulase from uncultured soil bacteria. *FEMS Microbiology Letters*, 282(1), 44–51. 10.1111/j.1574-6968.2008.01097.x

Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A., & Blanch, H. W. (2012). The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and Bioengineering*, 109(4), 1083–1087. 10.1002/bit.24370

- Kuhad, R. C., Gupta, R., & Singh, A. (2011). Microbial Cellulases and Their Industrial Applications. *Enzyme Research*, 2011, 1–10. 10.4061/2011/280696
- Kulkarni, C. P., & Maurya, C. B. (2017). Characterization of the Cellulase Enzyme Produced by Actinomycetes Isolated from the Mangrove Coastal Areas. *Biosciences, Biotechnology Research Asia*, 14(2), 685–690. 10.13005/bbra/2495
- Kumar, R., Singh, S., & Singh, O. V. (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of Industrial Microbiology & Biotechnology*, 35(5), 377–391. 10.1007/s10295-008-0327-8
- Kwon, Y., & Jewett, M. C. (2015). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Scientific Reports*, 5(1), 8663. 10.1038/srep08663
- Lafond, M., Sulzenbacher, G., Freyd, T., Henrissat, B., Berrin, J., & Garron, M. (2016). The Quaternary Structure of a Glycoside Hydrolase Dictates Specificity toward β -Glucans. *The Journal of Biological Chemistry*, 291(13), 7183–7194. 10.1074/jbc.M115.695999
- Lenchi, N. (2023). Draft Whole-Genome Sequences of *Microbacterium oxydans* and *Microbacterium maritypicum* Strains. *Microbiology Resource Announcements*, 12(1), e0108922. 10.1128/mra.01089-22
- Li, C., Lin, F., Li, Y., Wei, W., Wang, H., Qin, L., Zhou, Z., Li, B., Wu, F., & Chen, Z. (2016). A β -glucosidase hyper-production *Trichoderma reesei* mutant reveals a potential role of cel3D in cellulase production. *Microbial Cell Factories*, 15(1), 151. 10.1186/s12934-016-0550-3
- Li, F., Dong, J., Lv, X., Wen, Y., & Chen, S. (2020). Recombinant expression and characterization of two glycoside hydrolases from extreme alkalinophilic bacterium *Cellulomonas bogoriensis* 69B4T. *AMB Express*, 10(1), 44–10. 10.1186/s13568-020-00979-8

- Li, Y., Wang, Z., Zhou, Y., Zhu, G., & Lin, L. (2019). Enzymatic identification and functional sites study of a novel cold-active cellulase (MkCel5) from *Microbacterium kitamiensea*. *Biotechnology, Biotechnological Equipment*, 33(1), 739–747. 10.1080/13102818.2019.1612278
- Lin, L., Qin, N., & Guan, L. (2019). A Novel Cold-adapted Endoglucanase (M6A) from *Microbacterium kitamiense* S12 Isolated from Qinghai-Tibetan Plateau. *Biotechnology and Bioprocess Engineering*, 24(3), 544–551. 10.1007/s12257-019-0009-6
- Liu, L., Huang, W., Liu, Y., & Li, M. (2021). Diversity of cellulolytic microorganisms and microbial cellulases. *International Biodeterioration & Biodegradation*, 163, 105277. 10.1016/j.ibiod.2021.105277
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42(D1), D490–D495. 10.1093/nar/gkt1178
- Luan, T., Commichaux, S., Hoffmann, M., Jayeola, V., Jang, J. H., Pop, M., Rand, H., & Luo, Y. (2024). Benchmarking short and long read polishing tools for nanopore assemblies: achieving near-perfect genomes for outbreak isolates. *BMC Genomics*, 25(1), 679–17. 10.1186/s12864-024-10582-x
- Ma, L., Aizhan, R., Wang, X., Yi, Y., Shan, Y., Liu, B., Zhou, Y., & Lü, X. (2020). Cloning and characterization of low-temperature adapted GH5-CBM3 endo-cellulase from *Bacillus subtilis* 1AJ3 and their application in the saccharification of switchgrass and coffee grounds. *AMB Express*, 10(1), 42–11. 10.1186/s13568-020-00975-y

- Malik, W. A., & Javed, S. (2021). Biochemical Characterization of Cellulase from *Bacillus subtilis* Strain and its Effect on Digestibility and Structural Modifications of Lignocellulose Rich Biomass. *Frontiers in Bioengineering and Biotechnology*, 910.3389/fbioe.2021.800265
- Mbandi, S. K., Hesse, U., Rees, D. J. G., & Christoffels, A. (2014). A glance at quality score: implication for de novo transcriptome reconstruction of Illumina reads. *Frontiers in Genetics*, 5, 17. 10.3389/fgene.2014.00017
- McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), 404–405. 10.1093/bioinformatics/16.4.404
- Meddeb-Mouelhi, F., Moisan, J. K., & Beauregard, M. (2014). A comparison of plate assay methods for detecting extracellular cellulase and xylanase activity. *Enzyme and Microbial Technology*, 66, 16–19. 10.1016/j.enzmictec.2014.07.004
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, 14(1), 60. 10.1186/1471-2105-14-60
- Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*, 10(1), 2182–10. 10.1038/s41467-019-10210-3
- Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., & Göker, M. (2022). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Research*, 50(D1), 801. 10.1093/nar/gkab902
- Melnikova, N. V., Pushkova, E. N., Dvorianinova, E. M., Beniaminov, A. D., Novakovskiy, R. O., Povkhova, L. V., Bolsheva, N. L., Snezhkina, A. V., Kudryavtseva, A. V., Krasnov, G. S., & Dmitriev, A. A. (2021). Genome Assembly and Sex-Determining Region of Male and

- Female *Populus* × *sibirica*. *Frontiers in Plant Science*, 12, 625416. 10.3389/fpls.2021.625416
- Menshawy, M. N., Abdel-Hamid, A. M., Mohamed, S. K., & El-Katatny, M. H. (2022). Isolation and molecular identification of cellulose/hemicellulose degrading bacteria from agricultural compost and determination of their hydrolytic potential. *South African Journal of Botany*, 149, 617–621. 10.1016/j.sajb.2022.06.045
- Merda, D., Vila-Nova, M., Bonis, M., Boutigny, A., Brauge, T., Cavaiuolo, M., Cuntty, A., Regnier, A., Sayeb, M., Vingadassalon, N., Yvon, C., & Chesnais, V. (2024). Unraveling the impact of genome assembly on bacterial typing: a one health perspective. *BMC Genomics*, 25(1), 1059–13. 10.1186/s12864-024-10982-z
- Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, 34(13), i142–i150. 10.1093/bioinformatics/bty266
- Miller, G. L. (1959). Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, 31(3), 426–428. 10.1021/ac60147a030
- Mo, Y., Huang, X., Yue, M., Hu, L., & Hu, C. (2024). Preparation of nanocellulose and application of nanocellulose polyurethane composites. *RSC Advances*, 14, 18247–18257. <https://doi.org/10.1039/D4RA01412J>
- Mohamad Sobri, M. F., Abd-Aziz, S., Abu Bakar, F. D., & Ramli, N. (2020). In-Silico Characterization of Glycosyl Hydrolase Family 1 β -Glucosidase from *Trichoderma asperellum* UPM1. *International Journal of Molecular Sciences*, 21(11), 4035. 10.3390/ijms21114035

- Montero, M., Eydallin, G., Viale, A., Almagro, G., Muñoz, F., Rahimpour, M., Sesma, M., Baroja-Fernández, E., & Pozueta-Romero, J. (2009). *Escherichia coli* glycogen metabolism is controlled by the PhoP-PhoQ regulatory system at submillimolar environmental Mg²⁺ concentrations, and is highly interconnected with a wide variety of cellular processes. *Biochemical Journal*, 424(1), 129–141. 10.1042/BJ20090980
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16), 2785–2791. 10.1002/jcc.21256
- Myers, J. A., Curtis, B. S., & Curtis, W. R. (2013). Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophysics*, 6(1), 4–4. 10.1186/2046-1682-6-4
- Ogura, J., Toyoda, A., Kurosawa, T., Chong, A. L., Chohnan, S., & Masaki, T. (2006). Purification, Characterization, and Gene Analysis of Cellulase (Cel8A) from *Lysobacter* sp. IB-9374. *Bioscience, Biotechnology, and Biochemistry*, 70(10), 2420–2428. 10.1271/bbb.60157
- Okano, H., Kanaya, E., Ozaki, M., Angkawidjaja, C., & Kanaya, S. (2015). Structure, activity, and stability of metagenome-derived glycoside hydrolase family 9 endoglucanase with an N-terminal Ig-like domain. *Protein Science*, 24(3), 408–419. 10.1002/pro.2632
- Padilha, I. Q. M., Carvalho, L. C. T., Dias, P. V. S., Grisi, T. C. S. L., Silva, F. L. H. d., Santos, S. F. M., & Araújo, D. A. M. (2015). Production and Characterization of Thermophilic Carboxymethyl Cellulase Synthesized by *Bacillus* sp. Growing on Sugarcane Bagasse in Submerged Fermentation. *Brazilian Journal of Chemical Engineering*, 32(1), 35–42. 10.1590/0104-6632.20150321s00003298

- Palma, L. (2012). *Kinetic and bioinformatics characterization of cellulases from the metagenomic library of GEBIX specific to African oil palm (*Elaeis guineensis* Jacq.) empty fruit bunch* <https://tinyurl.com/557jxhsh>
- Park, S., Lee, B., & Park, K. (2017). Extremophilic Carbohydrate Active Enzymes (CAZymes). *Journal of Nutritional Health & Food Engineering*, 7(1)10.15406/jnhfe.2017.07.00230
- Park, Y. W., Tominaga, R., Sugiyama, J., Furuta, Y., Tanimoto, E., Samejima, M., Sakai, F., & Hayashi, T. (2003). Enhancement of growth by expression of poplar cellulase in *Arabidopsis thaliana*. *The Plant Journal*, 33(6), 1099–1106. 10.1046/j.1365-313X.2003.01696.x
- Paszkiwicz, K., & Studholme, D. J. (2010). De novo assembly of short sequence reads. *Briefings in Bioinformatics*, 11(5), 457–472. 10.1093/bib/bbq020
- Peracchi, A. (2018). The Limits of Enzyme Specificity and the Evolution of Metabolism. *Trends in Biochemical Sciences*, 43(12), 984–996. 10.1016/j.tibs.2018.09.015
- Petkun, S., Rozman Grinberg, I., Lamed, R., Jindou, S., Burstein, T., Yaniv, O., Shoham, Y., Shimon, L. J. W., Bayer, E. A., & Frolow, F. (2015). Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: structural and functional significance of the intermodular linker. *PeerJ (San Francisco, CA)*, 3, e1126. 10.7717/peerj.1126
- Phitsuwan, P., Lee, S., San, T., & Ratanakhanokchai, K. (2021). CalkGH9T: A Glycoside Hydrolase Family 9 Enzyme from *Clostridium alkalicellulosi*. *Catalysts*, 11(8), 1011. 10.3390/catal11081011
- Puentes-Cala, E., Atehortúa-Bueno, M., Tapia-Perdomo, V., Navarro-Escalante, L., Hernández-Torres, J., & Castillo-Villamizar, G. (2023). First insights into the gut microbiome of *Diatraea*

- saccharalis*: From a sugarcane pest to a reservoir of new bacteria with biotechnological potential. *Frontiers in Ecology and Evolution*, 1110.3389/fevo.2023.1027527
- Purohit, A., Pawar, L., & Yadav, S. K. (2023). Structural and functional insights of a cold-adaptive β -glucosidase with very high glucose tolerance from *Microbacterium* sp. CIAB417. *Enzyme and Microbial Technology*, 169, 110284. 10.1016/j.enzmictec.2023.110284
- Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. *Essays in Biochemistry*, 59, 1–41. 10.1042/bse0590001
- Sakon, J., Irwin, D., Wilson, D. B., & Karplus, P. A. (1997). Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Structural Biology*, 4(10), 810–818. 10.1038/nsb1097-810
- Sanjaya, R. E., Putri, K. D. A., Kurniati, A., Rohman, A., & Puspaningsih, N. N. T. (2021). In silico characterization of the GH5-cellulase family from uncultured microorganisms: physicochemical and structural studies. *Journal of Genetic Engineering and Biotechnology*, 19(1), 143–21. 10.1186/s43141-021-00236-w
- Saroj, P., P, M., & Narasimhulu, K. (2022). Biochemical Characterization of Thermostable Carboxymethyl Cellulase and β -Glucosidase from *Aspergillus fumigatus* JCM 10253. *Applied Biochemistry and Biotechnology*, 194(6), 2503–2527. 10.1007/s12010-022-03839-2
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. 10.1093/bioinformatics/btu153
- Sigova, E. A., Dvorianinova, E. M., Arkhipov, A. A., Rozhmina, T. A., Kudryavtseva, L. P., Kaplun, A. M., Bodrov, Y. V., Pavlova, V. A., Borkhert, E. V., Zhernova, D. A., Pushkova, E. N., Melnikova, N. V., & Dmitriev, A. A. (2024). Nanopore Data-Driven T2T Genome

- Assemblies of *Colletotrichum lini* Strains. *Journal of Fungi (Basel)*, 10(12), 874. 10.3390/jof10120874
- Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., & Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, 41(D1), D344–D347. 10.1093/nar/gks1067
- Singhania, R. R., Ruiz, H. A., Awasthi, M. K., Dong, C., Chen, C., & Patel, A. K. (2021). Challenges in cellulase bioprocess for biofuel applications. *Renewable and Sustainable Energy Reviews*, 151, 111622. 10.1016/j.rser.2021.111622
- Soueidan, H., Maurier, F., Groppi, A., Sirand-Pugnet, P., Tardy, F., Citti, C., Dupuy, V., & Nikolski, M. (2013). Finishing bacterial genome assemblies with Mix. *BMC Bioinformatics*, 14(S15), S16. 10.1186/1471-2105-14-S15-S16
- Sukumaran, R., Singhania, R., & Pandey, A. (2005). Microbial cellulases - Production, applications and challenges. *Journal of Scientific & Industrial Research*, 64(11), 832–844.
- Sulyman, A. O., Igunnu, A., & Malomo, S. O. (2020). Isolation, purification and characterization of cellulase produced by *Aspergillus niger* cultured on *Arachis hypogaea* shells. *Heliyon*, 6(12), e05668. 10.1016/j.heliyon.2020.e05668
- Swanson, N. L., Billard, B. D., & Gennaro, T. L. (1999). Limits of optical transmission measurements with application to particle sizing techniques. *Applied Optics*, 38(27), 5887–5893. 10.1364/AO.38.005887
- Teng, Y., Tao, X., Lu, W., Huang, Y., Xu, D., Li, M., & Fan, Y. (2020). Identification of Hidradenitis Suppurativa-Related mRNA Expression Patterns Through Analysis of Gene Expression Omnibus. *Dose-Response*, 18(3)10.1177/1559325820942646

- Teunissen, M. J., Kets, E. P. W., Op den Camp, H. J. M., Huis in't Veld, J. H. J., & Vogels, G. D. (1992). Effect of coculture of anaerobic fungi isolated from ruminants and non-ruminants with methanogenic bacteria on cellulolytic and xylanolytic enzyme activities. *Archives of Microbiology*, 157(2), 176–182. 10.1007/BF00245287
- The Galaxy Community. (2024). The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*, 52(W1), W83–W94. 10.1093/nar/gkae410
- Trisakul, K., Hinwan, Y., Eisiri, J., Salao, K., Chairprasert, A., Kamolwat, P., Tongsimma, S., Campino, S., Phelan, J., Clark, T. G., & Faksri, K. (2024). Comparisons of genome assembly tools for characterization of *Mycobacterium tuberculosis* genomes using hybrid sequencing technologies. *PeerJ (San Francisco, CA)*, 12, e17964. 10.7717/peerj.17964
- Ugras, S., Bicen, H. E. I., & Emire, Z. (2024). Determination of Cellulase Enzyme Produced by *Bacillus cereus* DU-1 Isolated from Soil, and its Effects on Cotton Fiber. *Brazilian Archives of Biology and Technology*, 6710.1590/1678-4324-2024230391
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE*, 9(11), e112963. 10.1371/journal.pone.0112963
- Wang, S., Li, W., Liu, S., & Xu, J. (2016). RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Research*, 44(W1), W430–W435. 10.1093/nar/gkw306
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL:

- homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. 10.1093/nar/gky427
- Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*, 13(6), e1005595. 10.1371/journal.pcbi.1005595
- Wilson, W. A., Roach, P. J., Montero, M., Baroja-Fernández, E., Muñoz, F. J., Eydallin, G., Viale, A. M., & Pozueta-Romero, J. (2010). Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiology Reviews*, 34(6), 952–985. 10.1111/j.1574-6976.2010.00220.x
- Wu, L., & Ma, J. (2019). The Global Catalogue of Microorganisms (GCM) 10K type strain sequencing project: providing services to taxonomists for standard genome sequencing and annotation. *International Journal of Systematic and Evolutionary Microbiology*, 69(4), 895–898. 10.1099/ijsem.0.003276
- Wu, S., & Zhang, Y. (2007). LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research*, 35(10), 3375–3382. 10.1093/nar/gkm251
- Xing, M., Zhang, X., & Huang, H. (2012). Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. *Biotechnology Advances*, 30(4), 920–929. 10.1016/j.biotechadv.2012.01.021
- Xu, D., & Zhang, Y. (2011). Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-Step Atomic-Level Energy Minimization. *Biophysical Journal*, 101(10), 2525–2534. 10.1016/j.bpj.2011.10.024
- Xu, J., Zhang, H., Zheng, J., Dovoedo, P., & Yin, Y. (2020). eCAMI: simultaneous classification and motif identification for enzyme annotation. *Bioinformatics*, 36(7), 2068–2075. 10.1093/bioinformatics/btz908

- Yennamalli, R. M., Rader, A. J., Wolt, J. D., & Sen, T. Z. (2011). Thermostability in endoglucanases is fold-specific. *BMC Structural Biology*, *11*(1), 10. 10.1186/1472-6807-11-10
- Yin, L., Huang, P., & Lin, H. (2010). Isolation of Cellulase-Producing Bacteria and Characterization of the Cellulase from the Isolated Bacterium *Cellulomonas* Sp. YJ5. *Journal of Agricultural and Food Chemistry*, *58*(17), 9833–9837. 10.1021/jf1019104
- Zhang, J., Lei, C., Liu, G., Bao, Y., Balan, V., & Bao, J. (2017). *In - Situ* Vacuum Distillation of Ethanol Helps to Recycle Cellulase and Yeast during SSF of Delignified Corncob Residues. *ACS Sustainable Chemistry & Engineering*, *5*(12), 11676–11685. 10.1021/acssuschemeng.7b03084
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(1), 40. 10.1186/1471-2105-9-40
- Zhao, Y., Olsen, J. E., Poulsen, L. L., & Christensen, H. (2025). *Comparison of genomic assembly and annotation based on two clones of avian pathogenic *Escherichia coli**. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. 10.1101/2024.11.22.624809
<https://www.proquest.com/docview/3131948743>
- Zheng, J., Ge, Q., Yan, Y., Zhang, X., Huang, L., & Yin, Y. (2023). dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research*, *51*(W1), W115–W121. 10.1093/nar/gkad328
- Zmasek, C. M., & Godzik, A. (2014). Phylogenomic analysis of glycogen branching and debranching enzymatic duo. *BMC Evolutionary Biology*, *14*(1), 183. 10.1186/s12862-014-0183-2